## Mathematical Approaches in Advanced Control Theories 2013

Guest Editors: Baocang Ding, Weihai Zhang, Tianming Liu, Xianxia Zhang, Dewei Li, and Tao Zou

# Mathematical Approaches in Advanced Control Theories 2013 

## Journal of Applied Mathematics

## Mathematical Approaches in Advanced Control Theories 2013

Guest Editors: Baocang Ding, Weihai Zhang, Tianming Liu, Xianxia Zhang, Dewei Li, and Tao Zou

Copyright © 2014 Hindawi Publishing Corporation. All rights reserved.
This is a special issue published in "Journal of Applied Mathematics." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Saeid Abbasbandy, Iran
Mina B. Abd-El-Malek, Egypt
Mohamed A. Abdou, Egypt
Subhas Abel, India
Carlos J. S. Alves, Portugal
Mohamad Alwash, USA
Igor Andrianov, Germany
Sabri Arik, Turkey
Ali R. Ashrafi, Iran
Francis T.K. Au, Hong Kong
Olivier Bahn, Canada
Antonio Bandera, Spain
Roberto Barrio, Spain
Alfredo Bellen, Italy
Jafar Biazar, Iran
Anjan Biswas, Saudi Arabia
Stephane P.A. Bordas, USA
James Robert Buchanan, USA
Alberto Cabada, Spain
Xiao Chuan Cai, USA
Zhenfu Cao, China
Jinde Cao, China
Yijia Cao, China
Alexandre Carvalho, Brazil
Song Cen, China
Tai-Ping Chang, Taiwan
Wei-Der Chang, Taiwan
Shih-sen Chang, China
Xinkai Chen, Japan
Rushan Chen, China
Ke Chen, UK
Jin Cheng, China
Eric Cheng, Hong Kong
Ching-Hsue Cheng, Taiwan
Qi Cheng, USA
Francisco Chiclana, UK
Jen-Tzung Chien, Taiwan
Han H. Choi, Republic of Korea
S. H. Chowdhury, Malaysia

Hung-Yuan Chung, Taiwan Carlos Conca, Chile
Vitor Costa, Portugal
Livija Cveticanin, Serbia
Orazio Descalzi, Chile
Kai Diethelm, Germany

Urmila Diwekar, USA
Vit Dolejsi, Czech Republic
Bo-Qing Dong, China
Magdy A. Ezzat, Egypt
Meng Fan, China
Ya Ping Fang, China
Didier Felbacq, France
Antonio J. M. Ferreira, Portugal
Michel Fliess, France
Marco A. Fontelos, Spain
Huijun Gao, China
Xin-Lin Gao, USA
B. J. Geurts, The Netherlands

Pablo González-Vera, Spain
Laurent Gosse, Italy
Keshlan S. Govinder, South Africa
Jose L. Gracia, Spain
Zhi-Hong Guan, China
Nicola Guglielmi, Italy
Kerim Guney, Turkey
Vijay Gupta, India
Saman K. Halgamuge, Australia
Abdelmagid S. Hamouda, Qatar
Maoan Han, China
Bo Han, China
Pierre Hansen, Canada
Ferenc Hartung, Hungary
Xiao-Qiao He, China
Luis J. Herrera, Spain
Ying Hu, France
Ning Hu, Japan
Jianguo Huang, China
Dan Huang, China
Zhilong L. Huang, China
Gerardo Iovane, Italy
Anuar Ishak, Malaysia
Takeshi Iwamoto, Japan
George Jaiani, Georgia
Zhongxiao Jia, China
Xing Jin, China
Zhen Jin, China
Zlatko Jovanoski, Australia
Tadeusz Kaczorek, Poland
Ido Kanter, Israel
Abdul Hamid Kara, South Africa

Hamid Reza Karimi, Norway
Dogan Kaya, Turkey
Chaudry Masood Khalique, South Africa
Khalil Khanafer, USA
Hyunsung Kim, Korea
Younjea Kim, Republic of Korea
Jong Hae Kim, Republic of Korea
Kazutake Komori, Japan
Vadim A. Krysko, Russia
Jin L. Kuang, Singapore
Miroslaw Lachowicz, Poland
Hak-Keung Lam, UK
Tak-Wah Lam, Hong Kong
Peter G L Leach, Cyprus
Jaehong Lee, Republic of Korea
Jinsong Leng, China
Edson D. Leonel, Brazil
Xiang Li, China
Yongkun Li, China
Wan-Tong Li, China
Jin Liang, China
Ching-Jong Liao, Taiwan
Chong Lin, China
Yansheng Liu, China
Shutian Liu, China
Kang Liu, USA
Weiqing Liu, China
Chongxin Liu, China
Chein-Shan Liu, Taiwan
Tao Liu, China
Zhijun Liu, China
Fei Liu, China
Jose L. López, Spain
Shiping Lu, China
Hongbing Lu, China
Li Ma, China
Ruyun Ma, China
Nazim I. Mahmudov, Turkey
Oluwole Daniel Makinde, South Africa
Francisco J. Marcellán, Spain
Guiomar Martín-Herrán, Spain
Nicola Mastronardi, Italy
Panayotis Takis Mathiopouloss, Greece
Michael McAleer, The Netherlands
Stephane Metens, France

Michael Meylan, Australia
Alain Miranville, France
Ram N. Mohapatra, USA
Cristinel Mortici, Romania
Jaime E. Munoz Rivera, Brazil
Javier Murillo, Spain
Roberto Natalini, Italy
Srinivasan Natesan, India
Roger Ohayon, France Javier Oliver, Spain Donal O'Regan, Ireland
M. Ostoja-Starzewski, USA

Turgut ÖziŞ, Turkey
Claudio Padra, Argentina
R. M. Palhares, Brazil

Francesco Pellicano, Italy
Juan Manuel Peña, Spain
Ricardo Perera, Spain
Malgorzata Peszynska, USA
Allan C. Peterson, USA
Vu Ngoc Phat, Vietnam
Andrew Pickering, Spain
Hector Pomares, Spain
Mario Primicerio, Italy
Morteza Rafei, The Netherlands
Laura Rebollo-Neira, UK
Roberto Renò, Italy
Ignacio Rojas, Spain
Carla Roque, Portugal
Debasish Roy, India
Samir Saker, Egypt
Miguel A. F. Sanjuan, Spain
Wolfgang Schmidt, Germany
Mehmet Sezer, Turkey
Naseer Shahzad, Saudi Arabia
Hui-Shen Shen, China

Jian Hua Shen, China
Yong Shi, China
Yasuhide Shindo, Japan
Patrick Siarry, France
Fernando Simões, Portugal
Theodore E. Simos, Greece
Francesco Soldovieri, Italy
A.-M. A. Soliman, Egypt

Qiankun Song, China
Xinyu Song, China
Yuri N. Sotskov, Belarus
Niclas Strömberg, Sweden
Ray K.L. Su, Hong Kong
Chengjun Sun, USA
Jitao Sun, China
Wenyu Sun, China
Ying Tan, China
XianHua Tang, China
San-Yi Tang, China
Zhidong Teng, China
Engang Tian, China
Alexander Timokha, Norway
Mariano Torrisi, Italy
Jung-Fa Tsai, Taiwan
Ch. Tsitouras, Greece
Kuppalapalle Vajravelu, USA
Alvaro Valencia, Chile
Erik Van Vleck, USA
Ezio Venturino, Italy
Jesus Vigo-Aguiar, Spain
Michael N. Vrahatis, Greece
Yaonan Wang, China
Mingxin Wang, China
Guangchen Wang, China
Qing-Wen Wang, China
Baolin Wang, China

Junjie Wei, China
Li Weili, China
Martin Weiser, Germany
Frank Werner, Germany
Shanhe Wu, China
Shi-Liang Wu, China
Yonghui Xia, China
Tiecheng Xia, China
Gongnan Xie, China
Daoyi Xu, China
Yuesheng Xu, USA
Chao Yan, USA
Chao Yang, China
Suh-Yuh Yang, Taiwan
Her-Terng Yau, Taiwan
Guan H. Yeoh, Australia
Bo Yu, China
Xiaohui Yuan, China
Jinyun Yuan, Brazil
Guisheng Zhai, Japan
Jianming Zhan, China
Jian-gang Zhang, China
Jifeng Zhang, China
Heping Zhang, China
Zhihua Zhang, China
Sheng Zhang, China
Liang Zhang, China
Jingxin Zhang, Australia
Chongbin Zhao, Australia
Shan Zhao, USA
Renat Zhdanov, USA
Bin Zhou, China
Goangseup Zi, Korea
J. D. Hoenderkamp, The Netherlands

## Contents

Mathematical Approaches in Advanced Control Theories 2013, Baocang Ding, Weihai Zhang, Tianming Liu, Xianxia Zhang, Dewei Li, and Tao Zou
Volume 2014, Article ID 718595, 2 pages

Innovation in Active Vibration Control Strategy of Intelligent Structures, A. Moutsopoulou,
G. E. Stavroulakis, and A. Pouliezos

Volume 2014, Article ID 430731, 14 pages

Nonlinear Fault Separation for Redundancy Process Variables Based on FNN in MKFDA Subspace,
Ying-ying Su, Shan Liang, Jing-zhe Li, Xiao-gang Deng, Tai-fu Li, and Cheng Zeng
Volume 2014, Article ID 729763, 9 pages

Optimal Manoeuvres of Underactuated Linear Mechanical Systems: The Case of Controlling Gantry Crane Operations, S. Woods and W. Szyszkowski
Volume 2014, Article ID 283565, 16 pages

Exponential $L_{2}-L_{\infty}$ Filtering for a Class of Stochastic System with Mixed Delays and Nonlinear Perturbations, Zhaohui Chen and Qi Huang Volume 2013, Article ID 659251, 10 pages

Improving the Asymptotic Properties of Discrete System Zeros in Fractional-Order Hold Case, Cheng Zeng, Shan Liang, and Yingying Su
Volume 2013, Article ID 268684, 16 pages

Modeling and Parameter Analysis of the OC3-Hywind Floating Wind Turbine with a Tuned Mass Damper in Nacelle, Yulin Si, Hamid Reza Karimi, and Huijun Gao Volume 2013, Article ID 679071, 10 pages

Reference Function Based Spatiotemporal Fuzzy Logic Control Design Using Support Vector Regression Learning, Xian-Xia Zhang, Ye Jiang, Shiwei Ma, and Bing Wang Volume 2013, Article ID 410279, 13 pages

Novel Observer-Based Suboptimal Digital Tracker for a Class of Time-Delay Singular Systems, Nien-Tsu Hu, Ter-Feng Wu, Sendren Sheng-Dun Xu, and Hsu-Chih Huang Volume 2013, Article ID 150590, 19 pages

Practical Aspects of Broken Rotor Bars Detection in PWM Voltage-Source-Inverter-Fed Squirrel-Cage Induction Motors, Hong-yu Zhu, Jing-tao Hu, Lei Gao, and Hao Huang Volume 2013, Article ID 128368, 11 pages

Multiple-Model Cardinality Balanced Multitarget Multi-Bernoulli Filter for Tracking Maneuvering
Targets, Xianghui Yuan, Feng Lian, and Chongzhao Han
Volume 2013, Article ID 727430, 16 pages

Sign Inference for Dynamic Signed Networks via Dictionary Learning, Yi Cen, Rentao Gu, and Yuefeng Ji Volume 2013, Article ID 708581, 10 pages

Adaptive Correction Forecasting Approach for Urban Traffic Flow Based on Fuzzy c-Mean Clustering and Advanced Neural Network, He Huang, Qifeng Tang, and Zhen Liu
Volume 2013, Article ID 195824, 7 pages

Robust Fault Detection of Linear Uncertain Time-Delay Systems Using Unknown Input Observers, Saeed Ahmadizadeh, Jafar Zarei, and Hamid Reza Karimi
Volume 2013, Article ID 829715, 12 pages

Adaptive Neural Network Dynamic Inversion with Prescribed Performance for Aircraft Flight Control, Wendong Gai, Honglun Wang, Jing Zhang, and Yuxia Li
Volume 2013, Article ID 452653, 12 pages

Control Strategy Based on Wavelet Transform and Neural Network for Hybrid Power System,
Y. D. Song, Qian Cao, Xiaoqiang Du, and Hamid Reza Karimi

Volume 2013, Article ID 375840, 8 pages

A Hybrid Multiobjective Differential Evolution Algorithm and Its Application to the Optimization of Grinding and Classification, Yalin Wang, Xiaofang Chen, Weihua Gui, Chunhua Yang, Lou Caccetta, and Honglei Xu
Volume 2013, Article ID 841780, 15 pages

Discrete Current Control Strategy of Permanent Magnet Synchronous Motors, Yan Dong, Kai Jing, Hexu Sun, and Yi Zheng
Volume 2013, Article ID 525014, 9 pages

Cylinder Position Servo Control Based on Fuzzy PID, Shibo Cai, Shunlei Wu, and Guanjun Bao Volume 2013, Article ID 375483, 10 pages

Parameter Matching Analysis of Hydraulic Hybrid Excavators Based on Dynamic Programming Algorithm, Wei Shen, Jihai Jiang, Xiaoyu Su, and Hamid Reza Karimi
Volume 2013, Article ID 615608, 10 pages

Neural Network Predictive Control for Vanadium Redox Flow Battery, Hai-Feng Shen, Xin-Jian Zhu, Meng Shao, and Hong-fei Cao
Volume 2013, Article ID 538237, 7 pages

Tree-Based Backtracking Orthogonal Matching Pursuit for Sparse Signal Reconstruction, Yigang Cen, Fangfei Wang, Ruizhen Zhao, Lihong Cui, Lihui Cen, Zhenjiang Miao, and Yanming Cen
Volume 2013, Article ID 864132, 8 pages

Nonlinear $H_{\infty}$ Optimal Control Scheme for an Underwater Vehicle with Regional Function Formulation, Zool H. Ismail and Matthew W. Dunnigan
Volume 2013, Article ID 732738, 7 pages

Parametric Approach to Trajectory Tracking Control of Robot Manipulators, Shijie Zhang and Yi Ning Volume 2013, Article ID 958597, 6 pages

Generalized Kalman-Yakubovich-Popov Lemma Based I-PD Controller Design for Ball and Plate System, Shuichi Mochizuki and Hiroyuki Ichihara
Volume 2013, Article ID 854631, 9 pages

Robust Optimal Sliding-Mode Tracking Control for a Class of Uncertain Nonlinear MIMO Systems, Haiping Pang and Xiuqin Yang
Volume 2013, Article ID 863168, 9 pages

Distributed Filter with Consensus Strategies for Sensor Networks, Xie Li, Huang Caimou, and Hu Haoji Volume 2013, Article ID 683249, 9 pages

Sliding-Mode Control of a Dc/Dc Postfilter for Ripple Reduction and Efficiency Improvement in POL Applications, S. I. Serna-Garcés, R. E. Jiménez, and C. A. Ramos-Paja
Volume 2013, Article ID 915146, 10 pages

Incomplete Phase Space Reconstruction Method Based on Subspace Adaptive Evolution Approximation, Tai-fu Li, Wei Jia, Wei Zhou, Ji-ke Ge, Yu-cheng Liu, and Li-zhong Yao Volume 2013, Article ID 983051, 9 pages

Two-Layer Predictive Control of a Continuous Biodiesel Transesterification Reactor, Hongyan Shi, Dingding Wang, Decheng Yuan, and Tianran Wang
Volume 2013, Article ID 587841, 14 pages

Research on the WSN Node Localization Based on TOA, Qing-hui Wang, Ting-ting Lu, Meng-long Liu, and Li-feng Wei
Volume 2013, Article ID 706064, 6 pages

Adaptive Semidiscrete Finite Element Methods for Semilinear Parabolic Integrodifferential Optimal Control Problem with Control Constraint, Zuliang Lu
Volume 2013, Article ID 302935, 11 pages

Switching Signal Design for Exponential Stability of Uncertain Discrete-Time Switched Time-Delay Systems, Zheng-Fan Liu, Chen-Xiao Cai, and Yun Zou
Volume 2013, Article ID 416292, 11 pages

Stability Analysis of Networked Control Systems with Random Time Delays and Packet Dropouts Modeled by Markov Chains, Li Qiu, Fengqi Yao, and Xiaopin Zhong
Volume 2013, Article ID 715072, 10 pages

Nonlinear Adaptive Equivalent Control Based on Interconnection Subsystems for Air-Breathing
Hypersonic Vehicles, Chaofang Hu and Yanwen Liu
Volume 2013, Article ID 560785, 10 pages
Stability of Nonlinear Stochastic Discrete-Time Systems, Yan Li, Weihai Zhang, and Xikui Liu Volume 2013, Article ID 356746, 8 pages

Output-Feedback and Inverse Optimal Control of a Class of Stochastic Nonlinear Systems with More
General Growth Conditions, Liu Jianwei, Guo Longchuan, Zuo Xin, and Liang Huaqing
Volume 2013, Article ID 251340, 10 pages
A Multilayer Feed Forward Small-World Neural Network Controller and Its Application on Electrohydraulic Actuation System, Xiaohu Li, Feng Xu, Jinhua Zhang, and Sunan Wang
Volume 2013, Article ID 872790 , 8 pages

A Novel Coherence Reduction Method in Compressed Sensing for DOA Estimation, Jing Liu,
ChongZhao Han, XiangHua Yao, and Feng Lian
Volume 2013, Article ID 548979, 9 pages
Splitting Matching Pursuit Method for Reconstructing Sparse Signal in Compressed Sensing, Liu Jing, Han ChongZhao, Yao XiangHua, and Lian Feng
Volume 2013, Article ID 804640, 8 pages

## Editorial

# Mathematical Approaches in Advanced Control Theories 2013 

Baocang Ding, ${ }^{1}$ Weihai Zhang, ${ }^{2}$ Tianming Liu, ${ }^{3}$ Xianxia Zhang, ${ }^{4}$ Dewei Li, ${ }^{5}$ and Tao Zou ${ }^{6}$<br>${ }^{1}$ Department of Automation, School of Electronic and Information Engineering, Xian Jiaotong University, Xian 710049, China<br>${ }^{2}$ College of Information and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266590, China<br>${ }^{3}$ Department of Computer Science and Bioimaging Research Center, University of Georgia, Boyd SRC 420, Athens, GA 30602, USA<br>${ }^{4}$ School of Mechatronics Engineering and Automation, Shanghai Key Laboratory of Power Station Automation Technology, Shanghai University, Shanghai 200072, China<br>${ }^{5}$ Department of Automation, Shanghai Jiaotong University, Shanghai 200240, China<br>${ }^{6}$ Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China<br>Correspondence should be addressed to Baocang Ding; baocang.ding@gmail.com

Received 17 April 2014; Accepted 17 April 2014; Published 28 April 2014
Copyright © 2014 Baocang Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Advanced control theory fills a gap between the mathematical control theory and modern control engineering practices. Conceptually, advanced control theories can include any theoretical problems related to the controller design. But in this issue it may include model predictive control, sliding mode control, robust control, real-time optimization, and identification and estimation, which are not limited to controller design. Advanced control technologies have become ubiquitous in various engineering applications (e.g., chemical process control, robot control, air traffic control, vehicle control, multiagent control, and networked control). The development of mathematical methods is essential for the applications of advanced control theories. Sometimes it lacks effective methods to tackle the computational issue (e.g., model predictive control of a fast process). Sometimes, a new application requires a brand-new solver for applying the advanced control theory (e.g., a new production line far exceeding the usual speed). The main focus of this special issue will be on the new research ideas and results for the mathematical problems in advanced control theories.

A total number of 83 papers were submitted for this special issue. Out of the submitted papers, 39 contributions have been included in this special issue. The 39 contributions consider several closely related and interesting topics.

The subjects in controller design/synthesis and system analysis have occupied 24 contributions. These contributions include, for example, adaptive control (see the work of C.

Hu and Y. Liu for the air-breathing hypersonic vehicles and the work of W. Gai et al. for the neural network dynamic inversion with prescribed performance in aircraft flight control), $H_{\infty}$ control (see the work of A. Moutsopoulou et al. for the active vibration control in intelligent structures and the work of Z . H. Ismail and M. W. Dunnigan for the robust technique for an autonomous underwater vehicle with region tracking function), model predictive control (see the work of H. Shen et al. for the vanadium redox flow battery modeled by neural network and the work of H . Shi et al. for the twolayered control of a continuous biodiesel transesterification reactor), sliding model control (see the work of H. Pang and X. Yang for robustifying the linear quadratic tracking controller and the work of S. I. Serna-Garcés et al. for an active postfilter based on two buck converters), networked control (see the work of L. Qiu et al. for the stability under random time delays and packet dropouts based on unified Markov jump model), backstepping technique (see the work of J. Liu et al. for output-feedback stabilization of stochastic nonlinear systems), fuzzy logic control (see the work of X.-X. Zhang et al. which presents a reference function based 3D design methodology using support vector regression learning), and neural network control (see the work of X. Li et al. which is designed under small world neural network model and is investigated in both linear and nonlinear controls).

Closely related to the controller design and synthesis are the 9 contributions on the estimation problem. These
contributions include, for example, time series prediction (see the work of H. Huang et al. for forecasting the urban traffic flow modeled by the fuzzy clustering and neural network), the compressed sensing (see the two works of J. Liu et al. for the direction of arrival estimation problem in phased array radar system and for discussing splitting matching pursuit method in reconstructing sparse signal), Kalman filter (see the work of X. Yuan et al. for integrating the cardinality balanced multitarget multi-Bernoulli filter with the interacting multiple models algorithm), and robust filter (see the work of Z. Chen and Q. Huang for the $L_{2}-L_{\infty}$ filter design for stochastic systems with mixed delays and nonlinear perturbations).

There are also 3 contributions on the fault diagnosis/detection/separation. These contributions can be seen as the extensions of the estimation problem, in the context of this special issue. For example, H. Zhu et al. propose a method for broken rotor bars detection in the voltage-source-inverter-fed squirrel-cage induction motors, and Y. Su et al. introduce an improved kernel Fisher distinguish analysis method for the nonlinear fault separation of redundancy process variables.

The last 3 contributions are for the mathematical programming (including the heuristic programming). For example, Q. Wang et al. consider the wireless sensor networks node localization based on the time of arrival; W. Shen et al. apply dynamic programming algorithm to the parameter matching analysis of hydraulic hybrid excavators; and Y. Wang et al. propose a hybrid differential evolution algorithm with multipopulation and apply it to solve a multiobjective optimization model of a grinding and classification process. Note that several other contributions mentioned above have also considered optimizations.

In the above, some contributions are included in the statistics but not mentioned, due to either more specific or more compounded technicalities. We hope the readers of this special issue will find it interesting and stimulating and expect that the included papers will contribute to further advance the area of advanced control.

## Acknowledgments

Finally, we would like to thank all the authors who have submitted papers to the special issue and the reviewers involved in the refereeing of the submissions.

Baocang Ding
Weihai Zhang
Tianming Liu
Xianxia Zhang
Dewei Li
Tao Zou

## Research Article

# Innovation in Active Vibration Control Strategy of Intelligent Structures 

A. Moutsopoulou, ${ }^{1}$ G. E. Stavroulakis, ${ }^{2}$ and A. Pouliezos ${ }^{2}$<br>${ }^{1}$ Department of Engineering (Civil), Technological Educational Institute of Crete, 71004 Heraklion, Crete, Greece<br>${ }^{2}$ Department of Production Engineering and Management, Technical University of Crete, 73100 Chania, Crete, Greece

Correspondence should be addressed to A. Moutsopoulou; amalia@staff.teicrete.gr
Received 19 July 2013; Revised 13 November 2013; Accepted 14 November 2013; Published 6 March 2014
Academic Editor: Weihai Zhang
Copyright © 2014 A. Moutsopoulou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Large amplitudes and attenuating vibration periods result in fatigue, instability, and poor structural performance. In light of past approaches in this field, this paper intends to discuss some innovative approaches in vibration control of intelligent structures, particularly in the case of structures with embedded piezoelectric materials. Control strategies are presented, such as the linear quadratic control theory, as well as more advanced theories, such as robust control theory. The paper presents sufficiently a recognizable advance in knowledge of active vibration control in intelligent structures.


## 1. Introduction

Modeling of an ideal system consisting of a beam with piezoelectric layers was done using one-dimensional finite elements with two degrees of freedom per node. Cubic and quadratic Hermit polynomials were employed for the representation of nodal rotations and vertical displacements. System differential equations are derived from Euler Bernoulli theory [1, 2].

Setting up the problem in a two-port diagram (inputoutput) was not a trivial task. The classic control problem was transformed into a two-port problem. The goal of nominal design was to keep error magnitude small, despite perturbations and noise in measurements. Moreover, controller size had to be contained so as to lower energy consumption and maintain piezoelectric materials within operating limits $( \pm 500 \mathrm{~V})$. By transforming transfer functions to state space equations and by using input and output equations, state space matrices have been derived; these matrices are used for finding the optimal controller according to the LQR and $H_{\infty}$ control criterion.

Selection of the weights involved in the controller we studied was done through optimization, while wind loading and noise in measurements were appropriately modelled for this particular problem. The obtained results were very
satisfactory; beam vibration is reduced even for realistic wind measurements. Beam response results, with as well as without control, were compared for all presented control strategies.

In this paper, we address the problem of vibrations of intelligent structures. Stimuli may come from external perturbations of the system, disturbances, or excitation that may cause structural vibrations, such as wind loading or earthquakes. An intelligent structure is expected to be able to sense the vibration and counteract it in a controlled fashion, so that vibration of the system can be reduced and contained. To that end, a number of intelligent materials may be used as actuators and sensors. Piezoelectric materials, memory materials, and electrostrictive and magnetostrictive materials are such materials. In this work, we focus on the use of piezoelectric materials, given that they exhibit good sensing and actuation properties.

## 2. Research on Intelligent Structures

The following paragraphs give a deep insight into the research work done on the intelligent structures so far. Culshaw discussed the concept of smart structure, its benefits, and applications [3]. Rao and Sunar explained the use of piezo materials as sensors and actuators in sensing vibrations in


Figure 1: Smart beam.
their survey paper [4]. Hubbard and Baily have studied the application of piezoelectric materials as sensor/actuator for flexible structures [5]. Hanagud developed a Finite Element Model (FEM) for a beam with many distributed piezoceramic sensors/actuators [6].

Hwang and Park presented a new finite element (FE) modeling technique for flexible beams [7]. Continuous time and discrete time algorithms were proposed to control a thin piezoelectric structure by Bona et al. [8]. Schiehlen and Schonerstedt reported the optimal control designs for the first few vibration modes of a cantilever beam using piezoelectric sensors/actuators [9]. Choi have shown a design of position tracking sliding mode control for a smart structure [10]. Distributed controllers for flexible structures can be seen in Pourki [11].

A FEM approach was used by Benjeddou to model a sandwich beam with shear and extension piezoelectric elements [12]. The finite element model employed the displacement field of Zhang and Sun [13]. It was shown that the finite element results agree quite well with the analytical results. Raja et al. extended the finite element model of Benjeddou's research team to include a vibration control scheme [14].

## 3. Mathematical Modeling

A cantilever slender beam with rectangular cross-sections is considered. Four pairs of piezoelectric patches are embedded symmetrically at the top and the bottom surfaces of the beam, as shown in Figure 1.

The beam is from graphite-epoxy T300-976 and the piezoelectric patches are PZT G1195N. The top patches act like sensors and the bottom like actuators. The resulting composite beam is modelled by means of the classical laminated technical theory of bending. Let us assume that the mechanical properties of both the piezoelectric material and the host beam are independent in time. The thermal effects are considered to be negligible as well [15].

The beam has length $L$, width $W$, and thickness $h$. The sensors and the actuators have width $b_{\mathrm{S}}$ and $b_{\mathrm{A}}$ and thickness $h_{\mathrm{S}}$ and $h_{\mathrm{A}}$, respectively. The electromechanical parameters of the beam of interest are given in Table 1.

Table 1: Parameters of the composite beam.

| Parameters | Values |
| :--- | :---: |
| Beam length, $L$ | 0.3 m |
| Beam width, $W$ | 0.04 m |
| Beam thickness, $h$ | 0.0096 m |
| Beam density, $\rho$ | $1600 \mathrm{~kg} / \mathrm{m}^{3}$ |
| Young's modulus of the beam, $E$ | $1.5 \times 10^{11} \mathrm{~N} / \mathrm{m}^{2}$ |
| Piezoelectric constant, $d_{31}$ | $254 \times 10^{-12} \mathrm{~m} / \mathrm{V}$ |
| Electric constant, $\xi_{33}$ | $11.5 \times 10^{-3} \mathrm{~V} \mathrm{~m} / \mathrm{N}$ |
| Young's modulus of the piezoelectric element | $1.5 \times 10^{11} \mathrm{~N} / \mathrm{m}^{2}$ |
| Width of the piezoelectric element | $b_{\mathrm{S}}=b_{\mathrm{A}}=0.04 \mathrm{~m}$ |
| Thickness of the piezoelectric element | $h_{\mathrm{S}}=h_{\mathrm{A}}=0.0002 \mathrm{~m}$ |

In order to derive the basic equations for piezoelectric sensors and actuators [1], we assume that
(i) the piezoelectric sensors actuators (S/A) are bonded perfectly on the host beam;
(ii) the piezoelectric layers are much thinner than the host beam;
(iii) the piezoelectric material is homogeneous, transversely isotropic, and linearly elastic;
(iv) the piezoelectric S/A are transversely polarized [1, 3, 16].
3.1. Finite Element Formulation. We consider a beam element of length $L_{e}$, which has two mechanical degrees of freedom at each node: one translational $\omega_{1}$ (resp. $\omega_{2}$ ) in direction $z$ and one rotational $\psi_{1}$ (resp., $\psi_{2}$ ). The vector of nodal displacements and rotations $q_{e}$ is defined as [17]

$$
\begin{equation*}
q_{e}^{r}=\left[\omega_{1}, \psi_{1}, \omega_{2}, \psi_{2}\right] \tag{1}
\end{equation*}
$$

The beam element transverse deflection $\omega(x, t)$ and the beam element rotation $\psi(x, t)$ of the beam are continuous and they are interpolated within by Hermitian linear shape functions $H_{i}^{\omega}$ and $H_{i}^{\psi}$ as follows [18, 19]:

$$
\begin{align*}
& \omega(x, t)=\sum_{i=1}^{4} H_{i}^{\omega}(x) q_{i}(t), \\
& \psi(x, t)=\sum_{i=1}^{4} H_{i}^{\psi}(x) q_{i}(t) . \tag{2}
\end{align*}
$$

This classical finite element procedure leads to the approximate discretized variational problem. For a finite element, the discrete differential equations are obtained by substituting the discretized expressions into the first variation of the kinetic energy and strain energy $[18,20]$ to evaluate the kinetic and strain energies. Integrating over spatial domains and using the Hamiltons principle [20], the equation of motion for a beam element is expressed in terms of nodal variable $q$ as follows:

$$
\begin{equation*}
M \ddot{q}(t)+D \dot{q}(t)+K q(t)=f_{m}(t)+f_{e}(t), \tag{3}
\end{equation*}
$$

where $M$ is the generalized mass matrix, $D$ the viscous damping matrix, $K$ the generalized stiffness matrix, $f_{m}$
the external loading vector, and $f_{e}$ the generalized control force vector, produced by electromechanical coupling effects. The independent variable $q(t)$ is composed of transversal deflections $\omega_{1}$ and rotations $\psi_{1}$; that is, $[16,18]$

$$
q(t)=\left[\begin{array}{c}
\omega_{1}  \tag{4}\\
\psi_{1} \\
\vdots \\
\omega_{n} \\
\psi_{n}
\end{array}\right],
$$

where $n$ is the number of nodes used in analysis. Vectors $\omega$ and $f_{m}$ are positive upwards. To transform to state-space control representation, let (in the usual manner)

$$
\dot{x}(t)=\left[\begin{array}{l}
q(t)  \tag{5}\\
\dot{q}(t)
\end{array}\right] .
$$

Furthermore, to express $f_{e}(t)$ as $B u(t)$, we write it as $f_{e}^{*} u$, where $f_{e}^{*}$ the piezoelectric force is for a unit applied on the corresponding actuator and $u$ represents the voltages on the actuators. Furthermore, $d(t)=f_{m}(t)$ is the disturbance vector $[16,18]$.

Then,

$$
\begin{align*}
\dot{x}(t)= & {\left[\begin{array}{cc}
O_{2 n \times 2 n} & I_{2 n \times 2 n} \\
-M^{-1} K & -M^{-1} D
\end{array}\right] x(t) } \\
& +\left[\begin{array}{c}
O_{2 n \times 2 n} \\
M^{-1} f_{e}^{*}
\end{array}\right] u(t)+\left[\begin{array}{c}
O_{2 n \times 2 n} \\
M^{-1}
\end{array}\right] \\
= & A x(t)+B u(t)+G d(t)  \tag{6}\\
= & A x(t)+\left[\begin{array}{ll}
B & G
\end{array}\right]\left[\begin{array}{l}
u(t) \\
d(t)
\end{array}\right] \\
= & A x(t)+\widetilde{B} \widetilde{u}(t) .
\end{align*}
$$

The previous description of the dynamical system will be augmented with the output equation (displacements only measured) [17] as follows:

$$
y(t)=\left[\begin{array}{llll}
x_{1}(t) & x_{3}(t) & \cdots & x_{n-1}(t) \tag{7}
\end{array}\right]^{T}=C x(t) .
$$

In this formulation, $u$ is $n \times 1$ (at most, but can be smaller), while d is $2 n \times 1$. The units used are compatible for instance $m$, rad, sec, and $N[21,22]$.

## 4. Linear Quadratic Regulator: LQR Control

It is well known $[23,24]$ that constant input disturbances can be eliminated only if the controller has a zero at infinity (i.e., it integrates). Another useful interpretation is that an integrator is a disturbance estimator. Hence we do not expect a zero steady-state error using an LQR controller.

The structure of LQR control with reduced order observer is shown in Figure 2.

Here, $d$ are the disturbances, $n$ is the noise, and the controller $K$ defines,

$$
\begin{equation*}
K=\lim _{t \rightarrow \infty} K(t), \tag{8}
\end{equation*}
$$



Figure 2: LQR controller with state estimator.
where

$$
\begin{equation*}
u(t)=-K(t) \widehat{x}(t) \tag{9}
\end{equation*}
$$

minimizes the weighted performance index as follows:

$$
\begin{equation*}
J=\int_{0}^{\infty}\left(\hat{x}^{T}(t) Q \widehat{x}(t)+u^{T}(t) R u(t)\right) \mathrm{d} t \tag{10}
\end{equation*}
$$

and $Q$, and $R$ are design weight matrices.
The controller $K$ is given by relation

$$
\begin{equation*}
K A-K B R^{-1} B^{T} K+Q+A^{T} K=0 \tag{11}
\end{equation*}
$$

which is the solution of the Riccati equation.
The weight matrices $Q$ and $R$ are used in order to:
(i) normalize the state and control vector;
(ii) assess the relative influence of deflection from equilibrium position and magnitude of control on the determination of a global criterion. Matrices $Q$ and $R$ are diagonal with positive diagonal inputs, so that

$$
\begin{align*}
& \sqrt{Q_{i}}=\frac{1}{\max \left(x_{i}\right)}, \quad i=1,2, \ldots, m \\
& \sqrt{R_{i}}=\frac{1}{\max \left(u_{i}\right)}, \quad i=1,2, \ldots, k \tag{12}
\end{align*}
$$

The value $\max \left(x_{i}\right)$ sets the maximum desirable output value $y$. The value $\max \left(u_{i}\right)$ has similar significance for input $u$.

Matrix $Q$ sets the weight for each state, while matrix $R$ holds the weight for each actuator's voltage. The LQR problem requires that the state be known [23].

## 5. Inputs

A typical wind load (Figure 3) Acting on the side of the structure. The wind load is a real-life wind speed measurement in relevance with time that took place in Estavromenos of Heraklion, Crete. We transform the wind speed in wind pressure.


Figure 3: Wind load.

Loading corresponds to the wind excitation. The function $f_{m}(t)$ has been obtained from the wind velocity record, through the relation

$$
\begin{equation*}
f_{m}(t)=\frac{1}{2} \rho C_{u} V^{2}(t) \tag{13}
\end{equation*}
$$

where $V=$ velocity, $\rho=$ density, and $C_{u}=1.5$.
Moreover, in all simulations, random noise has been introduced to measurements at system output locations within a probability interval of $\pm 1 \%$. Due to small displacements of system nodal points, noise amplitude is taken to be small, of the order of $5 \times 10^{-5}$. On the other hand, the signal is introduced at each node of the beam by a different percentage, that percentage being lower at the first node due to the fact that the beam end point is clamped.

## 6. Results of Application of LQR Control

The $Q$ and $R$ that were used are

$$
\begin{gather*}
R=0,0001 \times I_{4 \times 4},  \tag{14}\\
Q=100000 \times\left[\begin{array}{ccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & \vdots & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \vdots & \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & \vdots & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \vdots & 0_{7 \times 9} \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \vdots & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \vdots & \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \vdots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & \cdots \\
& & & 0_{9 \times 7} & & & & \vdots & 0_{9 \times 9}
\end{array}\right] . \tag{15}
\end{gather*}
$$

Since $\max \left(x_{i}\right)=0.00316228$ and $\max \left(u_{i}\right)=100$ (11), matrix $L$ is the design matrix. Its eigenvalues are chosen in such
a way that the observer subsystem can be about two times faster than the observed system. The selected values for our simulation are

$$
\lambda_{L}=10^{7} \times\left[\begin{array}{c}
2.7423556  \tag{16}\\
-0.430498 \\
-0.031873 \\
-0.000051+0.0001993 i \\
-0.000051-0.000199 i \\
-0.00045+0.000053 i \\
-0.00045-0.000053 i \\
-0.00039+0.00001 i \\
-0.00039-0.00001 i \\
-0.0004 \\
-0.0004 \\
-0.0004
\end{array}\right] .
$$

These values have been obtained by trial and error, given the poor numerical properties of the system. To find these values, we have used a robust pole computation algorithm included in MATLAB [25, 26].

The controller [Klqr] is given by relation (11) which is the solution of the Riccati equation, where $A$ and $B$ are respective state and control matrices of the system and $R$ and $Q$ are weight matrices of the performance criterion (regulator) (14) and (15), respectively.

For the simulation, beam nodal displacements and rotations with and without control are displayed in Figures 4 and 5, while Figure 6 presents actuator voltage values for control of all beam nodes.

### 6.1. Discussion of the Results of the Linear Quadratic Regulator

 ( $L Q R$ ). Using the linear quadratic controller criterion LQR, beam vibration reduction is observed at all nodal points, for both constant and sinusoidal mechanical input, as well as for realistic wind loading. LQR control achieves reduction of vibration but at the same time requires the entire system state time history as well as an extensive sensor distribution.We encounter the following difficulties:
(i) system disturbances are unknown and unpredictable;
(ii) the state vector is not measurable in its entirety, which in turn necessitates the use of an observer. This setup is problematic, as the observer has no information on the disturbance, which results in erroneous estimates. A way to circumvent this problem is the use of an unknown input observer. Unfortunately, this approach is not feasible, as one of the prerequisite conditions is not met. This situation complicates the problem, making the application of classic controllers such as LQR difficult, since its performance is directly related to the availability of the state vector, or at best of a reliable estimator of the state vector.

For the reasons mentioned above, we will continue with a discussion of more advanced control techniques for this particular problem, such as the $H_{\infty}$ control.


Figure 4: Displacement at all beam nodal points, with and without LQR control.

## 7. $H_{\text {Infinity }}$ Control

To relate the structures used in classical and $H_{\infty}$ control, let us look at Figure 7, in the frequency domain [16, 23, 24].

In this diagram, all inputs and outputs of interest are included, along with their respective weighs $W$, where $W_{d}$, $W_{u}, W_{n}$, and $W_{y}$ are the weighs for the disturbances, control, noise, and outputs, respectively. The exogenous inputs are the noise $n$ and the disturbances $d . K(s)$ is the controller, $B, G, x$, $y, C$ define at the relation $(6,7,8)$, and $F(s)$ is the transfer function of our system.

To find the necessary transfer functions consider the following:

$$
\begin{aligned}
y_{F w} & =W_{y} J x=W_{y} J F v \\
& =W_{y} J F\left(G W_{d} d+B u_{K}\right) \\
& =W_{y} J F G W_{d} d+W_{y} J F B u_{K}, \\
u_{w} & =W_{u} u_{K},
\end{aligned}
$$

$$
\begin{align*}
y_{n} & =C x+W_{n} n \\
& =C F v+W_{n} n \\
& =C F\left(G W_{d} d+B u_{K}\right)+W_{n} n \\
& =C F G W_{d} d+C F B u_{K}+W_{n} n . \tag{17}
\end{align*}
$$

Combining all these gives

$$
\left[\begin{array}{c}
u_{w}  \tag{18}\\
y_{F w} \\
\hline y_{n}
\end{array}\right]=\left[\begin{array}{cc|c}
0 & 0 & W_{u} \\
W_{y} J F G W_{d} & 0 & W_{y} J F B \\
\hline C F G W_{d} & W_{n} & C F B
\end{array}\right]\left[\begin{array}{c}
d \\
n \\
\hline u_{K}
\end{array}\right] .
$$

Note that the plant transfer function matrix, $F(s)$, is deduced from the suitably reformulated plant equations as follows:

$$
\begin{gather*}
\dot{x}(t)=A x(t)+I v(t), \\
y(t)=I x(t), \tag{19}
\end{gather*}
$$

where $v(t)=G d+B u_{k}$.


Figure 5: Angle of rotation at all beam nodal points, with and without LQR control.


Figure 6: Control produced voltage at all beam nodal points with LQR control.

Hence,

$$
\begin{equation*}
F(s)=(s I-A)^{-1} . \tag{20}
\end{equation*}
$$



Figure 7: $H_{\text {Infinity }}$ control bloc diagram in the frequency domain.

The equivalent two-port diagram in the state space form is shown in Figure 8 for the close loop, and with more details in Figure 9, with

$$
z=\left[\begin{array}{c}
u_{w}  \tag{21}\\
y_{F w}
\end{array}\right], \quad w=\left[\begin{array}{l}
d \\
n
\end{array}\right], \quad y=y_{n}, \quad u=u_{K},
$$

where $z$ are the output variables to be controlled and $w$ the exogenous inputs.


Figure 8: Two-port diagram.


Figure 9: Details of $H_{\infty}$ structure.

Given that $P$ has two inputs and two outputs, it is, as usual, naturally partitioned as

$$
\left[\begin{array}{l}
z(s)  \tag{22}\\
y(s)
\end{array}\right]=\left[\begin{array}{ll}
P_{z w}(s) & P_{z u}(s) \\
P_{y w}(s) & P_{y u}(s)
\end{array}\right]\left[\begin{array}{l}
w(s) \\
u(s)
\end{array}\right] \stackrel{\circ \rho}{=} P(s)\left[\begin{array}{l}
w(s) \\
u(s)
\end{array}\right] .
$$

Also,

$$
\begin{equation*}
u(s)=K(s) y(s) \tag{23}
\end{equation*}
$$

Using (18) the transfer function for $P$ which is

$$
P(s)=\left[\begin{array}{cc|c}
0 & 0 & W_{u}  \tag{24}\\
W_{y} J F G W_{d} & 0 & W_{y} J F B \\
\hline C F G W_{d} & W_{n} & C F B
\end{array}\right],
$$

while the closed loop transfer function for $M_{z w}(s)$ is

$$
\begin{equation*}
M_{z w}(s)=P_{z w}(s)+P_{z u}(s) K(s)\left(I-P_{y u}(s) K(s)\right)^{-1} P_{y w}(s), \tag{25}
\end{equation*}
$$

or

$$
\begin{equation*}
z=M_{z w} w=F_{l}(P, K) w \tag{26}
\end{equation*}
$$

Equation (25) is the well known lower LFT for $M_{z w}$.
To express $P$ in state space form, the natural partitioning

$$
P(s)=\left[\begin{array}{c|cc}
A & B_{1} & B_{2}  \tag{27}\\
\hline C_{1} & D_{11} & D_{12} \\
C_{2} & D_{21} & D_{22}
\end{array}\right]=\left[\begin{array}{cc}
P_{z w}(s) & P_{z u}(s) \\
P_{y w}(s) & P_{y u}(s)
\end{array}\right]
$$

is used (where the packed form has been used), while the corresponding form for the controller $K$ is [27-29]

$$
K(s)=\left[\begin{array}{c|c}
A_{K} & B_{K}  \tag{28}\\
\hline C_{K} & D_{K}
\end{array}\right] .
$$

Equation (27) defines the following equations:

$$
\begin{gather*}
\dot{x}(t)=A x(t)+\left[\begin{array}{ll}
B_{1} & B_{2}
\end{array}\right]\left[\begin{array}{l}
w(t) \\
u(t)
\end{array}\right], \\
{\left[\begin{array}{l}
z(t) \\
y(t)
\end{array}\right]=\left[\begin{array}{l}
C_{1} \\
C_{2}
\end{array}\right] x(t)+\left[\begin{array}{ll}
D_{11} & D_{12} \\
D_{21} & D_{22}
\end{array}\right]\left[\begin{array}{l}
w(t) \\
u(t)
\end{array}\right],}  \tag{29}\\
\dot{x}_{K}(t)=A_{K} x_{K}(t)+B_{K} y(t), \\
u(t)=C_{K} x_{K}(t)+D_{K} y(t) .
\end{gather*}
$$

To find the matrices involved, we break the feedback loop and use the relevant equations.

Therefore the equations relating the inputs, outputs, states, and input/output to the controller are

$$
\begin{gather*}
\dot{x}_{F}=A x_{F}+\left(G d_{w}+B u\right), \quad y_{F}=x_{F}, \\
\dot{x}_{u}=A_{u} x_{u}+B_{u} u, \quad u_{w}=C_{u} x_{u}+D_{u} u, \\
\dot{x}_{y F}=A_{y F} x_{y F}+B_{y F} J y_{F}, \quad y_{F w}=C_{y F} x_{y F}+D_{y F} y_{F}, \\
\dot{x}_{n}=A_{n} x_{n}+B_{n} n, \quad n_{w}=C_{n} x_{n}+D_{n} n, \\
\dot{x}_{d}=A_{d} x_{d}+G d, \quad d_{w}=C_{d} x_{d}+D_{d} d, \\
y_{n}=C y_{F}+n_{w}, \\
x=\left[\begin{array}{c}
x_{F} \\
x_{u} \\
y_{F w} \\
x_{n} \\
x_{d}
\end{array}\right], \quad y=y_{n}, \\
z=\left[\begin{array}{c}
u_{w} \\
y_{F w}
\end{array}\right], \tag{30}
\end{gather*}
$$

From (30), we use $d_{w}, n_{w} \kappa \alpha \iota y_{F w}$ and take our initial state space equation in the form of $(6,7,8)$, as follows:

$$
\begin{align*}
\dot{x}= & {\left[\begin{array}{ccccc}
A_{G} & 0 & 0 & 0 & G C_{d} \\
0 & A_{u} & 0 & 0 & 0 \\
B C_{F} & 0 & A_{y F} & 0 & 0 \\
0 & 0 & 0 & A_{n} & 0 \\
0 & 0 & 0 & 0 & A_{d}
\end{array}\right] x } \\
& +\left[\begin{array}{cc}
G D_{d} & 0 \\
0 & 0 \\
0 & 0 \\
0 & B_{n} \\
B_{d} & 0
\end{array}\right] w+\left[\begin{array}{c}
B \\
B_{u} \\
0 \\
0 \\
0
\end{array}\right] u, \tag{31}
\end{align*}
$$

$$
\begin{gathered}
z=\left[\begin{array}{ccccc}
0 & C_{u} & 0 & 0 & 0 \\
D_{y F} C_{F} & 0 & C_{y F} & 0 & 0
\end{array}\right] x+0 w+\left[\begin{array}{c}
D_{u} \\
0
\end{array}\right] u, \\
y=\left[\begin{array}{lllll}
C_{F} & 0 & 0 & C_{n} & 0
\end{array}\right] x+\left[\begin{array}{ll}
0 & D_{n}
\end{array}\right] w+0 u .
\end{gathered}
$$



Figure 10: Bode diagrams of diagonal elements of weight matrices.

Therefore the matrices are

$$
\begin{gather*}
A_{1}=\left[\begin{array}{ccccc}
A_{F} & 0 & 0 & 0 & G C_{d} \\
0 & A_{u} & 0 & 0 & 0 \\
B C_{F} & 0 & A_{y F} & 0 & 0 \\
0 & 0 & 0 & A_{n} & 0 \\
0 & 0 & 0 & 0 & A_{d}
\end{array}\right], \\
B_{1}=\left[\begin{array}{cc}
G D_{d} & 0 \\
0 & 0 \\
0 & 0 \\
0 & B_{n} \\
B_{d} & 0
\end{array}\right], \quad B_{2}=\left[\begin{array}{c}
B \\
B_{u} \\
0 \\
0 \\
0
\end{array}\right], \\
C_{1}=\left[\begin{array}{cccc}
0 & C_{u} & 0 & 0 \\
D_{y F} C_{F} & 0 & C_{y F} & 0 \\
0
\end{array}\right], \\
D_{11}=0, \\
C_{2}=\left[\begin{array}{lll}
C_{F} & 0 & D_{12}=\left[\begin{array}{c}
D_{u} \\
0
\end{array}\right], \\
C_{n} & 0
\end{array}\right], \quad D_{21}=\left[\begin{array}{ll}
0 & D_{n}
\end{array}\right], \quad D_{22}=0 . \tag{32}
\end{gather*}
$$

7.1. Results with $H_{\infty}$. Figure 10 presents the Bode diagrams of diagonal elements of the above weight matrices. These
matrices have been obtained through a number of tests, to ensure the feasibility of finding a controller $H_{\infty}$.

The controller obtained by applying $H_{\infty}$ control is 36 order. For this controller, $\gamma=0.074$. A plot of the maximum singular value of the weighted closed loop system (beam plus $H_{\infty}$ controller) is given in Figure 11, where we can clearly note that the value remains below $\gamma$ at all frequencies.

Figures 12, 13, and 14 further show the maximum singular values of transfer functions of the unweighted closed loop system (i.e., the initial one) that are of interest.

These figures show that the performance of the computed controller is satisfactory [30] since:
(i) as shown in Figure 12, there is a significant improvement in the effect of disturbance on error up to the frequency of 1000 Hz ;
(ii) as shown in Figure 13, there seems to be little effect of noise on error for frequencies beyond 1000 Hz ;
(iii) Figure 14 shows a satisfactory effect of the disturbance on the size of the control scheme (the design could be improved, if it were possible to reduce noise effect for frequencies of 1000 Hz ).


Figure 11: Maximum singular value of the unweighted closed loop system.


Figure 12: Maximum singular value disturbance to error.

To validate the above findings, system response time histories for the three input cases mentioned in this section are presented below.

Using the mechanical input, we get the following result.
Figure 15 shows the displacement time history at all nodal points of the beam, with and without control, while Figure 16 displays the angle of rotation time history at all beam nodal points, with and without control. By employing the $H_{\infty}$ control, vibration reduction is achieved; we observe vibration reduction of $90 \%$. Figure 17 presents the time evolution of the produced actuator voltage, which turns out to be lower than the piezoelectric voltage limit value of 500 V .
7.2. Order Reduction of Controller $H_{\infty}$. The $H_{\infty}$ controller found is of order 36. The fact that controller order, which is equal to the order of the system, is relatively higher than the order of classical controllers such as PI and LQR has led a number of researchers to develop order reduction algorithms.


Figure 13: Maximum singular value noise to error.


Figure 14: Maximum singular value plots for control.

The most widely used such algorithm, known as HIFOO, has been implemented in a Matlab environment, and is the one used in the following procedure [31].

The general problem is to compute a controller of reduced rank/order $n<36$ while retaining the performance of the $H_{\infty}$ criterion as well as the behaviour of a full order controller for the given system $[32,33]$ as follows:

$$
\begin{gather*}
\dot{x}(t)=A x(t)+B_{1} w(t)+B_{2} u(t), \\
z(t)=C_{1} x(t)+D_{11} w(t)+D_{12} u(t),  \tag{33}\\
y(t)=C_{2} x(t)+D_{21} w(t)+D_{22} u(t) .
\end{gather*}
$$

The state space equations for the controller $K$ are

$$
\begin{gather*}
\dot{x}_{K}(t)=A_{K}(t)+B_{K} y(t), \\
u(t)=C_{K}(t)+D_{K} y(t) . \tag{34}
\end{gather*}
$$



Figure 15: Displacement at all beam nodal points, with and without $H_{\infty}$ control.

Let $\alpha(X)$ be the spectral abscissa of a matrix $X$, that is, the maximum real part of its eigenvalues. Then, we require not only that $\alpha\left(A_{\mathrm{CL}}\right)<0$, where $A_{\mathrm{CL}}$ is the closed loop system matrix, but that $\alpha\left(A_{k}\right)<0$ as well. The feasible set of $A_{k}$, that is the set of stable matrices, is not a convex set and has a boundary that is not smooth $[34,35]$.

The HIFOO procedure has two phases: stability and performance optimization [31,36]. In the stability phase, HIFOO attempts to minimize

$$
\begin{equation*}
\max \left(\alpha\left(A_{\mathrm{CL}}, \in \alpha\left(A_{\mathrm{CL}}\right)\right)\right) \tag{35}
\end{equation*}
$$

where $\varepsilon$ is a positive parameter that will be described shortly, until a controller is found for which this quantity is negative; that is, the controller is stable and makes the closed loop system stable. In case it is unable to find such a controller, HIFOO terminates unsuccessfully.

In the performance optimization phase, HIFOO searches for a local minimizer of
$f(K)$

$$
= \begin{cases}\infty, & \text { if } \max \left(\alpha\left(A_{\mathrm{CL}}, \alpha\left(A_{K}\right)\right)\right) \geq 0  \tag{36}\\ \max \left(\left\|T_{z w}\right\|_{\infty}, \in\|K\|_{\infty}\right), & \text { if else }\end{cases}
$$

where

$$
\begin{equation*}
\|K\|_{\infty}=\sup _{R s=0}\left\|C_{k}\left(s I-A_{k}\right)^{-1} B_{K}+D_{K}\right\|_{2} . \tag{37}
\end{equation*}
$$

The introduction of $\varepsilon$ is motivated by the fact that the main design objective is to attain closed loop system stability and to minimize $\left\|T_{z w}\right\|_{\infty}$, by demonstrating that $\varepsilon$ should be relatively small; the term $\varepsilon\|K\|_{\infty}$, however, prevents the controller $H_{\infty}$ norm from becoming too large, in which case the stability constraint by itself would not exist. Given


Figure 16: Angle of rotation at all beam nodal points, with and without $H_{\infty}$ control.


Figure 17: Control produced voltage at all beam nodal points with $H_{\infty}$ control.
that it is preceded by the stability phase, the performance optimization phase is initialized with a finite value of $f(K)$. Consequently, when it reaches a value of $K$ for which $f(K)=$ $\infty$, that value is rejected, since an objective reduction is sought at each iteration [31, 36].
7.3. Results Using Controller HIFOO. As mentioned before, the HIFOO controller is implemented in Matlab by way of appropriate routines. It is called in the following manner:

$$
\begin{equation*}
\text { Kfoo }=\text { hifoo }(\text { plant, } 2), \tag{38}
\end{equation*}
$$

where plant is the system description in the form of (33) and $n=2$ is the controller order.

The resulting controller is described in state space in similar manner as $H_{\infty}$; that is,

$$
\begin{gather*}
\dot{x}_{K}(t)=A_{K}(t)+B_{K} y(t),  \tag{39}\\
u(t)=C_{K}(t)+D_{K} y(t) .
\end{gather*}
$$

The controller state space equation is given by (39), where controller matrices are equal to

$$
\begin{gather*}
A_{K}=\left[\begin{array}{cc}
728.1 & -5034 \\
207.5 & -1408
\end{array}\right], \\
B_{K}=\left[\begin{array}{cccc}
212.8 & 811.6 & 1716 & 2810 \\
-164.9 & -637.2 & -1348 & -2207
\end{array}\right], \\
C_{K}=\left[\begin{array}{ccc}
1557 & -916.7 \\
1013 & -592.3 \\
517 & -297.9 \\
144.3 & -82.59
\end{array}\right],  \tag{40}\\
D_{K}=\left[\begin{array}{cccc}
36.1 & 136.6 & 287.1 & 468.3 \\
23.5 & 87.69 & 186.5 & 303 \\
12.12 & 44.12 & 93.39 & 154.3 \\
4.204 & 12.53 & 26.92 & 43.51
\end{array}\right]
\end{gather*}
$$

For the purpose of comparison of HIFOO controller performance to that of $H_{\infty}$, the beam free end response is examined, for the mechanical input.

For the input in Figure 18, the beam free end response is shown, initially with and then without the HIFOO controller, while Figure 19 presents produced actuator voltage using the HIFOO controller.

Using the HIFOO controller for an actual wind loading, beam position control is effected with node displacements of order of $10^{-5}$, with lower produced voltage. We therefore maintain $H_{\infty}$ criterion performance with a lower order controller. The maximum produced voltage for the HIFOO controller is 7 V ; the respective value is 45 V for the $H_{\infty}$ controller. In other words, beam adjustment to its equilibrium position is achieved with a lower order controller that requires lower voltage; see Figure 19.

## 8. Results

In the present work, the use of active control technology in intelligent structures has been presented. The goal of control

(s)

Figure 18: Beam free end displacement, with and without HIFOO control.


Figure 19: Stress at beam nodal points, using HIFOO.
is vibration reduction, while sustaining low steady state error, short recovery time, and small maximum uplift; at the same time, control energy must remain within operating limits.

The beam that was used was discretized using 1dimensional finite elements with two degrees of freedom per node. Piezoelectric actuators were embedded in it with the objective of reducing vibrations under deterministic as well as stochastic loading conditions.

Initially, we examined the linear quadratic control criterion using a reduced rank observer, which makes the simulation more realistic. To find the observer, we employed a robust pole location algorithm. By selecting appropriate weights, beam vibration reduction was achieved for stochastic loading cases. In all simulations, random noise has been introduced in measurements, so that the system better approximate reality, given that displacement measurement by means of
piezoelectric sensors is not reliable. Next we applied more advanced control techniques, such as the $H_{\infty}$ criterion. The $H_{\infty}$ controller found is of order 36 .

In order to reduce computational requirements of the model, controller rank was reduced by means of nonparametric and nonconvex optimization, using the HIFOO controller. The controller exhibited good performance even for a significantly smaller system degree.

A natural consequence of the proposed research innovations is the acknowledgement of new scientific problems that can be used as the basis for further research beyond the scope of this work. The advantage of active control is the fact that it allows taking into account in the computation the worst case result of disturbances with uncertainty and system noise. Moreover, the active control can effectively cope with stronger input, permitting the design for a large frequency bandwidth. Results are noteworthy; vibration reduction is observed even for realistic wind loading, with piezoelectric component voltage kept within tolerance.

## Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to thank the Wind Energy and Synthesis of Energy Systems Laboratory, of the Technological Educational Institute of Crete, for its offer of real life wind speed measurements in relevance with time that took place in Estavromenos of Heraclion, Crete.

## References

[1] B. Bandyopadhyay, T. C. Manjunath, and M. Umapathy, Modeling, Control and Implementation of Smart Structures, vol. 350 of Lecture Notes in Control and Information Sciences, Springer, 2007.
[2] D. Halim, Vibration analysis and control of smart structures [Ph.D. thesis], 2002.
[3] B. Culshaw, "Smart structures-a concept or a reality?" Journal of Systems and Control Engineering, vol. 206, no. 1, pp. 1-8, 1992.
[4] S. S. Rao and M. Sunar, "Piezoelectricity and its use in disturbance sensing and control of flexible structures: a survey," Applied Mechanics Reviews, vol. 47, no. 4, pp. 113-123, 1994.
[5] T. Bailey and J. E. Ubbard Jr., "Distributed piezoelectricpolymer active vibration control of a cantilever beam," Journal of Guidance, Control, and Dynamics, vol. 8, no. 5, pp. 605-611, 1985.
[6] S. Hanagud, M. W. Obal, and A. J. Calise, "Optimal vibration control by the use of piezoceramic sensors and actuators," Journal of Guidance, Control, and Dynamics, vol. 15, no. 5, pp. 1199-1206, 1992.
[7] W.-S. Hwang and H. C. Park, "Finite element modeling of piezoelectric sensors and actuators," AIAA Journal, vol. 31, no. 5, pp. 930-937, 1993.
[8] B. Bona, M. Indri, and A. Tornambé, "Flexible piezoelectric structures-approximate motion equations and control algorithms," IEEE Transactions on Automatic Control, vol. 42, no. 1, pp. 94-101, 1997.
[9] W. Schiehlen and H. Schonerstedt, "Controller design for the active vibration damping of beam structure," in Proceedings of the Smart Mechanical Systems Adaptronics SAE International USA, pp. 1137-1146, 1998.
[10] S.-B. Choi, C.-C. Cheong, and C.-H. Lee, "Position tracking control of a smart flexible structure featuring a piezofilm actuator," Journal of Guidance, Control, and Dynamics, vol. 19, no. 6, pp. 1364-1369, 1996.
[11] F. Pourki, "Distributed controllers for flexible structures using piezoelectric actuators/sensors", in Proceedings of the 32nd IEEE Conference on Decision and Control (CDC '93), vol. 2, pp. 13671370, San Antonio, Tex, USA, December 1993.
[12] A. Benjeddou, M. A. Trindade, and R. Ohayon, "New shear actuated smart structure beam finite element," AIAA Journal, vol. 37, no. 3, pp. 378-383, 1999.
[13] X. D. Zhang and C. T. Sun, "Formulation of an adaptive sandwich beam," Smart Materials and Structures, vol. 5, no. 6, pp. 814-823, 1996.
[14] S. Raja, G. Prathap, and P. K. Sinha, "Active vibration control of composite sandwich beams with piezoelectric extensionbending and shear actuators," Smart Materials and Structures, vol. 11, no. 1, pp. 63-71, 2002.
[15] G. Zames and B. A. Francis, "Feedback, minimax sensitivity, and optimal robustness," IEEE Transactions on Automatic Control, vol. 28, no. 5, pp. 585-601, 1983.
[16] D. Halim and S. O. R. Moheimani, "Spatial resonant control of flexible structures-application to a piezoelectric laminate beam," IEEE Transactions on Control Systems Technology, vol. 9, no. 1, pp. 37-53, 2001.
[17] Z. Friedman and J. B. Kosmatka, "An improved two-node Timoshenko beam finite element," Computers and Structures, vol. 47, no. 3, pp. 473-481, 1993.
[18] G. E. Stavroulakis, G. Foutsitzi, E. Hadjigeorgiou, D. Marinova, and C. C. Baniotopoulos, "Design and robust optimal control of smart beams with application on vibrations suppression," Advances in Engineering Software, vol. 36, no. 11-12, pp. 806-813, 2005.
[19] B. Miara, G. Stavroulakis, and V. Valente, Topics on Mathematics for Smart Systems: Proceedings of the European Conference, Rome, Italy, 26-28 October 2006, World Scientific, Singapore, 2007.
[20] A. Moutsopoulou, A. Pouliezos, and G. E. Stavroulakis, "Modelling with uncertainty and robust control of smart beams," in Proceedings of the 9th International Conference on Computational Structures Technology (CST '09), B. H. V. Topping and M. Papadrakakis, Eds., paper 35, Civil Comp Press, Athens, Greece, September 2008.
[21] H. F. Tiersten, Linear Piezoelectric Plate Vibrations, Plenum Press, New York, NY, USA, 1st edition, 1969.
[22] S. M. Yang and Y. J. Lee, "Optimization of noncollocated sensor/actuator location and feedback gain in control systems," Smart Materials and Structures, vol. 2, no. 2, pp. 96-102, 1993.
[23] O. H. Bosgra and H. Kwakernaak, Design Methods for Control Systems, Course Notes, Dutch Institute for Systems and Control, 2001.
[24] H. Kwakernaak, "Robust control and $H_{\infty}$-optimizationtutorial paper," Automatica, vol. 29, no. 2, pp. 255-273, 1993.
[25] R. C. Ward, "Balancing the generalized eigenvalue problem," SIAM Journal on Scientific and Statistical Computing, vol. 2, no. 2, pp. 141-152, 1981.
[26] A. L. Tits and Y. Yang, "Globally convergent algorithms for robust pole assignment by state feedback," IEEE Transactions on Automatic Control, vol. 41, no. 10, pp. 1432-1452, 1996.
[27] H. Kimura, "Robust stabilizability for a class of transfer functions," IEEE Transactions on Automatic Control, vol. 29, no. 9, pp. 788-793, 1984.
[28] B. A. Francis, A Course on $H_{\infty}$ Control Theory, Lecture Notes in Control and Information Sciences, Springer, 1987.
[29] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis, "State-space solutions to standard $H_{2}$ and $H_{\infty}$ control problems," IEEE Transactions on Automatic Control, vol. 34, no. 8, pp. 831-847, 1989.
[30] A. Packard, J. Doyle, and G. Balas, "Linear, multivariable robust control with a $\mu$ perspective," Journal of Dynamic Systems, Measurement and Control, vol. 115, no. 2, pp. 426-438, 1993.
[31] J. V. Burke, D. Henrion, A. S. Lewis, and M. L. Overton, "HIFOO-a MATLAB package for fixed-order controller design and $H_{\infty}$ optimization," in Proceedings of the 5th IFAC Symposium on Robust Control Design (ROCOND '06), pp. 339344, Toulouse, France, July 2006.
[32] J. V. Burke, D. Henrion, A. S. Lewis, and M. L. Overton, "Stabilization via nonsmooth, nonconvex optimization," IEEE Transactions on Automatic Control, vol. 51, no. 11, pp. 1760-1769, 2006.
[33] J. V. Burke, A. S. Lewis, and M. L. Overton, "A robust gradient sampling algorithm for nonsmooth, nonconvex optimization," SIAM Journal on Optimization, vol. 15, no. 3, pp. 751-779, 2005.
[34] J. V. Burke and M. L. Overton, "Variational analysis of nonLipschitz spectral functions," Mathematical Programming B, vol. 90, no. 2, pp. 317-351, 2001.
[35] D. Hinrichsen and A. J. Pritchard, Mathematical Systems Theory I: Modelling, State Space Analysis, Stability and Robustness, vol. 48 of Texts in Applied Mathematics, Springer, 2006.
[36] M. Millston, HIFOO 1.5: structured control of linear systems with a non trivial feedthrough [M.S. thesis], Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, 2006.

## Research Article

# Nonlinear Fault Separation for Redundancy Process Variables Based on FNN in MKFDA Subspace 

Ying-ying Su, ${ }^{1,2}$ Shan Liang, ${ }^{1}$ Jing-zhe Li, ${ }^{3}$ Xiao-gang Deng, ${ }^{4}$ Tai-fu Li, ${ }^{2}$ and Cheng Zeng ${ }^{1}$<br>${ }^{1}$ College of Automation, Chongqing University, Chongqing 400044, China<br>${ }^{2}$ School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing 401331, China<br>${ }^{3}$ School of Safety Engineering, Chongqing University of Science and Technology, Chongqing 401331, China<br>${ }^{4}$ College of Mechanical and Power Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

Correspondence should be addressed to Shan Liang; lightsun@cqu.edu.cn
Received 19 July 2013; Revised 22 October 2013; Accepted 23 October 2013; Published 4 February 2014
Academic Editor: Xianxia Zhang
Copyright © 2014 Ying-ying Su et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Nonlinear faults are difficultly separated for amounts of redundancy process variables in process industry. This paper introduces an improved kernel fisher distinguish analysis method (KFDA). All the original process variables with faults are firstly optimally classified in multi-KFDA (MKFDA) subspace to obtain fisher criterion values. Multikernel is used to consider different distributions for variables. Then each variable is eliminated once from original sets, and new projection is computed with the same MKFDA direction. From this, differences between new Fisher criterion values and the original ones are tested. If it changed obviously, the effect of eliminated variable should be much important on faults called false nearest neighbors (FNN). The same test is applied to the remaining variables in turn. Two nonlinear faults crossed in Tennessee Eastman process are separated with lower observation variables for further study. Results show that the method in the paper can eliminate redundant and irrelevant nonlinear process variables as well as enhancing the accuracy of classification.


## 1. Introduction

With developments of modern process industry, multivariate monitor from sensors has showed their multicollinearity, nonlinear correlative coupling, time delay, and redundancy. It makes complexity increasing with exponent to fault separation and diagnosis, called "Curse of Dimension" [1, 2]. On the other hand, right ratio of fault classification decreases with multivariate and redundancy process variables. Therefore, many attentions have been paid on two points of view that are variable selection and dimension reduction [3, 4].

Among the study of variable selection, the existed methods can be broadly classified into three categories: random search techniques, measure-based method, and intelligent computation. In random search, each process variable is directly deleted or involved in the classification model one time in turn to search the most suitable input sets under a certain criterion, such as forward selection, backward selection, and stepwise that are simple and easily realized methods [5].

While it was studied by Masion and Gunst [6] that these methods would result in mistaken results, variable set appears multicollinearity. Measure-based method appears to select variable with computing relevancy among all variables, as well as that between variables and labels. The variables with highest similar characteristic will be gathered in one kind. According to different definition, K-L information measure, minimum description length, and mutual information are used [7-9]. Intelligent computation deepens to solve nonlinear variable selection problem, such as neurnal network that is once used to nonlinear model, while its selection criterion is uncertain [10].

Dimension reduction is different from variable selection, which mainly depends on transformation and information extraction of original variable matrix. It projects original variables with a certain mapping to a new subspace and extracts information in lower dimension, such as principal component analysis (PCA) [11] and partial least squares (PLS) [12].


Figure 1: The fault diagnosis with multivariate.

Original variables with linear-relative process variables are linearly projected according to the maximum direction of covariance matrix. Meanwhile, the maximum original information can be kept as most as possible. Contribution chart method is the way to calculate contribution of each variable to certain fault with $T^{2}$ statics and SPE $[13,14]$ for PCA. The above linear methods have been extended to nonlinear ones after kernel method presented [15-20], such as kernel principal component analysis (KPCA), kernel partial least squares (KPLS), and kernel fisher discriminant analysis (KFDA). Kernel method converts a linear classification learning algorithm into nonlinear one, by mapping the original observations into a higher-dimensional space. So that linear classifier in the new space equals to a nonlinear classifier in the original space.

However, nonlinear information projected to the new feature space has higher dimension, and data matrix has lost their original physical meaning in original sample space. If we separated nonlinear faults crossed together in original space, the dimension of classifier with kernel method would become huge, while right ratio would decrease with redundancy and multicollinearity variables.

The objective of this paper is to deepen dimension reduction method for the above problems with measure method in variable selection called MKFDA-FNN. Nonlinear process variables are projected in higher-dimension space with MKFDA. Discriminant vector and its corresponding feature vector with maximum separation are computed to cluster original variables with highest similarity. With embeddimension increasing, false nearest neighbors (FNN) with high similarity are able to be removed in turn. Thus, nonlinear redundancy and multicollinearity process variables can be removed from input sets to nonlinear classifier. Finally, we give an actual fault separation problem in classical chemical process Tennessee Eastman (TE) to further study.

## 2. Problem Description

In fault separation problem presented above, it equals to screen original process variables related to certain faults as most as possible. Multivariate data matrix considered initially with normal and fault information is described in Figure 1, where $X_{1}, X_{2}, \ldots, X_{n}$ are process variables with $n$-dimension,


Figure 2: Nonlinear fault diagnosis with redundancy process variables based on FNN in MKFDA subspace.
$Z_{11}^{-1}, Z_{12}^{-1}, \ldots, Z_{1 f}^{-1}$ present time-delay variables of $X_{1}$ at different sample time, $Z_{21}^{-1}, Z_{22}^{-1}, \ldots, Z_{2 g}^{-1}$ present time-delay variables of $X_{2}$, and $Z_{n 1}^{-1}, Z_{n 2}^{-1}, \ldots, Z_{n h}^{-1}$ present time-delay variables of $X_{n}$ at different sample-time. In this way, original data matrix is composed of $n$-dimension process/control variables and their delay variables in

$$
\begin{align*}
& X_{(f+g+h) \times l}(k) \\
& \quad=\left[\begin{array}{cccc}
x_{1}(k-1) & x_{1}(k-2) & \cdots & x_{1}(k-l) \\
x_{1}(k-2) & x_{1}(k-3) & \cdots & x_{1}(k-l-1) \\
\vdots & \vdots & \vdots & \vdots \\
x_{1}(k-f) & x_{1}(k-f-1) & \cdots & x_{1}(k-f-l+1) \\
x_{2}(k-1) & x_{2}(k-2) & \cdots & x_{2}(k-l) \\
x_{2}(k-2) & x_{2}(k-3) & \cdots & x_{2}(k-l-1) \\
\vdots & \vdots & \vdots & \vdots \\
x_{2}(k-g) & x_{2}(k-g-1) & \cdots & x_{2}(k-g-l+1) \\
\vdots & \vdots & \vdots & \vdots \\
x_{n}(k-1) & x_{n}(k-2) & \cdots & x_{n}(k-l) \\
x_{n}(k-2) & x_{n}(k-3) & \cdots & x_{n}(k-l-1) \\
\vdots & \vdots & \vdots & \vdots \\
x_{n}(k-h) & x_{n}(k-h-1) & \cdots & x_{n}(k-h-l+1)
\end{array}\right], \tag{1}
\end{align*}
$$

where $f, g$, and $h$ present maximum delay order of process/ control variables $x_{1}, x_{2}$, and $x_{n}, k$ presents current sample time, and $l$ is sample length.

## 3. Multivariate Fault Separation Based on MKFDA-FNN

To fault separation problem with nonlinear redundancy process/control variables, an approach is proposed in Figure 2. Correlated nonlinear variables are firstly projected to a higher-dimension MKFDA subspace. Furthermore, in order to find fairly useful variables, the importance of each input is measured in subspace with distance measure inspired by FNN. Accordingly, redundant variables are recognized. It makes separation of faults crossed together easily.
3.1. False Nearest Neighbors. FNN is the feature selection method on the basis of phase space reconstruction (PSR) in high-dimension data space [21]. With embed-dimension
increasing, movement locus becomes open, and false nearest neighbors with high similarity are able to be removed in turn. It restores the locus of chaos. Its algorithm is as follows.

In $d$-dimension phase space including original variables and their time delay, each phase vector $x(i)=\{x(i), x(i+\tau)$, $\ldots, x(i+(d-1) \tau)\}$ has one nearest neighbors $x^{N N}(i)$. Their 2 -norm distance is

$$
\begin{equation*}
R_{d+1}^{2}(i)=R_{d}^{2}(i)+\left\|x(i+\tau d)-x^{N N}(i+\tau d)\right\| \tag{2}
\end{equation*}
$$

When $d$-dimension is increased to $d+1$, the above phase vector is changed as new one, noted as $R_{d+1}(i)$ in

$$
\begin{equation*}
R_{d}(i)=\left\|x(i)-x^{N N}(i)\right\| \tag{3}
\end{equation*}
$$

If $R_{d+1}(i)$ was much bigger than $R_{d}(i)$, it means the projection of two nonneighbor phase vector from higher dimension to lower one. So the two neighbors are the false nearest neighbors.

Note that

$$
\begin{equation*}
a_{1}(i, d)=\frac{\left\|x(i+\tau d)-x^{N N}(i+\tau d)\right\|}{R_{d}(i)} . \tag{4}
\end{equation*}
$$

If $a_{1}(i, d)$ is larger than $R_{\tau}, x^{N N}(i)$ should be fault nearest neighbor of $x(i)$. Threshold $R_{\tau}$ is determined between inter$\operatorname{val}(10,50)$. Once there appeared noise in process data, the following judge criterion should be involved. If $R_{d+1}(i) / R_{A} \geq 2$, $x^{N N}(i)$ should be nearest fault neighbor of $x(i)$, where $R_{A}$ is

$$
\begin{equation*}
R_{A}=\frac{1}{N} \sum_{i=1}^{N}[x(i)-\bar{x}], \quad \bar{x}=\frac{1}{N} \sum_{i=1}^{N} x(i) . \tag{5}
\end{equation*}
$$

The distance measure between vectors can explain the similarity of false nearest neighbors factually in (6). Assume that there was a data space $Q$ with $d$-dimension variable, and one sample vector is $A=\left(q_{1}, q_{2}, \ldots, q_{i}, \ldots, q_{d}\right)$. We set variable $q_{i}$ as zero, standing for vector $A$ without variable $q_{i}$ that is noted as $B=\left(q_{1}, q_{2}, \ldots, 0, \ldots, q_{d}\right)$ in Figure 3.

The similarity between $A$ and $B$ is

$$
\begin{equation*}
\delta=\|A-B\|^{2} . \tag{6}
\end{equation*}
$$

If distance measure is small, it shows that vectors $A$ and $B$ have highly similarity. That is, the removed variable $q_{i}$ makes little impact on nonlinear pattern, and process variable $q_{i}$ has low interpreting ability. Otherwise, if it was much bigger, it reveals that $B$ much differs from $A$. Process variable $q_{i}$ is important to interpreting of nonlinear pattern. $B$ is false nearest neighbors of $A$.
3.2. Kernel Fisher Discriminant Analysis. KFDA is most useful to nonlinear classification problems [22]. Nonlinear discriminant vector in original space is extracted to linear optimal discriminate vector in high-dimension feature space $H$ with conventional fisher discriminant analysis (FDA). Since dimension of $H$ is much higher, it is hard to directly confirm nonlinear mapping function from original space to the feature space. Reproducing kernel-based method widely


Figure 3: Data space $Q$ with $d$-dimension variable.
developed in machine learning (ML) can achieve this goal. Nonlinear mapping is indirectly found according to $k(\mathbf{x}, \mathbf{y})=$ $\Phi(\mathbf{x})^{T} \Phi(\mathbf{y})$ in Gram-space [23], where $\Phi: \mathbf{R}^{d} \rightarrow H$.

Conventional kernel function can be selected as follows [6].
(i) Polynomial kernel function $K\left(\mathbf{x}, \mathbf{x}^{\prime}\right)=\left(\mathbf{x} \cdot \mathbf{x}^{\prime}+c\right)^{d}, d=$ $1,2, \ldots, N, c$ is constant.
(ii) Gaussian kernel function $K\left(\mathbf{x}, \mathbf{x}^{\prime}\right)=\exp (-(\| \mathbf{x}-$ $\left.\left.\mathbf{x}^{\prime} \| / 2 \sigma^{2}\right)\right), \delta$ is the parameter of breadth.
(iii) Sigmoid kernel function $K\left(\mathbf{x}, \mathbf{x}^{\prime}\right)=\tanh \left(f\left\langle\mathbf{x}, \mathbf{x}^{\prime}\right\rangle+\theta\right)$.

Assume that original sample set was $X=\left\{x_{1}, x_{2}, \ldots, x_{N}\right\}$ with $d$-dimension and $N$-samples, where $X_{i}$ is the sample of $i$ th type, $N_{i}=\left|X_{i}\right|$, and $i=1,2$. There exists nonlinear mapping function $\Phi: \mathbf{R}^{d} \rightarrow H$. It transforms nonlinear original sample space $\mathbf{R}^{d}$ to linear classification in high-dimension data space $H$; that is, $\Phi(\mathbf{x}) \in H, x \in \mathbf{R}^{d}$. In space $H$, distance scatter of intraclass and classes with training data is $\mathbf{S}_{\omega}$ and $\mathbf{S}_{b}$ in (7) and (9), respectively,

$$
\begin{gather*}
\mathbf{S}_{\omega}=\mathbf{S}_{1}+\mathbf{S}_{2},  \tag{7}\\
\mathbf{S}_{i}=\sum_{x \in X_{i}}\left(\Phi(\mathbf{x})-\mathbf{m}_{i}\right)\left(\Phi(\mathbf{x})-\mathbf{m}_{i}\right)^{T}, \quad i=1,2,  \tag{8}\\
\mathbf{S}_{b}=\left(\mathbf{m}_{1}-\mathbf{m}_{2}\right)\left(\mathbf{m}_{1}-\mathbf{m}_{2}\right)^{T}, \tag{9}
\end{gather*}
$$

where $\mathbf{m}_{i}$ is the mean of $i$ th type in feature space. KFDA is to find a projection direction $\mathbf{w}$, which meets the following two properties: (1) data that has similar characteristic should be gathered together as most as possible; (2) the ones with different characteristic should be gathered as far as possible. So a key is to search projection direction $\mathbf{w}^{*}$ and its corresponding discriminant function $g(\mathbf{x})=\left(\mathbf{w}^{*}\right)^{T} \mathbf{x}-\mathbf{y}_{0}$. Similarly with linear FDA, the optimal projection direction $\mathbf{w}^{*}$ is to search vector $\mathbf{w}$, which maximizes fisher criterion function (10), where $\mathbf{w}^{*}$ is optimal projection direction:

$$
\begin{equation*}
J_{H}(\mathbf{w})=\frac{\mathbf{w}^{T} \mathbf{S}_{b} \mathbf{w}}{\mathbf{w}^{T} \mathbf{S}_{\omega} \mathbf{w}} \tag{10}
\end{equation*}
$$

Since dimension of feature space $H$ is usually high and $\Phi$ is indirect mapping function, discriminant vector is hard to compute directly. Thus, each solution $\mathbf{w}$ is expressed as linear

Table 1: Steps designed in this paper.

| Inputs | $T=\left\{\left(\mathbf{x}_{1}, y_{1}\right),\left(\mathbf{x}_{2}, y_{2}\right), \ldots,\left(\mathbf{x}_{N}, y_{N}\right)\right\}, \mathbf{x}_{i} \in \mathbf{R}^{d}, y_{i} \in\{+1,-1\}, i=1,2, \ldots, N$ |
| :--- | :--- |
| Step 1 | Initiate $\lambda \in[0,1]$ and compute $\mathbf{u}_{i}, i=1,2$ |
| Step 2 | Select suitable multikernel function |
| Step 3 | Compute the kernel mean vector between two kinds with $\mathbf{k}_{w i}=\sum_{j=1}^{N_{i}}\left(\mathbf{K}_{\mathbf{x}_{j}^{i}}-\mathbf{u}_{i}\right)\left(\mathbf{K}_{\mathbf{x}_{j}^{i}}-\mathbf{u}_{i}\right)^{T}$ |
| Step 4 | Compute the kernel scatter matrix of intraclass $\mathbf{k}_{w}=\lambda \mathbf{k}_{w_{1}}+(1-\lambda) \mathbf{k}_{w_{2}}$ |
| Step 5 | Compute $\boldsymbol{\alpha}^{*}=\mathbf{k}_{w}^{-1}\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right), \mathbf{y}_{0}=\boldsymbol{\alpha}^{T}\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right)$ |
| Step 6 | Get the optimal solution of $(16)$ |
| Step 7 | Place the inspected process variable as zero in original samples |
| Step 8 | Project the new samples into the feature space |
| Step 9 | Compute the contribution of one variable at one time with FNN in MKFDA |
| Step 10 | Repeat the above course for the remaining variables |
| Outputs | The distance measure $\boldsymbol{\delta}=\left[\delta_{1}, \delta_{2}, \ldots, \delta_{n}\right]$ of each original variable is obtained |

combination of samples in (11), according to kernel-based method,

$$
\begin{equation*}
\mathbf{w}=\sum_{i=1}^{N} \alpha_{i} \Phi\left(\mathbf{x}_{i}\right)=\boldsymbol{\psi} \boldsymbol{\alpha} \tag{11}
\end{equation*}
$$

where $\boldsymbol{\alpha}=\left(\alpha_{1}, \alpha_{2}, \ldots, \alpha_{N}\right)^{T}$.
Moreover, nonlinear transformation function $\Phi\left(\mathbf{x}_{1}\right)$, $\Phi\left(\mathbf{x}_{2}\right), \ldots, \Phi\left(\mathbf{x}_{N}\right)$ of samples can be projected to feature space $H$ with direction $\mathbf{w}$ in

$$
\begin{align*}
\mathbf{w}^{T} \Phi\left(\mathbf{x}_{i}\right) & =\boldsymbol{\alpha}^{T} \psi^{T} \Phi\left(\mathbf{x}_{\mathbf{i}}\right) \\
& =\boldsymbol{\alpha}^{T}\left(\Phi\left(\mathbf{x}_{1}\right), \Phi\left(\mathbf{x}_{2}\right), \ldots, \Phi\left(\mathbf{x}_{N}\right)\right)^{T} \Phi\left(\mathbf{x}_{i}\right) \\
& =\boldsymbol{\alpha}^{T}\left(k\left(\mathbf{x}_{1}, \mathbf{x}_{i}\right), k\left(\mathbf{x}_{2}, \mathbf{x}_{i}\right), \ldots, k\left(\mathbf{x}_{N}, \mathbf{x}_{i}\right)\right),  \tag{12}\\
& =\boldsymbol{\alpha}^{T} \mathbf{k}_{x_{i}}
\end{align*}
$$

From (11), for all $\mathbf{x} \in \mathbf{R}^{d}$, assume that $\mathbf{k}_{x}=\left(k\left(\mathbf{x}_{1}, \mathbf{x}\right)\right.$, $\left.k\left(\mathbf{x}_{2}, \mathbf{x}\right), \ldots, k\left(\mathbf{x}_{2}, \mathbf{x}\right)\right)^{T}$ and projection of mean vector $\mathbf{m}_{i}$ with direction $\mathbf{w}^{*}$ in feature space $H$ is

$$
\begin{equation*}
\mathbf{w}^{*} \mathbf{m}_{i}=\boldsymbol{\alpha}^{T} \psi^{T}\left(\frac{1}{N_{i}} \sum_{j=1}^{N_{i}} \Phi\left(\mathbf{x}_{j}^{i}\right)\right)=\boldsymbol{\alpha}^{T} \mathbf{u}_{i} \tag{13}
\end{equation*}
$$

where $\mathbf{u}_{i}=\left(\left(1 / N_{i}\right) \sum_{j=1}^{N_{i}} \Phi\left(\mathbf{x}_{1}\right) \Phi\left(\mathbf{x}_{j}^{i}\right),\left(1 / N_{i}\right) \sum_{j=1}^{N_{i}} \Phi\left(\mathbf{x}_{2} \Phi\left(\mathbf{x}_{j}^{i}\right)\right.\right.$, $\ldots,)^{T}$.

From (12) and (13), we have

$$
\begin{align*}
\mathbf{w}^{T} \mathbf{S}_{b} \mathbf{w} & =\boldsymbol{\alpha}^{T} \mathbf{k}_{b} \boldsymbol{\alpha} \\
\mathbf{w}^{T} \mathbf{S}_{\omega} \mathbf{w} & =\boldsymbol{\alpha}^{T} \mathbf{k}_{\omega} \boldsymbol{\alpha} \tag{14}
\end{align*}
$$

where $\mathbf{k}_{b}=\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right)\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right)^{T}, \mathbf{k}_{w}=\mathbf{k}_{w_{1}}+\mathbf{k}_{w_{2}}, \mathbf{k}_{w_{i}}=$ $\sum_{j=1}^{N_{i}}\left(\mathbf{k}_{x_{j}^{i}}-\mathbf{u}_{i}\right)\left(\mathbf{k}_{x_{j}^{i}}-\mathbf{u}_{i}\right)^{T}$.

Since fisher criterion function is optimal solution of (15), vector $\mathbf{w}$ can be resolved as $\boldsymbol{\alpha}$ in the following fisher criterion (16) $[24]$ :

$$
\begin{align*}
& J_{H}(\mathbf{w})=\frac{\mathbf{w}^{T} \mathbf{S}_{b} \mathbf{w}}{\mathbf{w}^{T} \mathbf{S}_{b} \mathbf{w}},  \tag{15}\\
& J_{H}(\boldsymbol{\alpha})=\frac{\boldsymbol{\alpha}^{T} \mathbf{k}_{b} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^{T} \mathbf{k}_{w} \boldsymbol{\alpha}} \tag{16}
\end{align*}
$$

Furthermore, the solution of optimal vector $\boldsymbol{\alpha}^{*}$ and $\mathbf{y}_{0}$ can be solved [25] with

$$
\begin{gather*}
\boldsymbol{\alpha}^{*}=\mathbf{k}_{w}^{-1}\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right) \\
\mathbf{y}_{0}=\frac{\left(\mathbf{w}^{*}\right)^{T} \mathbf{m}_{1}+\left(\mathbf{w}^{*}\right)^{T} \mathbf{m}_{2}}{2}=\frac{\left(\boldsymbol{\alpha}^{*}\right)^{T}\left(\mathbf{u}_{1}+\mathbf{u}_{2}\right)}{2} \tag{17}
\end{gather*}
$$

Thus, the corresponding function of kernel fisher discriminant function is obtained as

$$
\begin{equation*}
g(\mathbf{x})=\left(a^{*}\right)^{T} \mathbf{k}_{\mathbf{x}}-\mathbf{y}_{0} \tag{18}
\end{equation*}
$$

3.3. Multikernel Fisher Discriminant Analysis. From Section 3.2, the solution of maximizing (15) equals to the solution of maximizing (16). Assume that $\boldsymbol{\alpha}^{*}=\mathbf{k}_{w}^{-1}\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right)$ is optimal solution to classification effect, whereas $\boldsymbol{\alpha}^{*}$ is both determined by kernel scatter matrix $\mathbf{k}_{w}$ and difference of kernel mean vector $\left(\mathbf{u}_{1}-\mathbf{u}_{2}\right)$. In the condition of independent and identically distributed, kernel mean of samples is independent with number of samples. It indicates difference of kernel mean vector ( $\mathbf{u}_{1}-\mathbf{u}_{2}$ ) doing nothing with the unbalance of samples. So $\boldsymbol{\alpha}^{*}$ is only determined by kernel scatter matrix $\mathbf{k}_{w}$ for intraclass. If distribution of different variables differed, it should result in the contributions $\mathbf{k}_{w_{1}}, \mathbf{k}_{w_{2}}, \ldots, \mathbf{k}_{w}$ not in the similar interval. Besides that the solution of $\boldsymbol{\alpha}^{*}$ is not the optimal one. Hence, in order to avoid the influence of different distribution for samples, we presented multikernel fisher discriminant analysis method. It advances the kernel criterion function $\mathbf{k}_{w}=\left(\mathbf{k}_{w_{1}}+\mathbf{k}_{w_{2}}\right)$ into

$$
\begin{equation*}
\mathbf{k}_{w}=\lambda \mathbf{k}_{w_{1}}+(1-\lambda) \mathbf{k}_{w_{2}} \tag{19}
\end{equation*}
$$



Figure 4: The technological process of Tennessee Eastman.
where $\lambda(\lambda \in[0,1])$ is the adjustable MATLAB parameter and $\mathbf{k}_{\omega_{1}}$ and $\mathbf{k}_{\omega_{2}}$ are the kernel matrix computed with each suitable kernel function from Section 3.2 (i)/(ii)/(iii).

In this way, the influence with different sample distributions is considered with the suitable kernel function.

The above algorithm in this paper can be chiefly described in Table 1. In this way, the contribution of each original process variable $q_{i}$ to the certain fault is measured.

## 4. Fault Separation of Tennessee Eastman with Redundancy Variables

4.1. Tennessee Eastman Chemical Process. Tennessee Eastman (TE) is a classical chemical process created by Eastman Chemical Company in 1993 [26]. Its technological process is shown in Figure 4. There are four reactants (A, C, D, and E) and two products $(\mathrm{G}, \mathrm{H})$. Besides that, there is one inert material B and byproduct F .

In TE process, the dynamic TE model is composed of five major units: a reactor, a separator, a stripper, a condenser, and a compressor. Each unit can be expressed with some equations, in all of 148 algebraic equations and 30 differential equations. So it becomes one of the most complex models and is widely used to test study algorithm with control, system
monitor, fault diagnosis, and so forth. Here, we take Tennessee Eastman as the study object to measure its fault separation ability with our method.
4.2. Nonlinear Fault Separation of Redundancy Variables. In TE process, there are 41 observed variables and 12 manipulated variables from controller, some of which are nonlinear redundancy variables. Moreover, there are 20 types of classical fault in TE process shown in Table 2. Since Fault9 and Fault11 are nonlinear overlapped together shown in Figure 5, we take their fault separation as the study goal, meanwhile, 53 process variables must be screened for their multicollinearity and nonlinear redundancy. Process data of TE is simulated at one-minute sampling time in MATLAB software from Downs [27]. All the measurements have Gaussian noise. A total of 1000 samples are collected for training, where 800 data are collected for Fault9 and 200 for Fault1l. In addition, 835 samples are applied to test separation validity with 644 for Fault9 and 171 for Fault11.
4.3. Results and Discussion. If we distinguished Fault9 and Fault11, there are 53 variables to be considered in all. Therefore, we compute the contribution of 53 variables with mentioned method to see the importance of each process variables


Figure 5: The distribution of Fault9 and Fault11, described in 2dimension diagram with $x$-axis of process variable Vab. 13 and $y$-axis of process variable Vab.21.


Figure 6: Contribution of all the 53 process variables to distinguish Fault9 and Fault11.
on faults. Multikernel function is selected as Gaussian kernel and polynomial kernel, each comprised of $50 \%$. The contributions of each variable to the faults are computed with steps in Section 3.3 that is shown in Figure 6 and Table 3. From large to small, the proper importance of all the 53 process variables is reordered as \{Vab.21, Vab.13, Vab.9, Vab.16, Vab.7, Vab.20, Vab.11, Vab.2, Vab.12, Vab.8, Vab.19, Vab.5, Vab.22, Vab.6, Vab.3, Vab.18, Vab.14, Vab.15, Vab.17, Vab.10, Vab.41, Vab.40, Vab.27, Vab.23, Vab.29, Vab.31, Vab.26, Vab.33, Vab.25, Vab.32, Vab.4, Vab.24, Vab.30, Vab.35, Vab.34, Vab.37, Vab.36, Vab.28, Vab.39, Vab.38, Vab.1, Vab.53, Vab.52, Vab.51, Vab.50, Vab.49, Vab.48, Vab.47, Vab.46, Vab.45, Vab.44, Vab.43, Vab.42\}.

In the Following, the curves of the first two important Vab. 21 and Vab. 13 in TE process are given in Figures 7(a) and 7 (b) and Figures 8(a) and 8(b), respectively. It expresses the

(a) Process Vab. 21 (reactor coolant temperature) in Fault9

(b) Process Vab. 21 (reactor coolant temperature) in Fault11

Figure 7: The changing of process Vab. 21 in actual TE.
strong variation of process variables Vab. 21 and Vab.13, actually.

According to the sequence of each process variable, the different feature sets are constructed as $\{$ Vab.21\}, \{Vab.21, Vab.13\}, \{Vab.21, Vab.13, Vab.9\}, and so on. Nonlinear pattern classification of Fault9 and Fault11 is tested with support vector machine (SVM), which is widely used in pattern recognition. The parameters of SVM are optimized with crossvalidation $c=2035$ and $g=1024$. With the above variable sets, the accuracy of fault separation between Fault9 and Fault11 is successively tested. The results are shown in Figure 9 and Table 4. It reveals that the separation accuracy becomes lower when the considered variables increase.

From the above results, we conclude that (1) if all the 53 process variables were used to separate Fault9 and Fault 11, right ratio is merely $72.12 \%$. It indicates that not all of the variables are directly related to certain fault. Some redundancy or irrelevant variables may decrease the classification accuracy


Figure 8: The changing of process Vab. 13 in actual TE.


Figure 9: The accuracy with different feature sets to indentify Fault9 and Fault11 with testing data.

TABLE 2: State distribution in TE process.

| Fault | Disturbance | Type |
| :--- | :---: | :---: |
| 1 | $A / C$ feed ratio, $B$ composition constant | Step |
| 2 | $B$ composition, $A / C$ ratio constant | Step |
| 3 | $D$ feed temperature | Step |
| 4 | Reactor cooling water inlet temperature | Step |
| 5 | Condenser cooling water inlet temperature | Step |
| 6 | $A$ feed loss | Step |
| 7 | $C$ header pressure loss-reduced availability | Step |
| 8 | $A, B, C$ feed composition | Random |
| 9 | $D$ feed temperature | Random |
| 10 | $C$ feed temperature | Random |
| $\mathbf{1 1}$ | Reactor cooling water inlet temperature | Random |
| 12 | Condenser cooling water inlet temperature | Random |
| 13 | Reaction kinetics | Slow drift |
| 14 | Reactor cooling water valve | Sticking |
| 15 | Condenser cooling water valve | Sticking |
| $16-20$ | Unknown | Unknown |

Table 3: The contributions of 53 process variables to fault separation.

| Process variable | Contribution | Process variable | Contribution |
| :--- | :---: | :---: | :---: |
| Vab. 21 | 2.8273 | Vab. 17 | 0.0000 |
| Vab. 13 | 2.1145 | Vab. 10 | 0.0000 |
| Vab. 9 | 1.2318 | Vab. 41 | 0.0000 |
| Vab. 16 | 1.1313 | Vab. 40 | 0.0000 |
| Vab. 7 | 0.2687 | Vab.27 | 0.0000 |
| Vab. 20 | 0.1319 | Vab.23 | 0.0000 |
| Vab. 11 | 0.0522 | Vab.29 | 0.0000 |
| Vab.2 | 0.0355 | Vab.31 | 0.0000 |
| Vab. 12 | 0.0259 | Vab.26 | 0.0000 |
| Vab. 8 | 0.0191 | Vab.33 | 0.0000 |
| Vab. 19 | 0.0092 | Vab. 25 | 0.0000 |
| Vab. 5 | 0.0012 | Vab.32 | 0.0000 |
| Vab. 22 | 0.0006 | Vab. 4 | 0.0000 |
| Vab. 6 | 0.0004 | Vab.24 | 0.0000 |
| Vab. 3 | 0.0002 | Vab.30 | 0.0000 |
| Vab. 18 | 0.0000 | Vab.35 | 0.0000 |
| Vab. 14 | 0.0000 | $\vdots$ | $\vdots$ |
| Vab. 15 | 0.0000 | Vab. 42 | 0.000 |

and must be eliminated. (2) If the feature were selected as the first five process variables \{Vab.21, Vab.13, Vab.9, Vab.16, Vab. 7$\}$, the accuracy increases to the highest as $94.55 \%$. It means that the above five process variables are key to the fault separation. (3) If the model should be simplified at most, the process variable $\{$ Vab.21\} is the best feature variable. We can recognize Fault9 and Fault11 according to the process changing of Vab.21.

On the other hand, Fault9 stands for the random disturbance to $D$ feed temperature. Fault11 is random disturbance to reactor cooling water inlet temperature. While \{Vab.21,

Table 4: The accuracy with different feature sets with testing data.

| Feature set | Combination of variables | Accuracy |
| :---: | :---: | :---: |
| Set ${ }_{1}$ | Vab. 21 | 83.521\% |
| $\mathrm{Set}_{2}$ | Vab.21, Vab. 13 | 85.731\% |
| $\mathrm{Set}_{3}$ | Vab.21, Vab.13, Vab. 9 | 89.652\% |
| $\mathrm{Set}_{4}$ | Vab.21, Vab.13, Vab.9, Vab. 16 | 92.123\% |
| Set ${ }_{5}$ | Vab.21, Vab.13, Vab.9, Vab.16, Vab. 7 | 94.547\% |
| Set $_{6}$ | Vab.21, Vab.13, Vab.9, Vab.16, Vab.7, Vab. 20 | 92.532\% |
| : | : | : |
| Set $_{53}$ | Vab.21, Vab.13, Vab.9, Vab.16, Vab.7, Vab.20, ... Vab. 42 | 72.1199 |

Vab.13, Vab.9, Vab.16, Vab.7\} are the reactor coolant temperature, product separation pressure, reactor temperature, stripper pressure, reactor pressure, respectively, it is easy to see that the five selected variables are fairly relative to Fault9 and Fault11. The simulation results keep pace with the reality.

## 5. Conclusions

Nonlinear redundancy and multicollinearity variables can decrease the accuracy in classifier that must be eliminated. For the problem, FNN in MKFDA subspace is studied in the paper. Nonlinear variables are projected to a new linear higher dimension subspace with single-kernel fisher discriment analysis to get optimal classification with the intra-class nearest and inter-class farthest as most as possible. Furthermore, conventional single-kernel KFDA is expanded to multikernel method to solve the influence of each process variable with different distribution function. In order to reduce the higher dimension emerging in multi-KFDA subspace, FNN is composed to recognize the importance of each process variables on faults. According to simulation results in TE process, original variables are reduced to 5 in this paper, and the accuracy of tested right ratio reaches to $94.55 \%$ compared with tested right ratio $72.12 \%$ in the classifier between Fault 9 and Fault11.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research is supported by National Natural Science Foundation of China (no. 51376204, no. 51075418, no. 50905194, and no. 61174015), the Natural Science Foundation Project (no. CQCSTC2012jjA40026, no. CQCSTC2012jjA90011), and the Research Foundation of Chongqing University of Sci. and Tec. (no. CK2011Z01, no. CK2011B04).

## References

[1] D. H. Lim, S. H. Lee, and M. G. Na, "Smart soft-sensing for the feedwater flowrate at PWRs using a GMDH algorithm," IEEE Transactions on Nuclear Science, vol. 57, no. 1, pp. 340-347, 2010.
[2] M. J. Brusco and D. Steinley, "Exact and approximate algorithms for variable selection in linear discriminant analysis," Computational Statistics and Data Analysis, vol. 55, no. 1, pp. 123-131, 2011.
[3] F. A. Michelsen, B. F. Lund, and I. J. Halvorsen, "Selection of optimal, controlled variables for the TEALARC LNG process," Industrial and Engineering Chemistry Research, vol. 49, no. 18, pp. 8624-8632, 2010.
[4] F. Cipollini and G. M. Gallo, "Automated variable selection in vector multiplicative error models," Computational Statistics and Data Analysis, vol. 54, no. 11, pp. 2470-2486, 2010.
[5] A. J. Miller, Subset Selection in Regression, Chapman and Hall, London, UK, 2002.
[6] R. L. Masion and R. F. Gunst, Statistical Design and Analysis of Experiments with Applications to Engineering and Science, John Wiley \& Sons, Hoboken, NJ, USA, 2004.
[7] Y. Yang and J. O. Pederson, "A Comparative study on feature selection in text categorization," in Proceedings of the 14th International Conference on Machine Learning, pp. 412-420, 1997.
[8] K. Kira and L. A. Rendell, "The Feature selection problem: traditional methods and a new algorithm," in Proceedings of the 9th National Conference on Artificial Intelligence (AAAI '92), pp. 129-134, July 1992.
[9] B. Pfahringer, "Compression-based feature subset selection," in Proceedings of the Workshop on Data Engineering for Inductive Learning (IJCAI '95), pp. 101-106, 1995.
[10] J. C. Isaac, Kernel methods and component analysis for pattern recognition [Ph.D. thesis], 2007.
[11] J. C. Huang, J. S. Zhao, W. Sun, and Y. K. Ding, "PCA-based early fault diagnosis of solid waste incinerator," Chemical Industry and Engineering Progress, vol. 25, no. 12, pp. 1489-1492, 2006.
[12] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," Chemometrics and Intelligent Laboratory Systems, vol. 58, no. 2, pp. 109-130, 2001.
[13] J. D. Wu, P. H. Chiang, Y. W. Chang, and Y. J. Shiao, "An expert system for fault diagnosis in internal combustion engines using probability neural network," Expert Systems with Applications, vol. 34, no. 4, pp. 2704-2713, 2008.
[14] D. F. Wang, S. J. Wang, and J. He, "Maintaining and fault removing on hydraulic system of CAK6140," Machinery Design and Manufacture, vol. 7, pp. 161-162, 2010.
[15] J. H. Li and P. L. Cui, "Improved kernel fisher discriminant analysis for fault diagnosis," Expert Systems With Applications, vol. 36, no. 2, pp. 1423-1432, 2009.
[16] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre, "Generalized power method for sparse principal component analysis," Journal of Machine Learning Research, vol. 11, pp. 517-553, 2010.
[17] K. Kim, J.-M. Lee, and I.-B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," Chemometrics and Intelligent Laboratory Systems, vol. 79, no. 1-2, pp. 22-30, 2005.
[18] R. Jenssen, "Kernel entropy component analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pp. 847-860, 2010.
[19] N. Otopal, "Restricted kernel canonical correlation analysis," Linear Algebra and Its Applications, vol. 437, no. 1, pp. 1-13, 2012.
[20] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.
[21] H. Y. Wang and Z. H. Sheng, "Choice of the parameters for the phase space reconstruction of chaotic time series," Journal of Southeast University, vol. 30, no. 5, pp. 113-117, 2000.
[22] Z. B. Zhu and Z. H. Song, "A novel fault diagnosis system using pattern classification on kernel FDAsubspace," Expert Systems with Applications, vol. 38, no. 6, pp. 6895-6905, 2011.
[23] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Muller, "Fisher discriminant analysis with kernels," in Proceedings of the 9th IEEE Workshop on Neural Networks for Signal Processing (NNSP'99), pp. 41-48, August 1999.
[24] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, vol. 10, no. 5, pp. 1299-1319, 1998.
[25] J. Y. Gan and Y. W. Zhang, "Generalized kernel fisher optimal discriminant in pattern recognition," Pattern Recognition and Artificial Intelligence, vol. 15, no. 4, pp. 429-434, 2002.
[26] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," Computers and Chemical Engineering, vol. 17, no. 3, pp. 245-255, 1993.
[27] N. Lawrence Kicker, "Tennessee eastman," 2013, http://www .cheme.washington.edu/facresearch/faculty/ricker.html.

## Research Article

# Optimal Manoeuvres of Underactuated Linear Mechanical Systems: The Case of Controlling Gantry Crane Operations 

S. Woods and W. Szyszkowski<br>Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK, Canada S7N 5A9<br>Correspondence should be addressed to W. Szyszkowski; walerian.szyszkowski@usask.ca

Received 12 June 2013; Revised 26 September 2013; Accepted 27 September 2013; Published 12 January 2014
Academic Editor: Weihai Zhang
Copyright © 2014 S. Woods and W. Szyszkowski. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

A method of solving optimal manoeuvre control of linear underactuated mechanical systems is presented. The nonintegrable constraints present in such systems are handled by adding dummy actuators and then by applying Lagrange multipliers to reduce their action to zero. The open- and closed-loop control schemes can be analyzed. The method, referred to as the constrained modal space optimal control (CMSOC), is illustrated in the examples of gantry crane operations.


## 1. Introduction

Underactuated mechanical systems have fewer independent actuators than degrees of freedom (DOFs) to be controlled [1]. Typical nonlinear examples of such systems, usually with only several DOFs, are rigid multilink robotic manipulators with passive joints or any manipulator with flexible links (described by at least one mode of vibration). Linear examples include vibrating structures with continuously distributed mass (i.e., with theoretically infinite number of DOFs to describe them) such as masts, antennas, buildings, brides, and car suspension, controlled by discrete actuators. This paper presents a method of analyzing and simulating optimal manoeuvres between two given configurations (often referred to as point-to-point manoeuvres) for linear underactuated systems. The method combines optimal control theory with computational mechanics and the finite element (FE) technique, in particular.

The number of DOFs equal to the number of actuators will be referred to as actuated (after [1]), while all remaining DOFs will be referred to as underactuated (however, all DOFs are in fact controlled). The actuated and unactuated DOFs must satisfy a number of constraints equal to the number of unactuated DOFs and resulting from the equations governing the motion of such systems. For mechanical systems we assume that these constraints may be
nonintegrable (nonholonomic), meaning unactuated DOFs cannot be explicitly eliminated. Many of the techniques presented in the literature deal with underactuated problems by applying the constraints to eliminate the unactuated DOFs and then by solving the reduced fully actuated problems [24]. These approaches are limited to particular problems where the constraints can be simplified to a form making such mathematical manipulations possible. The method presented here is capable of dealing with any linear system, as it does not require the elimination of unactuated DOFs. Instead, the underactuated system is formulated as if it were fully actuated by adding "dummy" (zero-valued) actuators to all unactuated DOFs. The modal space is used in modelling the system motions. The method can be considered as an extension of the independent modal space control (IMSC; e.g., see [5]) into the underactuated problems, therefore it will be referred to as the constrained modal space optimal control (CMSOC) method. The system constraints resulting from underactuation are then determined by eliminating these dummy actuators. The constraints are algebraic in terms of controls but differential (nonintegrable) in terms of the DOFs. The algebraic form of the constraints is used to generate the so-called matrix of constraints, which is utilized to handle the nonintegrable constraints with the help of timevarying Lagrange multipliers. Pontryagin's principle is used to optimize the trajectory and actuation forces.

This paper presents the CMSOC method in a general form and then explains some details of the corresponding numerical procedure in the examples of standard two or three-DOF gantry crane operations. The method is verified by recreating the closed-loop control of the two-DOF gantry crane problem obtained in [1] via applying the classical technique and the open-loop optimal control considered in [3].

## 2. Problem Formulation

2.1. Dynamics of a General Underactuated System. The computational model for the motion of a linear mechanical system is represented by a standard form used in FE analysis:

$$
\begin{equation*}
\mathbf{M} \ddot{\mathbf{q}}+\mathbf{C} \dot{\mathbf{q}}+\mathbf{K q}=\mathbf{B F} \mathbf{F}_{a} \tag{1}
\end{equation*}
$$

where $\mathbf{q}$ and $\mathbf{F}_{a}$ are vectors of DOFs and activation forces, respectively, and $\mathbf{M}, \mathbf{C}$, and $\mathbf{K}$ are constant mass, damping, and stiffness matrices, respectively. In particular, (1) is suitable to model the dynamics of a range of actively controlled structural members undergoing small amplitude oscillations and finite translations. In underactuated systems $n_{a}$ independent actuation forces are to control $n>n_{a}$ number of DOFs. Matrix B of dimensions $n \times n_{a}$ assigns components of vector $\mathbf{F}_{\text {a }}$ to particular DOFs and obviously is not invertible if $n \neq n_{a}$. Clearly, the actuators via (1) control all DOFs of the system. For the purpose of analysis the DOFs can be divided into actuated $\left(\mathbf{q}_{a}\right)$ and unactuated $\left(\mathbf{q}_{r}\right)$ ones by rearranging these equations as follows:

$$
\begin{align*}
& {\left[\begin{array}{ll}
\mathbf{M}_{a a} & \mathbf{M}_{a r} \\
\mathbf{M}_{r a} & \mathbf{M}_{r r}
\end{array}\right]\left[\begin{array}{l}
\ddot{\mathbf{q}}_{a} \\
\ddot{\mathbf{q}}_{r}
\end{array}\right]+\left[\begin{array}{ll}
\mathbf{C}_{a a} & \mathbf{C}_{a r} \\
\mathbf{C}_{r a} & \mathbf{C}_{r r}
\end{array}\right]\left[\begin{array}{l}
\dot{\mathbf{q}}_{a} \\
\dot{\mathbf{q}}_{r}
\end{array}\right]+\left[\begin{array}{ll}
\mathbf{K}_{a a} & \mathbf{K}_{a r} \\
\mathbf{K}_{r a} & \mathbf{K}_{r r}
\end{array}\right]\left[\begin{array}{l}
\mathbf{q}_{a} \\
\mathbf{q}_{r}
\end{array}\right]} \\
& \quad=\left[\begin{array}{c}
\mathbf{F}_{a} \\
\mathbf{0}
\end{array}\right] . \tag{2a}
\end{align*}
$$

The bottom row represents the equations constraining the actuated and unactuated DOFs in the following form:

$$
\begin{equation*}
\mathbf{M}_{r a} \ddot{\mathbf{q}}_{a}+\mathbf{C}_{r a} \dot{\mathbf{q}}_{a}+\mathbf{K}_{r a} \mathbf{q}_{a}+\mathbf{M}_{r r} \ddot{\mathbf{q}}_{r}+\mathbf{C}_{r r} \dot{\mathbf{q}}_{r}+\mathbf{K}_{r r} \mathbf{q}_{r}=0 . \tag{3}
\end{equation*}
$$

The system can formally be converted to a fully actuated one by using (3) to explicitly determine vector $\mathbf{q}_{r}$ in terms of $\mathbf{q}_{a}$ (i.e, $\mathbf{q}_{r}=g\left(\mathbf{q}_{a}\right)$ ), and then by substituting this vector to the top row of (2a) to obtain $\mathbf{F}_{a}=\mathbf{F}_{a}\left(\mathbf{q}_{a}, g\left(\mathbf{q}_{a}\right)\right)$. Unless some matrices in (3) vanish, it is not generally possible, and therefore these constraints are considered as nonholonomic.

The control task for vector $\mathbf{F}_{a}$ in (1) is to manoeuvre the system from an initial state to a final state described by the following boundary conditions (point-to-point manoeuvre):

$$
\begin{align*}
\mathbf{q}(0)=\mathbf{q}_{\mathbf{0}}, & \dot{\mathbf{q}}(0)=\dot{\mathbf{q}}_{\mathbf{0}} \\
\mathbf{q}\left(t_{f}\right)=\mathbf{q}_{\mathbf{f}}, & \dot{\mathbf{q}}\left(t_{f}\right)=\dot{\mathbf{q}}_{\mathbf{f}}
\end{align*}
$$

It should be emphasised that no trajectory is specified in this task. A particular trajectory satisfying (1) and (4) may be determined if extra conditions are imposed on the system. We
will identify such a trajectory by optimizing the performance index as discussed in the next section. Note that this problem is different from a typical trajectory tracking problem in which instead of (4) the task is specified as the system output in the form

$$
\begin{equation*}
\mathbf{y}=\mathbf{h}(\mathbf{q}) \tag{5a}
\end{equation*}
$$

Several methods have been proposed to solve the inverse problems of finding the input $\mathbf{F}_{a}$ for the output $\mathbf{y}$ as defined by (1) and (5a), notably the servo-constraint approach [6-9] and the flatness method [10, 11]. In particular, differentiating (5a) with $n_{a}$ outputs twice one obtains $\ddot{\mathbf{y}}=\mathbf{H} \ddot{\mathbf{q}}+\overline{\mathbf{h}}$, where the size of matrix $\mathbf{H}$ is $n_{a} \times n$; then square matrix $\mathbf{H M}^{-1} \mathbf{B}$ is required to be nonsingular to solve the problem. This condition does not apply in the method presented here since our output is given only in terms of (4), that is, the system's initial and final configurations.

The set of (1) or (2a) is uncoupled when mapped into modal space, where vector of DOFs $\mathbf{q}$ (size $n$ ) is transformed to the equally sized vector of modal variables $\boldsymbol{\eta}=\left[\eta_{1} \cdots \eta_{n}\right]^{T}$. Similarly, vector $\mathbf{F}$ is related to an equally sized vector of modal controls $\mathbf{U}=\left[u_{1} \cdots u_{n}\right]^{T}$. These transformations are

$$
\begin{gather*}
\mathbf{q}=\boldsymbol{\Phi} \boldsymbol{\eta}  \tag{6a}\\
\mathbf{U}=\left(\boldsymbol{\Phi}^{T} \mathbf{B}\right) \mathbf{F}_{a}=\widehat{\mathbf{B}} \mathbf{F}_{a} \tag{6b}
\end{gather*}
$$

where $\widehat{\mathbf{B}}=\boldsymbol{\Phi}^{T} \mathbf{B}$ is the transfer matrix of size $n \times n_{a}$ between vectors $\mathbf{F}_{a}$ and $\mathbf{U}$ and mode shape matrix $\boldsymbol{\Phi}=\left[\boldsymbol{\varphi}_{1} \cdots \boldsymbol{\varphi}_{n}\right]$ relates vectors $\mathbf{q}$ and $\boldsymbol{\eta}$. The $M$-normalized matrix $\boldsymbol{\Phi}$, consisting of $n$ modal shape vectors $\boldsymbol{\varphi}_{i}$ (each with $n$ components), satisfies the following orthogonality conditions:

$$
\begin{align*}
& \boldsymbol{\Phi}^{T} \mathbf{M} \boldsymbol{\Phi}=\mathbf{I}  \tag{7a}\\
& \boldsymbol{\Phi}^{T} \mathbf{K} \boldsymbol{\Phi}=\mathbf{\Omega} \tag{7b}
\end{align*}
$$

where $\mathbf{I}$ is the unitary matrix and $\boldsymbol{\Omega}$ is the diagonal matrix of ordered frequencies with the terms $\Omega_{i i}=\omega_{i}^{2}$. Each mode shape vector $\boldsymbol{\varphi}_{i}$ and frequency $\omega_{i}$ are solutions to the eigenvalues problem $\left(\mathbf{K}-\omega_{i}^{2} \mathbf{M}\right) \boldsymbol{\varphi}_{i}=0(i=1, \ldots, n)$. The above modal analysis (or operations defined by (6a)-(7b)) is carried out routinely in the FE approach, even for problems with a very large number of DOFs (large $n$ ).

The equations of motion (1) become uncoupled when applying transformations (6a) and (6b) subject to orthogonality conditions (7a) and (7b) and take the following form:

$$
\begin{equation*}
\mathbf{I} \ddot{\eta}+\Delta \dot{\eta}+\Omega \boldsymbol{\eta}=\mathbf{U} \tag{8}
\end{equation*}
$$

where for the Rayleigh damping (i.e., $\mathbf{C}=\alpha \mathbf{M}+\beta \mathbf{K}$ ) the diagonal terms of $\Delta$ are $\Delta_{i i}=2 \varsigma_{i} \omega_{i}=\boldsymbol{\varphi}_{i}^{T} \mathbf{C} \boldsymbol{\varphi}_{i}$ and where $\varsigma_{i}=\alpha / 2 \omega_{i}+\beta \omega_{i} / 2$ are the modal damping ratios. Note that a rigid body translation, for which $\omega_{i}=0$, is also included in the above equation.

A continuous system, or an FE model (1) of the system described by $n$ DOFs (where $n$ may be a large number), can be approximated by (8) with only $n_{m}$ significant modes
considered, where usually $n_{m} \ll n$. The number of significant modes that should be sufficient to represent such a system is generally problem related and depends mainly on its physical characteristics, the spatial distribution, and frequency content of the loading [12].

In the system approximated by $n_{m}$ modes (similarly as for the system's DOFs) one can consider $n_{a}$ modes as actuated and $n_{r}=n_{m}-n_{a}$ modes as unactuated. Then matrix $\Phi$ will be reduced to only $n_{m}$ columns, and transfer matrix $\widehat{\mathbf{B}}$ in (6a) and ( 6 b ) will be of dimensions $n_{m} \times n_{a}$.

In order to control all $n_{m}$ modes this system can be made artificially fully actuated by adding $n_{r}=n_{m}-n_{a}$ dummy actuation forces (zero valued) forming subvector $\mathbf{F}_{\mathbf{d}}$. For the purpose of analysis vector $\mathbf{F}_{\mathbf{a}}$ in (1) is replaced by the augmented force vector $\mathbf{F}_{\mathrm{a}}^{\prime}=\left[\begin{array}{ll}\mathbf{F}_{\mathbf{a}}^{T} & \mathbf{F}_{\mathrm{d}}^{T}\end{array}\right]^{T}$ containing $n_{a}$ real actuation forces forming subvector $\mathbf{F}_{\mathbf{a}}$ and $n_{r}$ dummy actuators forming subvector $\mathbf{F}_{\mathbf{d}}$. Then, in (1) and (6a) and (6b), matrix $\mathbf{B}$ of dimensions $n \times n_{a}$ is replaced by matrix $\mathbf{B}^{\prime}$ of dimensions $n \times n_{m}$ (this matrix assigns the component of $\mathbf{F}_{\mathrm{a}}^{\prime}$ to particular nodes). Consequently in (6b) matrix $\widehat{\mathbf{B}}$ of dimensions $n \times n_{a}$ is replaced by a square matrix $\widehat{\mathbf{B}}^{\prime}$ of dimensions $n_{m} \times n_{m}$ ( $n_{m}$ modes controlled by $n_{m}$ actuators). The dummy actuators $\mathbf{F}_{\mathbf{d}}$ should be placed in such a way that $\widehat{\mathbf{B}}^{\prime}$ is nonsingular.

In the inverse dynamics based control analysis, each control $\mathbf{U}_{i}$ can be obtained from (8) by substituting the corresponding prescribed mode $\boldsymbol{\eta}_{i}$. Then, for known vector $\mathbf{U}$, the actuation forces $\mathbf{F}_{\mathbf{a}}^{\prime}$ should be determined by inverting transformation (6b) in which matrix $\widehat{\mathbf{B}}^{\prime}$ (instead of $\widehat{\mathbf{B}}$ ) is now square and nonsingular (the dummy actuators were added to the system only to ensure that this inversion is possible). In the next step, after computing the inverse of operation (6b), the dummy actuators will be eliminated by giving them zero values. For that purpose the inverse matrix $\left(\widehat{\mathbf{B}}^{\prime}\right)^{-1}$, representing the mapping from modal controls $\mathbf{U}=$ $\left[\mathbf{U}_{\mathbf{a}}^{T} \mid \mathbf{U}_{\mathbf{r}}^{T}\right]^{T}$ to actuation forces $\mathbf{F}_{a}^{\prime}$ for any augmented system of size $n_{m} \times n_{m}$, is partitioned as follows:

$$
\left(\widehat{\mathbf{B}}^{\prime}\right)^{-1} \mathbf{U}=\mathbf{F}_{a}^{\prime} \Longrightarrow\left[\begin{array}{c|c|c}
\widetilde{\mathbf{B}}_{\mathbf{a}} & \widetilde{\mathbf{B}}_{\mathbf{r}}  \tag{9}\\
\hline \mathbf{A}_{\mathbf{a}} & \mathbf{A}_{\mathbf{r}}
\end{array}\right]\left[\begin{array}{c}
\mathbf{U}_{\mathbf{a}} \\
\hline \mathbf{U}_{\mathbf{r}}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{F}_{\mathbf{a}} \\
\hline \mathbf{F}_{\mathrm{d}}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{F}_{\mathbf{a}} \\
\hline \mathbf{0}
\end{array}\right]
$$

Square submatrix $\widetilde{\mathbf{B}}_{\mathrm{a}}$ is of size $n_{a} \times n_{a}$ and square submatrix $\mathbf{A}_{\mathbf{r}}$ is of size $n_{r} \times n_{r}$. To be consistent with modes classifications (actuated and unactuated), vectors $\mathbf{U}_{\mathbf{a}}=\left[u_{1} \cdots u_{n_{a}}\right]^{T}$ and $\mathbf{U}_{\mathbf{r}}=\left[u_{n_{a}+1} \cdots u_{n_{m}}\right]^{T}$ are referred to as actuated and unactuated modal controls, respectively. Given the null-valued dummy force vector $\mathbf{F}_{\mathbf{d}}=[0 \cdots 0]^{T}$ (size $\left.n_{r} \times 1\right)$ the bottom $n_{r}$ rows of operation (9) (lower partition) define constraints on the system in terms of all modal controls, in the following form:

$$
\begin{equation*}
\mathbf{A} \mathbf{U}=\mathbf{A}_{a} \mathbf{U}_{a}+\mathbf{A}_{r} \mathbf{U}_{r}=\mathbf{0} \tag{10}
\end{equation*}
$$

Matrix $\mathbf{A}=\left[\mathbf{A}_{\mathbf{a}} \mid \mathbf{A}_{\mathbf{r}}\right]\left(\right.$ size $\left.n_{r} \times n_{m}\right)$ defines the system constraints written algebraically in terms of modal controls. Since (10) is homogeneous matrix $\mathbf{A}$ can be normalized such that the diagonal terms corresponding to controls $\mathbf{U}_{a}$ are set
to unity (i.e., $A_{i i}=1$ ). In this form $\mathbf{A}$ becomes independent of the choice of dummy actuators, which reflects the fact that these zero-force actuators were added somewhat arbitrarily only to facilitate the elimination process, that is, to satisfy the constraints in (9). Matrix $\mathbf{A}$ is discussed with more details in [13, 14].

Real actuation force(s) may be obtained from the top partition of operation (9) in terms of all modal controls in vector $\mathbf{U}$. They can also be obtained in terms of only actuated modal controls in vector $\mathbf{U}_{\mathbf{a}}$ by applying $n_{r}$ constraints (10) to eliminate unactuated modal controls $\mathbf{U}_{\mathbf{r}}$. Thus, $n_{a}$ components of actuator forces in vector $\mathbf{F}_{\mathrm{a}}$ can be obtained in terms of $n_{a}$ actuated modal controls in vector $\mathbf{U}$ a from the following operation:

$$
\begin{equation*}
\mathbf{F}_{\mathbf{a}}=\overline{\mathbf{B}} \mathbf{U}_{\mathbf{a}} \tag{11}
\end{equation*}
$$

Square matrix $\overline{\mathbf{B}}=\widetilde{\mathbf{B}}_{\mathbf{a}}-\widetilde{\mathbf{B}}_{\mathbf{r}} \mathbf{A}_{\mathbf{r}}^{-1} \mathbf{A}_{\mathbf{a}}\left(\right.$ size $\left.n_{a} \times n_{a}\right)$ is referred to as the pseudotransfer matrix, and it relates actuated modal controls to real actuator forces. Similar to the normalized constraint matrix $\mathbf{A}$, the pseudotransfer matrix $\overline{\mathbf{B}}$ is independent of the choice of dummy actuators.
2.2. Optimal Manoeuvres of Underactuated Systems. In linear optimal control [15], the manoeuvre is optimal if, for a given task, it minimizes the performance index:

$$
\begin{equation*}
J=\frac{1}{2} \int_{0}^{t_{f}}\left(\left(\boldsymbol{\eta}^{T} \widehat{\mathbf{Q}}_{\mathbf{d}} \boldsymbol{\eta}\right)+\left(\dot{\boldsymbol{\eta}}^{T} \widehat{\mathbf{Q}}_{\mathbf{v}} \dot{\boldsymbol{\eta}}\right)+\left(\mathbf{U}^{T} \widehat{\mathbf{R}} \mathbf{U}\right)\right) d t \longrightarrow \min \tag{12}
\end{equation*}
$$

where $\widehat{\mathbf{Q}}_{\mathbf{d}}, \widehat{\mathbf{Q}}_{\mathbf{v}}$, and $\widehat{\mathbf{R}}$ are matrices, with the diagonal terms $\widehat{Q}_{d i i}, \widehat{Q}_{v i i}$, and $\widehat{R}_{i i}\left(i=1, \ldots, n_{m}\right)$, that are weights for the system's potential energy, kinetic energy, and actuator work, respectively. Note that $n_{m}$ modal variables and modal controls are included in (12); however, these modes are not independent because of constraint (10), resulting from underactuation. Such a problem can be solved by applying Pontryagin's principle. Here we use the procedure described in [15]. Hamiltonian $H$ for the constrained optimization problem involving performance index (12), uncoupled equations of motion (8), and constraints (10) is defined in the following form:

$$
\begin{align*}
H= & -\frac{1}{2}\left(\boldsymbol{\eta}^{T} \widehat{\mathbf{Q}}_{\mathbf{d}} \boldsymbol{\eta}+\dot{\boldsymbol{\eta}}^{T} \widehat{\mathbf{Q}}_{\mathbf{v}} \dot{\boldsymbol{\eta}}+\mathbf{U}^{T} \widehat{\mathbf{R}} \mathbf{U}\right)+\mathbf{P}_{\mathbf{d}} \dot{\boldsymbol{\eta}}  \tag{13}\\
& +\mathbf{P}_{\mathbf{v}}(-\Delta \dot{\boldsymbol{\eta}}-\boldsymbol{\Omega} \boldsymbol{\eta}+\mathbf{U})+\mathbf{v}^{T} \mathbf{A} \mathbf{U} .
\end{align*}
$$

$\mathbf{P}_{\mathbf{d}}$ and $\mathbf{P}_{\mathbf{v}}$ are standard costate vectors related to modal position and velocity states ( $\boldsymbol{\eta}$ and $\boldsymbol{\eta}$ ) of a system, respectively. Vector $\mathbf{v}^{T}=\left[v_{1} \cdots v_{n_{r}}\right]$ represents the set of time-dependent Lagrange multipliers introduced to enforce constraints (10). These multipliers play a similar role to, for example, that of the multipliers used in the servo-constraint approach [6-9]
mentioned before. According to Pontryagin's principle the costate equations take the following form:

$$
\begin{gather*}
\dot{\mathbf{P}}_{\mathrm{d}}=-\frac{\partial H}{\partial \boldsymbol{\eta}}=\widehat{\mathbf{Q}}_{\mathrm{d}} \boldsymbol{\eta}+\boldsymbol{\Omega} \mathbf{P}_{\mathbf{v}}  \tag{14a}\\
\dot{\mathbf{P}}_{\mathrm{v}}=-\frac{\partial H}{\partial \dot{\boldsymbol{\eta}}}=\widehat{\mathbf{Q}}_{\mathrm{v}} \dot{\boldsymbol{\eta}}-\mathbf{P}_{\mathrm{d}}+\Delta \mathbf{P}_{\mathrm{v}} \tag{14b}
\end{gather*}
$$

The Hamiltonian is stationary with respect to modal control if

$$
\begin{equation*}
\frac{\partial H}{\partial \mathbf{U}}=-\widehat{\mathbf{R}} \mathbf{U}+\mathbf{P}_{\mathbf{v}}+\mathbf{A}^{T} \mathbf{v}=\mathbf{0} . \tag{15}
\end{equation*}
$$

Substituting (8) into (15) gives

$$
\begin{equation*}
\mathbf{P}_{\mathbf{v}}=\widehat{\mathbf{R}}(\mathbf{I} \ddot{\boldsymbol{\eta}}+\Delta \dot{\boldsymbol{\eta}}+\boldsymbol{\Omega} \boldsymbol{\eta})-\mathbf{A}^{T} \mathbf{v} \tag{16}
\end{equation*}
$$

Substituting (16) into (14b) yields

$$
\begin{align*}
\mathbf{P}_{\mathrm{d}}= & \widehat{\mathbf{Q}}_{\mathrm{v}} \dot{\boldsymbol{\eta}}-\widehat{\mathbf{R}}(\dot{\mathrm{I}} \dot{\boldsymbol{\eta}}+\Delta \ddot{\boldsymbol{\eta}}+\boldsymbol{\Omega} \dot{\boldsymbol{\eta}})  \tag{17}\\
& +\Delta \widehat{\mathbf{R}}(\mathrm{I} \ddot{\boldsymbol{\eta}}+\Delta \dot{\boldsymbol{\eta}}+\boldsymbol{\Omega} \boldsymbol{\eta})+\mathrm{A}^{T} \dot{\mathbf{v}}-\Delta \mathrm{A}^{T} \mathbf{v} .
\end{align*}
$$

Finally, substituting (17) into (14a) generates the following set of optimality equations:

$$
\begin{align*}
\widehat{\mathbf{R}} \ddot{\boldsymbol{\eta}} & +\left(\mathbf{2} \widehat{\mathbf{R}}-\widehat{\mathbf{Q}}_{\mathbf{v}}-\widehat{\mathbf{R}} \Delta^{2}\right) \ddot{\boldsymbol{\eta}}+\left(\widehat{\mathbf{R}} \boldsymbol{\Omega}^{2}+\widehat{\mathbf{Q}}_{\mathbf{d}}\right) \boldsymbol{\eta} \\
& -\left(\mathbf{A}^{T} \ddot{\mathbf{v}}-\Delta \mathbf{A}^{T} \dot{\mathbf{v}}+\boldsymbol{\Omega} \mathbf{A}^{T} \mathbf{v}\right)=\mathbf{0} . \tag{18}
\end{align*}
$$

Note that $n_{m}$ optimality equations (18) contain $n_{m}$ unknown components in $\boldsymbol{\eta}$ and $n_{r}$ unknown components in $\mathbf{v}$. Therefore, additional $n_{r}$ constraint equations (10) are required in order to obtain all the unknown modal variable functions in vector $\boldsymbol{\eta}$ and Lagrange multiplier functions in vector $\mathbf{v}$. However, the constraints must be written in terms of $\boldsymbol{\eta}$ not in terms of $\mathbf{U}$ (note the change in the constraints' form from algebraic to differential). The uncoupled equations of motion (8) are substituted into algebraic constraints (10) to obtain

$$
\begin{equation*}
\mathrm{A}(\mathrm{I} \ddot{\boldsymbol{\eta}}+\Delta \dot{\boldsymbol{\eta}}+\boldsymbol{\Omega} \boldsymbol{\eta})=\mathbf{0} . \tag{19}
\end{equation*}
$$

The number of $n_{m}+n_{r}((18)$ and (19)) is equal the unknown components in vectors $\boldsymbol{\eta}$ and $\mathbf{v}$.

Boundary conditions (4) are mapped into modal space by using the inverse of transformation (6a) or through the relation $\boldsymbol{\eta}=\boldsymbol{\Phi}^{T} \mathbf{M q}$ (obtained by additional substitution of condition (7a)). These transformed boundary conditions are

$$
\begin{array}{rlrl}
\boldsymbol{\eta}(0) & =\boldsymbol{\Phi}^{T} \mathbf{M} \mathbf{q}_{\mathbf{0}}, & \dot{\boldsymbol{\eta}}(0)=\boldsymbol{\Phi}^{T} \mathbf{M} \dot{\mathbf{q}}_{\mathbf{0}} \\
\boldsymbol{\eta}\left(t_{f}\right) & =\boldsymbol{\Phi}^{T} \mathbf{M} \mathbf{q}_{\mathbf{f}}, & & \dot{\boldsymbol{\eta}}\left(t_{f}\right)=\boldsymbol{\Phi}^{T} \mathbf{M} \dot{\mathbf{q}}_{\mathbf{f}} \tag{20}
\end{array}
$$

For fully actuated problems, the last term ( $\mathbf{A}^{T} \ddot{\mathbf{v}}-\Delta \mathbf{A}^{T} \dot{\mathbf{v}}+$ $\boldsymbol{\Omega} \mathbf{A}^{T} \mathbf{v}$ ) in optimality equations (18) vanishes because there are no constraints or Lagrange multipliers needed to enforce them. Therefore, a fully actuated problem involves only $n_{m}$ optimality equations (18) to be solved in terms of $n_{m}$ uncoupled modal variables in vector $\boldsymbol{\eta}$.

The solution to the combined set of (18), (19), and (20) can be efficiently obtained using symbolic differential operator $D^{n}=d^{n} / d t^{n}$. Substituting this operator into (19) and (20) and rewriting in matrix notation give

$$
\left[\begin{array}{c|c}
\mathbf{E} & -\widehat{\mathbf{E}}^{T}  \tag{21}\\
\hline \widetilde{\mathbf{E}} & \mathbf{0}
\end{array}\right]\left[\begin{array}{c}
\boldsymbol{\eta} \\
\boldsymbol{\nu}
\end{array}\right]=\mathbf{0} \quad \text { or } \quad \mathbf{E}_{\mathbf{p}} \mathbf{Y}=0
$$

where

$$
\begin{gather*}
\mathbf{E}=\widehat{\mathbf{R}} D^{4}+\left(2 \widehat{\mathbf{R}} \boldsymbol{\Omega}-\widehat{\mathbf{Q}}_{\mathbf{v}}-\widehat{\mathbf{R}} \Delta^{2}\right) D^{2}+\left(\widehat{\mathbf{R}} \Omega^{2}+\widehat{\mathbf{Q}}_{\mathbf{d}}\right), \\
\widehat{\mathbf{E}}=\mathbf{A}\left(\mathbf{I} D^{2}+\Delta D+\boldsymbol{\Omega}\right), \quad \widetilde{\mathbf{E}}=\mathbf{A}\left(\mathbf{I} D^{2}-\Delta D+\boldsymbol{\Omega}\right) . \tag{22}
\end{gather*}
$$

Matrix $\mathbf{E}_{\mathbf{p}}$ contains submatrices $\mathbf{E}, \widetilde{\mathbf{E}}$, and $-\widehat{\mathbf{E}}^{T}$. Vector $\mathbf{Y}=\left[\boldsymbol{\eta}^{T} \mid \boldsymbol{\nu}^{T}\right]^{T}$ contains all unknown modal variables and Lagrange multipliers. Note that in a fully actuated case, matrix $\mathbf{E}_{\mathbf{p}}$ in (21) consists only of submatrix $\mathbf{E}$ and vector $\mathbf{Y}=\boldsymbol{\eta}$.

The solution to a system described in form (21) involves the roots $r_{l}\left(l=1, \ldots, 4 n_{m}\right)$ of the characteristic equation for the determinant of $\mathbf{E}_{\mathbf{p}}$ [16], where operator $D$ is replaced by the auxiliary variable $r$ rendering a $4 n_{m}^{\text {th }}$ order polynomial. This operation is written as

$$
\begin{equation*}
\left.\operatorname{det} \mathbf{E}_{\mathbf{p}}\right|_{D \rightarrow r}=0 \tag{23}
\end{equation*}
$$

Generally, the $4 n_{m}$ roots of the characteristic equation (23) take the following form:

$$
\begin{equation*}
r_{l}= \pm \alpha_{k} \pm i \beta_{k} \quad\left(k=1, \ldots, n_{m}, l=1, \ldots, 4 n_{m}\right) . \tag{24}
\end{equation*}
$$

The positive real numbers $\alpha_{k}$ and $\beta_{k}$ characterize the response of the $k$ th mode of motion. For nonzero, unique roots, solution vector $\mathbf{Y}$ consists of $n_{m}+n_{r}$ components $Y_{j}$ that can be written in terms of $4 n_{m}$ independent elementary functions related to the roots (24), in the form [16]:

$$
\begin{array}{r}
Y_{j}=\sum_{k=1}^{n_{m}}\left[e^{\alpha_{k} t}\left(c_{k j}^{1} \sin \left(\beta_{k} t\right)+c_{k j}^{2} \cos \left(\beta_{k} t\right)\right)\right. \\
\left.+e^{-\alpha_{k} t}\left(c_{k j}^{3} \sin \left(\beta_{k} t\right)+c_{k j}^{4} \cos \left(\beta_{k} t\right)\right)\right]  \tag{25}\\
\text { where } j=1, \ldots, n_{m}+n_{r}
\end{array}
$$

Obviously, the frequency of $k$ th mode controlled by the actuators can be interpreted as $\omega_{k}^{a}=\beta_{k}$ and its rate of active attenuation (or amplification) as $\varsigma_{k}^{a}=\alpha_{k} / \beta_{k}$. If multiple roots and zero-valued roots are obtained from (23), then solution functions (25) must be modified to mathematically accommodate these situations. There are $4 n_{m}\left(n_{m}+n_{r}\right)$ unknown integration constants $c_{k j}^{1}, \ldots, c_{k j}^{4}$ contained in the solution functions (25).

Integration constants $c_{k j}^{1}, \ldots, c_{k j}^{4}$ are obtained by substituting the assumed form (25) into differential equations (18) and (19) and using the method of undetermined coefficients to generate $n_{m}+n_{r}$ sets of $4 n_{m}$ linear algebraic equations relating these constants. By replacing one set of $4 n_{m}$ equations


Figure 1: Gantry crane system.
with the set of $4 n_{m}$ boundary conditions (20), the integration constants can be solved simultaneously. All these symbolic operations, including the determination of the roots (24) and constants in (25), can be done automatically using the MAPLE mathematical software.

For closed-loop control, asymptotically convergent solution functions are required such that the control task is met over an infinite period of time $\left(t_{f} \rightarrow \infty\right)$. The resulting number of integration constants is reduced by half, as terms involving positive exponential $e^{\alpha_{k} t}$ in the solution form (25) disappear $\left(c_{k j}^{1}=c_{k j}^{2}=0\right)$.

To quantitatively measure the performance of closedloop control schemes settling time $t_{f}^{3 \%}$ is defined as the time needed for various variables to be reduced to within 3\% of their initial value (i.e., $e^{-\alpha_{k}^{\min } t_{f}^{350}}=0.03 \rightarrow t_{f}^{3 \%}=3.5 / \alpha_{k}^{\mathrm{min}}$ ).

The above procedure was applied to actively suppress vibrations of a spatial antenna mast in [13] and of plane frames in [14], the cases with the number of DOFs much greater than the number of significant modes included in the analysis (i.e., with $n \gg n_{m}$ ). In both cases only oscillating modes were controlled. Here the application of the above methodology is focused on various control schemes, which are demonstrated in controlling the translational and oscillating modes of a gantry crane.

## 3. Dynamics and Optimal Control of the Gantry Crane System

The gantry crane problem is one of the simplest underactuated mechanical systems involving two DOFs-cart translation and suspended load rotation-and a single actuator-a cart-driving force ( $n_{m}=2, n_{a}=n_{r}=1$ ).

The gantry crane model is shown in Figure 1. The model includes the mass of the cart $M$, the mass of the suspended load $m$, swing angle $\theta$, gravitational acceleration $g$, horizontal distance $a$ from the cart's initial position to the origin, and
length $L$ of the massless rigid link connecting the cart and load. The task is to manoeuvre the system from an initial resting state at some nonzero horizontal distance $(x=a$, $\theta=0$ ) to a final resting equilibrium state at the origin $(x=0$, $\theta=0$ ) by applying time-varying force $F_{a}$. Any finite cart translations are permitted, but swings of the suspended load are assumed to be sufficiently small for a linearized model to be valid. In modal space rigid body translation for such a manoeuvre is easily separated from the oscillatory motion of the suspended load. Dummy force $F_{d}$ is added to artificially make the system fully actuated and formulate the augmented gantry crane system.

This same gantry crane model was used in several papers dealing with control or/and optimization. Notably, a Lyapunov function was used in [1] to obtain an asymptotically stable (closed-loop) control (linear and nonlinear) for attenuating disturbances (nonzero initial positions) in the system, and optimal control by applying Pontryagin's principle was considered in [3]. Results for the linearized system are of interest because they serve as a useful comparison for the controls obtained in this paper. Similar problems of controlling the plane motion of gantry cranes were presented in [10, 11] using the concept of flatness. Various aspects of controlling gantry cranes, 3D operations were considered in [9, 17-19].

The gantry crane system shown in Figure 1 and its coordinate system are chosen to mimic those used in [1]. Matrices and vectors in the general equation of motion (1) take the following forms:

$$
\begin{gather*}
\mathbf{M}=\left[\begin{array}{cc}
M+m & -m \\
-m & m
\end{array}\right], \quad \mathbf{K}=\left[\begin{array}{cc}
0 & 0 \\
0 & \frac{m g}{L}
\end{array}\right], \\
\mathbf{B}=\left[\begin{array}{c}
1 \\
0
\end{array}\right], \quad \mathbf{B}^{\prime}=\left[\begin{array}{cc}
1 & 1 \\
0 & -1
\end{array}\right],  \tag{26}\\
\mathbf{q}=\left[\begin{array}{c}
x \\
L \theta
\end{array}\right], \quad \mathbf{F}_{a}^{\prime}=\left[\begin{array}{l}
F_{a} \\
F_{d}
\end{array}\right] .
\end{gather*}
$$

To be consistent with the assumptions made in $[1,3]$ no dissipative effects (i.e., friction, etc.) are considered ( $\mathbf{C = 0}=0$.

The initial and final conditions (consistent with [1]) take the following forms:

$$
\begin{align*}
\mathbf{q}(0) & =\left[\begin{array}{ll}
a & 0
\end{array}\right]^{T}  \tag{27}\\
\dot{\mathbf{q}}(0)=\mathbf{q}\left(t_{f}\right) & =\dot{\mathbf{q}}\left(t_{f}\right)=\left[\begin{array}{ll}
0 & 0
\end{array}\right]^{T} .
\end{align*}
$$

The modal analysis $\left(n_{m}=2\right)$ gives

$$
\begin{gather*}
\boldsymbol{\Omega}=\left[\begin{array}{cc}
0 & 0 \\
0 & \left(1+\frac{m}{M}\right) \frac{g}{L}
\end{array}\right],  \tag{28a}\\
\boldsymbol{\Phi}=\left[\begin{array}{cc}
\frac{1}{\sqrt{M+m}} & \sqrt{\frac{m}{M(M+m)}} \\
0 & \sqrt{\frac{M+m}{M m}}
\end{array}\right] . \tag{28b}
\end{gather*}
$$

The rigid body translational mode of motion is represented in (28a) by the frequency $\omega_{1}=0$ and the second vibrating mode (load swinging) is represented by the frequency $\omega_{2}=$ $\sqrt{(1+m / M)(g / L)}$.

The uncoupled modal equations of motion (8) become:

$$
\begin{equation*}
\ddot{\eta}_{1}=u_{1}, \quad \ddot{\eta}_{2}+\omega_{2}^{2} \eta_{2}=u_{2} . \tag{29}
\end{equation*}
$$

The augmented system transfer matrix $\widehat{\mathbf{B}}^{\prime}=\boldsymbol{\Phi}^{T} \mathbf{B}^{\prime}$ is obtained by the appropriate substitutions from (26) and (28b) into the general partitioned form (9):

$$
\left[\begin{array}{c|c}
\frac{M}{\sqrt{M+m}} & \sqrt{\frac{M m}{M+m}}  \tag{30}\\
\frac{m}{\sqrt{M+m}} & -\sqrt{\frac{M m}{M+m}}
\end{array}\right]\left[\begin{array}{l}
u_{1} \\
\hline u_{2}
\end{array}\right]=\left[\begin{array}{c}
F_{a} \\
\hline F_{d}
\end{array}\right]=\left[\begin{array}{c}
F_{a} \\
\hline 0
\end{array}\right]
$$

Modal controls $u_{1}$ and $u_{2}$ are considered actuated and unactuated, respectively. The $\left(n_{r}=1\right)$ constraint equation is obtained by normalizing the bottom row of matrix $\left(\widehat{\mathbf{B}}^{\prime}\right)^{-1}$ in (30) to obtain

$$
\mathbf{A U}=\left[1-\sqrt{\frac{M}{m}}\right]\left[\begin{array}{l}
u_{1}  \tag{31}\\
u_{2}
\end{array}\right]=u_{1}-\sqrt{\frac{M}{m}} u_{2}=0 .
$$

The constraint (31) may be applied to eliminate redundant modal control $u_{2}$ from the top row operation of (30) to obtain force $F_{a}$ as a function of independent modal control $u_{1}$, giving

$$
\begin{equation*}
F_{a}=\overline{\mathbf{B}} \mathbf{U}_{a}=(\sqrt{M+m}) u_{1} \tag{32}
\end{equation*}
$$

where $\overline{\mathbf{B}}=\sqrt{M+m}$ is the pseudotransfer matrix (since $n_{a}=1$, this matrix has only one term). Cart-driving force $F_{a}$ may be applied using open-loop control (as a known function of time) or using closed-loop control through a set of gains in full-state feedback. Both schemes will be analyzed and simulated using the CMSOC method.

The performance index (12) takes the following form, consisting of the gantry crane system's four states $\left(\eta_{1}, \eta_{2}, \dot{\eta}_{1}, \dot{\eta}_{2}\right)$ and two modal controls $\left(u_{1}, u_{2}\right)$ :

$$
\begin{align*}
J=\frac{1}{2} \int_{0}^{t_{f}} & \left(\widehat{Q}_{d 11} \eta_{1}^{2}+\widehat{Q}_{d 22} \eta_{2}^{2}+\widehat{Q}_{v 11} \dot{\eta}_{1}^{2}\right.  \tag{33}\\
& \left.+\widehat{Q}_{v 22} \dot{\eta}_{2}^{2}+\widehat{R}_{11} u_{1}^{2}+\widehat{R}_{22} u_{2}^{2}\right) d t \longrightarrow \min
\end{align*}
$$

The $n_{m}=2$ coupled optimality equations (18) take the following forms:

$$
\begin{gather*}
\widehat{R}_{11} \ddot{\eta}_{1}-\widehat{Q}_{v 11} \ddot{\eta}_{1}+\widehat{Q}_{d 11} \eta_{1}-\ddot{v}=0,  \tag{34a}\\
\widehat{R}_{22} \ddot{\eta}_{2}+\left(2 \widehat{R}_{22} \omega_{2}^{2}-\widehat{Q}_{v 22}\right) \ddot{\eta}_{2}+\left(\widehat{R}_{22} \omega_{2}^{4}+\widehat{Q}_{d 22}\right) \eta_{2} \\
+\sqrt{\frac{M}{m}}\left(\ddot{v}+\omega_{2}^{2} \nu\right)=0, \tag{34b}
\end{gather*}
$$

where $v$ is the Lagrange multiplier used to meet the $n_{r}=$ 1 constraint (31). The differential form (19) of constraint equation (31) is written as

$$
\begin{equation*}
\ddot{\eta}_{1}-\sqrt{\frac{M}{m}}\left(\ddot{\eta}_{2}+\omega_{2}^{2} \eta_{2}\right)=0 . \tag{35a}
\end{equation*}
$$

In modal space, the boundary conditions (27) are

$$
\begin{gather*}
\eta_{1}(0)=a \sqrt{M+m}, \quad \eta_{1}\left(t_{f}\right)=0, \\
\eta_{2}(0)=\dot{\eta}_{1}(0)=\dot{\eta}_{2}(0)=0,  \tag{36a}\\
\eta_{2}\left(t_{f}\right)=\dot{\eta}_{1}\left(t_{f}\right)=\dot{\eta}_{2}\left(t_{f}\right)=0 .
\end{gather*}
$$

Equations (34a), (34b), and (35a) written according to form (21) (with $D^{n}=d^{n} / d t^{n}$ ) yield

$$
\mathbf{E}_{\mathbf{p}} \mathbf{Y}=\left[\begin{array}{cc|c}
E_{1} & 0 & -\widehat{E}_{11}  \tag{37}\\
0 & E_{2} & -\widehat{E}_{21} \\
\hline \widetilde{E}_{11} & \widetilde{E}_{21} & 0
\end{array}\right]\left[\begin{array}{c}
\eta_{1} \\
\eta_{2} \\
\hline v
\end{array}\right]=0,
$$

where

$$
\begin{gather*}
E_{1}=\widehat{R}_{11} D^{4}-\widehat{Q}_{v 11} D^{2}+\widehat{Q}_{d 11} \\
E_{2}=\widehat{R}_{22} D^{4}+\left(2 \widehat{R}_{22} \omega_{2}^{2}-\widehat{Q}_{v 22}\right) D^{2}+\left(\widehat{R}_{22} \omega_{2}^{4}+\widehat{Q}_{d 22}\right) \\
\widehat{E}_{11}=\widetilde{E}_{11}=D^{2}  \tag{38}\\
\widehat{E}_{21}=\widetilde{E}_{21}=-\sqrt{\frac{M}{m}}\left(D^{2}+\omega_{2}^{2}\right)
\end{gather*}
$$

The characteristic equation of the system represented in (37) is obtained through operation (23), giving the 8th order polynomial equation:

$$
\begin{align*}
& \left.\operatorname{det} \mathbf{E}_{\mathbf{p}}\right|_{D \rightarrow r} \\
& =E_{1} \widehat{E}_{21}^{2}+\left.E_{2} \widehat{E}_{11}^{2}\right|_{D \rightarrow r} \\
& =\frac{M}{m}\left(\widehat{R}_{11} r^{4}-\widehat{Q}_{v 11} r^{2}+\widehat{Q}_{d 11}\right)\left(r^{2}+\omega_{2}^{2}\right)^{2}  \tag{39}\\
& \quad+r^{4}\left(\widehat{R}_{22} r^{4}+\left(2 \widehat{R}_{22} \omega_{2}^{2}-\widehat{Q}_{v 22}\right) r^{2}\right. \\
& \left.\quad \quad+\left(\widehat{R}_{22} \omega_{2}^{4}+\widehat{Q}_{d 22}\right)\right)=0 .
\end{align*}
$$

Eight roots may be obtained from the characteristic equation (39), which are then substituted into an appropriate assumed solution form (if the roots take the full complex form (24), then the assumed function takes form (25)) to characterize the three unknown solution functions $\left(\eta_{1}, \eta_{2}, \nu\right)$. This leaves twenty-four unknown integration constants to be determined by substituting the appropriate solution form into the three equations (34a), (34b), and (35a). By relating the coefficients corresponding to each of the eight independent elementary functions (i.e., in (25) each is in the form $e^{\left( \pm \alpha_{i} \pm \beta_{i}\right) t}$ ), eight algebraic equations are obtained for each differential equation in the set (34a), (34b) and (35a), resulting in a total of twentyfour equations in terms of twenty-four unknown integration constants $c_{k j}^{i}$. However, these twenty-four equations are linearly dependent. To obtain a unique solution, any one set of eight algebraic equations (obtained from either (34a), (34b), or (35a)) must be replaced with the eight boundary conditions (36a).

The optimal actuation forces needed to drive the gantry crane from an initially disturbed position $(x=a, \theta=0)$ to the origin $(x=0, \theta=0)$ will be derived for four cases using the CMSOC method. These cases are as follows
(A) an open-loop control that minimizes actuation forces for a fixed time interval as in [3];
(B) a closed-loop control that mimics the control presented in [1];
(C) a closed-loop control with response improved over that presented in [1];
(D) a closed-loop control of the fully actuated system (two actuators).

For each case, the gantry crane's physical parameters are chosen to match those given in [1]; namely, $M=m=1 \mathrm{~kg}$, $L=1 \mathrm{~m}, g=9.8 \mathrm{~m} / \mathrm{s}^{2}, a=-5 \mathrm{~m}$, and $\omega_{2}=4.43 \mathrm{~s}^{-1}$.

As a final case, the CMSOC method is applied to a modified three-DOF gantry crane, with an additional linkmass hinge attached to the existing model in Figure 1 and controlled by one or two actuators. This final case involves two subcases.
(E1) A closed-loop control that manoeuvres a modified gantry crane to the origin using the cart-driving force as well as a torque applied to the first rigid link (two actuators).
(E2) An open-loop control that manoeuvres the modified gantry crane to the origin using only the cart-driving force (one actuator) over a fixed time interval.
(A) Open-Loop Control of Gantry Crane Manoeuvre in a Finite Time Interval. The first control manoeuvres the gantry crane from a known initial position to the origin in a finite time interval $t_{f}$ in an open-loop scheme. The performance index is chosen to be consistent with that presented in [3], corresponding to the weightings $\widehat{R}_{11}=\widehat{R}_{22}=1$ in the general
form (33) with all other weightings null valued. Thus, the optimal control minimizes

$$
\begin{equation*}
J=\int_{0}^{t_{f}}\left(u_{1}^{2}+u_{2}^{2}\right) d t=\frac{1}{M} \int_{0}^{t_{f}} F_{a}^{2} d t \longrightarrow \min \tag{40}
\end{equation*}
$$

Performance index (40) minimizes the modal controls or the actuation force over the finite manoeuvre time $t_{f}$, which is chosen as $t_{f}=4 \mathrm{~s}$ to represent again one of the cases considered in [3]. The gantry crane's characteristic polynomial equation (39) is simplified to

$$
\begin{equation*}
\left(1+\frac{M}{m}\right)\left(r^{2}+\omega_{2}^{2}\right)^{2} r^{4}=0 \tag{41}
\end{equation*}
$$

The roots of (41) are $r_{1, \ldots, 8}=0,0,0,0, \pm i \omega_{2}, \pm i \omega_{2}$. There are four zero roots $r_{1, \ldots, 4}=0$, two imaginary roots $r_{5,7}=i \omega_{2}$, and two imaginary roots $r_{6,8}=-i \omega_{2}$. When written in form (24), these roots correspond to $\beta_{2}=\omega_{2}$ and $\alpha_{1}=\beta_{1}=\alpha_{2}=0$. Because of the zero roots and repeating roots, the solution functions take the following form:

$$
\begin{align*}
Y_{j}= & c_{1 j}+c_{2 j} t+c_{3 j} t^{2}+c_{4 j} t^{3}  \tag{42}\\
& +\left(c_{5 j}+c_{7 j} t\right) \sin \left(\omega_{2} t\right)+\left(c_{6 j}+c_{8 j} t\right) \cos \left(\omega_{2} t\right)
\end{align*}
$$

Each solution function (42) $(j=1, \ldots, 3)$ contains eight unknown integration constants $c_{k j}(k=1, \ldots, 8)$, which are determined through substitution and comparison of similar terms in any two differential equations in the set (34a), (34b), and (35a) and by substitution of the eight boundary conditions (36a). With the integration constants determined, the resulting solution functions are

$$
\begin{gather*}
\eta_{1}=-7.09-.104 t+1.41 t^{2}-.235 t^{3}  \tag{43}\\
\quad+.0235 \sin (4.43 t)+.0151 \cos (4.43 t) \\
\eta_{2}= \\
\quad .144-.0719 t+(.00451-.0333 t) \sin (4.43 t)  \tag{44}\\
+(-.144+.0520 t) \cos (4.43 t)  \tag{45}\\
v=-2.82+1.41 t-.460 \sin (4.43 t)-.295 \cos (4.43 t) .
\end{gather*}
$$

Substituting (43) into (29) yields

$$
\begin{equation*}
u_{1}=2.82-1.41 t-.460 \sin (4.43 t)-.295 \cos (4.43 t) \tag{46}
\end{equation*}
$$

The Lagrange multiplier function $v(t)(45)$, which represents a "modal force" enforcing the modal constraints, is not used in further analysis and is shown here only for completeness of the solution.

Mapping modal variables $\eta_{1}$ and $\eta_{2}((43)$ and (44)) into DOFs $x$ and $\theta$ via transformation (6a), the trajectories shown in Figures 2(a) and 2(b) are obtained. Optimal force $F_{a}$, shown in Figure 2(c), is obtained by substituting modal control $u_{1}$ (46) into transformation (32).

This phase of the solution was done automatically using MAPLE. The solution procedure accepts any problem with $n_{m}$ modes (obtained from FE analysis for more complex structures) controlled by $n_{a} \leq n_{m}$ actuators. The modal-toDOF transformations for the gantry crane are indicated in


Figure 2: Histograms of (a) cart trajectory $x$, (b) swing angle $\theta$, and (c) force $F_{a}$ (open loop).

Figures 2(a), 2(b), and 2(c). As shown, the open-loop control is able to perform the task in exactly four seconds, with a peak force of about 3.6 N and a maximum load swing angle of about $0.28 \mathrm{rad}\left(16^{\circ}\right)$. The optimal force accelerates the gantry crane over the first half of the manoeuvre ( 2 s ) and decelerates the cart over the last half with identical, but opposite and mirrored, forces.

Similar plots for the closed-loop control presented in [1] are shown in Figure 3. This control requires an effective manoeuvre time of $t_{f}^{3 \%} \approx 6 \mathrm{~s}$ to reach the origin, a maximum load rotation angle of 0.73 rad ( 42 deg ), and a maximum force of 15 N . It should be noted that this relatively large rotation angle is mentioned here (and other angles quoted in the sequel) for the purpose of comparison only.

From Figure 2 and Figure 3, one can conclude that the open-loop control performs the manoeuvre in a shorter period of time $\left(t_{f}=4 \mathrm{~s}\right.$ versus $\left.t_{f}^{3 \%} \approx 6 \mathrm{~s}\right)$ with much smaller peak force requirements ( 3.6 N versus 15 N ) and much smaller angles of oscillation ( $16^{\circ}$ versus $42^{\circ}$ ). Also, the open-loop control brings the system to a complete stop after 4 s , while the closed-loop control produces overshoot and the system takes longer to effectively come to rest.

Calculations show that if the finite manoeuvre time for the open-loop control is extended (or shortened), the peak force requirement and maximum swing angle are reduced (or increased)—approximately proportional to $t_{f}^{-2}$. For example, if the open-loop control is modified to settle over the same effective period of time as that of the closed-loop control $\left(t_{f} \approx\right.$ 6 ), the maximum force is reduced to approximately 1.6 N with a maximum swing of about 7 deg .

The open-loop control can always provide a faster and more efficient manoeuvre. However, it is possible only when the initial positions and manoeuvre times are known in advance. Closed-loop control is necessary if any initial configuration (unknown explicitly) is treated as disturbance, and its automatic reduction/removal is desired (the final position is at rest). Case (B) demonstrates how the CMSOC method is applied to analyze and simulate a closed-loop system that approximately produces the same dynamic responses as given in [1].
(B) Closed-Loop Control of Gantry Crane: Reproducing Control from [1]. A closed-loop control can perform the same task as that of the open-loop control (case (A)); however it does so automatically, without prior knowledge of initial conditions involved. Any disturbance from its resting configuration at the origin $(x=0, \theta=0)$ is relayed through a set of constant gains to generate the cart-driving force $F_{a}$ to attenuate this disturbance.

In general, to simulate the closed-loop process analytically the manoeuvre time $t_{f}$ is infinite and all parameters are driven asymptotically to the origin with increasing time. For the gantry crane, this requires that all roots of the characteristic equation (39) be nonzero complex numbers in the left half of the complex plane (unlike the open-loop system of case (A), which contained zero roots and purely imaginary roots). It can be verified that the weightings $\widehat{Q}_{d 11}$ and $\widehat{Q}_{d 22}$ in the performance index (33) must be nonzero in order to meet these criteria.

The gantry crane control as given in [1] is closely reproduced by choosing the weightings in the performance index (33) equal to $\widehat{Q}_{v 11}=\widehat{Q}_{v 22}=0, \widehat{Q}_{d 11}=4.5, \widehat{Q}_{d 22}=42$, and $\widehat{R}_{11}=\widehat{R}_{22}=1$. The resulting characteristic polynomial equation (39) has eight roots that take form (24), with real and complex parts equal to

$$
\begin{array}{ll}
\alpha_{1}=0.853, & \beta_{1}=0.856  \tag{47}\\
\alpha_{2}=0.513, & \beta_{2}=4.46
\end{array}
$$

Note that the first actively controlled mode of frequency $\omega_{1}^{a}=$ $0.856 \mathrm{~s}^{-1}\left(\omega_{1}=0\right.$ for uncontrolled system $)$ is damped with the ratio $\varsigma_{1}^{a}=0.996$, while the second mode of frequency $\omega_{2}^{a}=$ $4.46 \mathrm{~s}^{-1}\left(\omega_{2}=4.43 \mathrm{~s}^{-1}\right.$ for uncontrolled system $)$ is damped with the ratio $\varsigma_{1}^{a}=0.115$.

Similar to case (A), modal variables $\eta_{1}$ and $\eta_{2}$ are determined by substituting the parameters from (47) into the assumed solution function (25) and then solving for the unknown coefficients by comparing similar terms in two of the three optimality/constraint equations (34a), (34b) and (35a), and substituting the boundary conditions (36a). Unlike case (A), the closed-loop problem requires that only half


Figure 3: Histograms of (a) cart trajectory, (b) swing angle, and (c) optimal force from [1].
as many integration constants must be solved because the coefficients preceding exponential growth functions $\left(e^{\alpha_{i} t}\right)$ are assumed to be null valued. This gives

$$
\begin{align*}
\eta_{1}= & e^{-\alpha_{1} t}\left(-7.70 \sin \left(\beta_{1} t\right)-7.08 \cos \left(\beta_{1} t\right)\right) \\
& +e^{-\alpha_{2} t}\left(.126 \sin \left(\beta_{2} t\right)-.00564 \cos \left(\beta_{2} t\right)\right) \\
\eta_{2}= & e^{-\alpha_{1} t}\left(-.565 \sin \left(\beta_{1} t\right)+.534 \cos \left(\beta_{1} t\right)\right) \\
& +e^{-\alpha_{2} t}\left(.149 \sin \left(\beta_{2} t\right)-.534 \cos \left(\beta_{2} t\right)\right)  \tag{48}\\
u_{1}= & e^{-\alpha_{1} t}\left(-10.3 \sin \left(\beta_{1} t\right)+11.3 \cos \left(\beta_{1} t\right)\right) \\
& +e^{-\alpha_{2} t}\left(-2.44 \sin \left(\beta_{2} t\right)-.686 \cos \left(\beta_{2} t\right)\right)
\end{align*}
$$

Using the appropriate transformations (see Figure 2) the modal space variables (48) are mapped into the DOF space variables. The resulting system trajectories and the optimal force histogram are visually indistinguishable from those shown in Figure 3.

The CMSOC method can also generate the closed-loop gains from the assumed weighting coefficients to demonstrate
that the gains corresponding to the solution (48) are obtained and compared with the gains used in [1].

In full-state feedback control the active force is a function of all system states in the following form:

$$
\begin{equation*}
\mathbf{F}_{\mathbf{a}}=-\mathbf{G}_{\mathbf{d}} \mathbf{q}-\mathbf{G}_{\mathbf{v}} \dot{\mathbf{q}} . \tag{49}
\end{equation*}
$$

For the general CMSOC method, gains $\mathbf{G}_{\mathbf{d}}=\left[\begin{array}{lll}g_{1 d} & \cdots & g_{n_{m} d}\end{array}\right]$ and $\mathbf{G}_{\mathbf{v}}=\left[\begin{array}{lll}g_{1 v} & \cdots & g_{n_{m} v}\end{array}\right]$ correspond to the observed positions and velocities of all $n_{m}$ DOFs of a system. For the gantry crane, (49) takes the following form:

$$
\begin{equation*}
F_{a}=-g_{1 d} x-g_{2 d} L \theta-g_{1 v} \dot{x}-g_{2 v} L \dot{\theta} \tag{50}
\end{equation*}
$$

By substituting known DOF trajectories $\left(x=(1 / \sqrt{2})\left(\eta_{1}+\right.\right.$ $\eta_{2}$ ) and $\theta=\sqrt{2} \eta_{2}$ ) and the known force function $\left(F_{a}=\right.$ $\sqrt{2} u_{1}$ ) into (50) and grouping the terms related to the four independent elementary solution functions (operations are done in MAPLE automatically), one obtains:

$$
\begin{align*}
& e^{-\alpha_{1} t}\left\{\begin{array}{l}
\left(-14.6-5.85 g_{1 d}+8.95 g_{1 v}-.799 g_{2 d}+.0352 g_{2 v}\right) \sin \left(\beta_{1} t\right) \\
+\left(16.0-4.63 g_{1 d}-1.06 g_{1 v}+.755 g_{2 d}-1.33 g_{2 v}\right) \cos \left(\beta_{1} t\right)
\end{array}\right\} \\
& \quad+e^{-\alpha_{2} t}\left\{\begin{array}{c}
\left(-3.45+.195 g_{1 d}+1.56 g_{1 v}+.211 g_{2 d}+3.26 g_{2 v}\right) \sin \left(\beta_{2} t\right) \\
+\left(-.970-.374 g_{1 d}+1.06 g_{1 v}-.755 g_{2 d}+1.33 g_{2 v}\right) \cos \left(\beta_{2} t\right)
\end{array}\right\}=0 \tag{51}
\end{align*}
$$

Each of the bracketed terms in (51) (containing the unknown gains) must equal to zero for the equation to be true at any time. This gives four equations in terms of four unknown gains, which may be solved to obtain

$$
\begin{align*}
& \mathbf{G}_{\mathbf{d}}=\left[\begin{array}{ll}
g_{1 d} & g_{2 d}
\end{array}\right]=\left[\begin{array}{ll}
3.00 & .732
\end{array}\right] \\
& \mathbf{G}_{\mathbf{v}}=\left[\begin{array}{ll}
g_{1 v} & g_{2 v}
\end{array}\right]=\left[\begin{array}{ll}
3.66 & -.924
\end{array}\right] . \tag{52}
\end{align*}
$$

Though initial conditions were assumed in determining the trajectories $x$ and $\theta$ and force $F_{a}$, it can be verified that gains
(52) remain invariant towards any choice of these assumed conditions.

The control gains used in [1] were

$$
\begin{align*}
& \mathbf{G}_{\mathbf{d}}^{*}=\left[\begin{array}{ll}
g_{1 d} & g_{2 d}
\end{array}\right]=\left[\begin{array}{ll}
3.0 & .71
\end{array}\right] \\
& \mathbf{G}_{\mathbf{v}}^{*}=\left[\begin{array}{ll}
g_{1 d} & g_{2 d}
\end{array}\right]=\left[\begin{array}{ll}
3.69 & -.87
\end{array}\right] \tag{53}
\end{align*}
$$

Comparing the gains (52) and (53) confirms that the CMSOC method is able to closely reproduce the closed-loop control in [1] by careful selection of the weighting parameters in performance index (33). However, as shown next in case (C),

Table 1: Weightings for five different performance indices in form (33).

| Index \# | $\widehat{Q}_{d 11}$ | $\widehat{Q}_{d 22}$ | $\widehat{Q}_{v 11}$ | $\widehat{Q}_{v 22}$ | $\widehat{R}_{11}$ | $\widehat{R}_{22}$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| P0 | 0.01 | 0.01 | 0 | 0 | 1 | 1 |
| P1 | 6 | 0.01 | 0 | 0 | 1 | 1 |
| P2 | 6 | 50 | 0 | 0 | 1 | 1 |
| P3 | 6 | 50 | 4 | 0 | 1 | 1 |
| P4 | 6 | 50 | 4 | 50 | 1 | 1 |

the performance of this closed-loop control may be improved through better selection of these weighting parameters to produce faster convergence without an increase in the required peak actuation forces.
(C) Closed-Loop Control of Gantry Crane: Improving Performance. Case (B) developed a control that closely reproduced the control given in [1] by minimizing a performance index that gave no weight $\left(\widehat{Q}_{\nu 11}=\widehat{Q}_{\nu 22}=0\right)$ to states $\dot{\eta}_{1}$ and $\dot{\eta}_{2}$, representing the gantry crane's velocity. Consequently, the control caused the gantry to gain too much speed and then overshoot its target and produce large persistent load swings. These problems are mitigated by a more careful choice of the performance index weighting parameters in (33). To demonstrate the effect these parameters have on the gantry crane's dynamics and to illustrate how they might be meaningfully selected, several cases, labelled P1 to P5 (each with different performance indices as listed in Table 1), are considered.

Each case reflects a performance index which gives significant weightings to an incrementally increasing number of system states (of four possible states $\eta_{1}, \eta_{2}, \dot{\eta}_{1}, \dot{\eta}_{2}$ ), while holding the weighting on both modal controls $\left(u_{1}, u_{2}\right)$ at unity. Case P0 gives none of the states a significant weighting, case Pl gives a significant weighting to a single state $\left(\eta_{1}\right)$, case P2 gives significant weightings to two states $\left(\eta_{1}, \eta_{2}\right)$, and so on until case P4 significantly weights all four states. Table 1 summarizes how these weightings are chosen for each case. Since the gantry crane's asymptotic convergence mathematically requires that weightings $\widehat{Q}_{d 11}$ and $\widehat{Q}_{d 22}$ in the index (33) are nonzero, a small value (0.01) is used instead of zero in cases P0 and P1 to demonstrate how the system behaves when these weightings are negligible. The DOF trajectories $(x$ and $\theta)$ and force histogram $\left(F_{a}\right)$ for the manoeuvres minimizing the performance indices for cases P0-P4 are presented in Table 2. The settling times $t_{f}^{3 \%}$ of the DOFs are also listed for each case. All plots in Table 2 are shown over the first 8 s of the manoeuvre period except for P0 (30 s).

Note that the first modal variable $\eta_{1}$ primarily influences the cart's rigid body mode of motion, while the second modal variable $\eta_{2}$ influences the suspended load rotation. In fact there is a direct relationship between the angle of the load rotation and the second modal variable $\left(\theta=\sqrt{2} \eta_{2}\right)$ such that this angular trajectory is directly affected by varying the weights given to $\eta_{2}\left(\widehat{Q}_{d 22}\right)$ and its derivative $\dot{\eta}_{2}\left(\widehat{Q}_{v 22}\right)$ in the performance index (33). Likewise, the speed at which the
cart can be made to reach its target is affected through the weightings given to $\eta_{1}\left(\widehat{Q}_{d 11}\right)$ and its derivative $\dot{\eta}_{1}\left(\widehat{Q}_{v 11}\right)$.

The performance index in case P0 heavily weights the modal controls $u_{1}$ and $u_{2}$ in comparison to modal variables $\eta_{1}$ and $\eta_{2}$ ( 100 times more) and neglects the modal velocities $\dot{\eta}_{1}$ and $\dot{\eta}_{2}$. The resulting control requires a small peak force $(0.7 \mathrm{~N})$, producing small maximum load swing angles ( 0.06 rad or 3.4 deg ), but requires a very long manoeuvre time to converge to the origin ( $t_{f}^{3 \%} \approx 440 \mathrm{~s}$ ). If the weightings $\widehat{R}_{11}$, and $\widehat{R}_{22}$ were increased even further relative to the weighting $\widehat{Q}_{d 11}$, the maximum force requirements and angular rotations would become infinitesimal while the settling times would approach infinity.

In case Pl a significant weighting value is given to the first modal variable $\eta_{1}\left(\widehat{Q}_{d 11}=6\right)$, while other weightings remain unchanged from case P0. This control greatly increases the speed at which the cart reaches its target position at $x=0(\sim$ 2 s ), but upon reaching this position the load undergoes large swing angles ( 1.0 rad or 57 deg ) that persist for a very long time ( $t_{f}^{3 \%} \approx 440 \mathrm{~s}$ ).

The maximum force increases significantly ( 17.3 N ) in comparison to case P0 because the rigid body cart motion requires much larger accelerations during the initial 2 s of the manoeuvre in order to quickly attenuate $\eta_{1}$ due to its significant weighting value.

Case P2 improves the load swing attenuation, which was poorly dampened in case P1, by including a large weighting value to the second modal variable $\eta_{2}\left(\widehat{Q}_{d 22}=50\right)$ (other weightings remain the same as in the previous case). The maximum load swing angle is reduced ( 0.8 rad or 46 deg ) and the load swinging motion is damped much more quickly ( $\sim 6.5 \mathrm{~s}$ ). The cart translation requires similar accelerations and thus approximately the same maximum force $(17.3 \mathrm{~N})$ is needed. By inspection, one can see that case P2 produces similar behaviour to the control given in [1] shown in Figure 3 and likewise shares the problem of target overshoot and large persistent load swings.

Case P3 reduces the tendency of the cart to overshoot the target by also giving a significant weighting to the first modal velocity $\dot{\eta}_{1}\left(\widehat{Q}_{v 11}=4\right)$. However, large persistent load swings are still present, and so convergence is not significantly improved over that produced in case P2. The maximum required force ( 17.3 N ) remains essentially unchanged, while the load swing angles are reduced slightly ( 0.75 rad or 43 deg ).

The performance index in case P 4 includes a large weight on the second modal velocity $\dot{\eta}_{2}\left(\widehat{Q}_{\nu 22}=50\right)$, while keeping all other weightings unchanged from case P3. This produces a control that reduces the magnitude of load swing angles ( 0.45 rad or 25.8 deg ) while attenuating the swinging motion more quickly ( $t_{f}^{3 \%} \approx 4.2 \mathrm{~s}$ ). The gantry crane performs the manoeuvre in essentially a single load swing cycle, with similar initial cart accelerations and thus maximum forces ( 17.3 N ) as in previous cases. Case P4 produces faster convergence then previous cases because, from an optimal control perspective, it incorporates all of the gantry crane's states in the minimization by assigning all weightings in the performance index (33) with significant numerical values.

TABLE 2: DOF responses and force histograms that optimize performance indices $\mathrm{P} 0-\mathrm{P} 4$.
Index

It is essential that the second modal velocity $\dot{\eta}_{2}$ is given a significant weighting value to yield fast convergence, because the energy of the suspended load oscillates equally between potential and kinetic energy. Since potential energy and kinetic energy are proportional to the squares of displacement and velocity, respectively, both of the corresponding states $\eta_{2}$ and $\dot{\eta}_{2}$ should carry a significant and approximately equal weight in the performance index (33). Without
weighting the load swing velocity state, the control focuses on eliminating swing angles but not swing velocities. However, when the load is near the bottom of its swing, its velocity is near maximum ( $\dot{\eta}_{2}=\sqrt{2} \dot{\theta} \rightarrow$ max $)$, while its displacement is near minimum $\left(\eta_{2}=\sqrt{2} \theta \rightarrow \mathrm{~min}\right)$. Therefore, the optimal force derived without consideration for the load swing velocity is unable to eliminate any significant portion of the load swing energy when the load is near the bottom of its
swing $(\theta \rightarrow 0)$. Cases P2 and P3 failed to adequately weight $\dot{\eta}_{2}$, resulting in larger, more persistent load swings than in case P4.

The control produced in case P4 provides a significant improvement over the control presented in [1], as it converges more quickly to the origin, while reducing load swing magnitudes, without any increase in the required maximum forces. To complete the design of this closed-loop control, the gains are obtained from (50) in a similar fashion as in case (B), giving

$$
\begin{align*}
& \mathbf{G}_{\mathbf{d}}=\left[\begin{array}{ll}
g_{1 d} & g_{2 d}
\end{array}\right]=\left[\begin{array}{ll}
3.46 & 9.10
\end{array}\right] \\
& \mathbf{G}_{\mathbf{v}}=\left[\begin{array}{ll}
g_{1 v} & g_{2 v}
\end{array}\right]=\left[\begin{array}{ll}
5.43 & 1.79
\end{array}\right] \tag{54}
\end{align*}
$$

Note that the gains $g_{1 d}$ and $g_{1 v}$ are somewhat close to the gains for the control presented in [1] (53), but gains $g_{2 d}$ and $g_{2 v}$ are substantially different.
(D) Closed-Loop Control of Gantry Crane: Fully Actuated Control. The CMSOC method can also be applied to fully actuated systems. To illustrate this, consider the same gantry crane system, now with both actuators $F_{a}$ and $F_{d}$ acting as real actuators (no dummy actuator). This situation may arise practically when a person is employed to guide the suspended load while the cart performs its translations.

Since the problem is fully actuated there, are no additional constraints on the system motion and consequently no Lagrange multipliers needed to enforce them. The optimal forces can be solved by calculating the inverse dynamics directly from (6b), which takes form (30) (except $F_{d} \neq 0$ ), written as

$$
\mathbf{F}=\widehat{\mathbf{B}}^{-1} \mathbf{U} \Longrightarrow\left[\begin{array}{l}
F_{a}  \tag{55}\\
F_{d}
\end{array}\right]=\left[\begin{array}{cc}
\frac{M}{\sqrt{M+m}} & \sqrt{\frac{M m}{M+m}} \\
\frac{m}{\sqrt{M+m}} & -\sqrt{\frac{M m}{M+m}}
\end{array}\right]\left[\begin{array}{l}
u_{1} \\
u_{2}
\end{array}\right] .
$$

The optimality equations in the differential operator form (21) become

$$
\mathbf{E}_{\mathbf{p}} \mathbf{Y}=[\mathbf{E}][\boldsymbol{\eta}]=\mathbf{0} \Longrightarrow\left[\begin{array}{cc}
E_{1} & 0  \tag{56}\\
0 & E_{2}
\end{array}\right]\left[\begin{array}{l}
\eta_{1} \\
\eta_{2}
\end{array}\right]=0
$$

where

$$
\begin{gather*}
E_{1}=\widehat{R}_{11} D^{4}-\widehat{Q}_{v 11} D^{2}+\widehat{Q}_{d 11} \\
E_{2}=\widehat{R}_{22} D^{4}+\left(2 \widehat{R}_{22} \omega_{2}^{2}-\widehat{Q}_{v 22}\right) D^{2}+\left(\widehat{R}_{22} \omega_{2}^{4}+\widehat{Q}_{d 22}\right) \tag{57}
\end{gather*}
$$

With weightings chosen according to the performance index in case $P 4\left(\widehat{Q}_{d 11}=6, \widehat{Q}_{d 22}=50, \widehat{Q}_{v 11}=4, \widehat{Q}_{v 22}=50\right.$, and $\widehat{R}_{11}=\widehat{R}_{22}=1$ ), the roots of the characteristic equation (39) for the system given by $(56)\left(\left.E_{1} E_{2}\right|_{D \rightarrow r}=0\right)$ take form (24) with the following real and imaginary parts:

$$
\begin{array}{ll}
\alpha_{1}=3.62, & \beta_{1}=2.78 \\
\alpha_{2}=1.49, & \beta_{2}=0.474 \tag{58}
\end{array}
$$

For any fully actuated system, each modal variable $\eta_{i}$ is independently controlled by a single modal control $u_{i}$, resulting in uncoupled solution functions of the following form:

$$
\begin{equation*}
\eta_{i}=e^{\alpha_{i} t}\left(c_{i}^{1} \sin \left(\beta_{i} t\right)+c_{i}^{2} \cos \left(\beta_{i} t\right)\right) \tag{59}
\end{equation*}
$$

For the gantry crane $(i=1,2)$ the four unknown integration constants $c_{i}^{1,2}$ are obtained by substituting the four initial conditions for $\eta_{i}(0)$ and $\dot{\eta}_{i}(0)$ given by (36a). As in the previous cases, the solved modal variables in form (59) are mapped into the original coordinates to obtain the DOF trajectories and optimal forces. Figure 4 shows the cart trajectory $x$ and the optimal forces on the cart and suspended load $F_{a}$ and $F_{d}$, respectively. The angular trajectory $\theta$ of the load is not shown because it remains zero $(\theta=0)$ all the time. Practically, this means that for the optimal manoeuvre the person (actuator) guiding the suspended load must simply act to prevent it from swinging. Fast convergence $\left(t_{f}^{3 \%}=0.78 \mathrm{~s}\right)$ to the origin is obtained; however the task requires relatively large maximum forces ( 104 N ) compared to previous cases. The required actions of cart-driving force $F_{a}$ and suspended load guiding force $F_{d}$ are identical, as the whole gantry crane system moves as a single rigid body. If smaller forces are desired, then a larger weight may be given to the modal controls ( $\widehat{R}_{11}, \widehat{R}_{22}$ ) in the performance index. Note that only a 30 N maximum force would be required to execute the manoeuvre in 1 s by applying an open-loop control scheme.
(E1) Closed-Loop Control of Modified Three-DOF Gantry Crane (Two Actuators). To illustrate the application of the CMSOC method to a problem of a higher dimension $\left(n_{m}=\right.$ 3 ), the gantry crane is modified by adding an additional link with an end load, as shown in Figure 5 (a case of a doublependulum gantry crane in 3D was presented in [19]). In comparison to the previous cases considered the control task is unchanged except that the oscillations of the additional suspended load must also be damped. Consider the control that uses two actuators ( $n_{a}=2$ )-the standard cart-driving force $F_{a}$ and torque $T_{a}$, produced by a motor fixed to the cart and applied to the first rigid link of length $L_{1}$ which supports the mass $m_{1}$. Dummy torque $T_{d}$ to be used in formulating the augmented system is applied to the second link of length $L_{2}$ which carries mass $m_{2}$. All other physical variables are the same as in the original gantry crane model with the exception of $\theta_{1}$ and $\theta_{2}$, which denote the angles of links of lengths $L_{1}$ and $L_{2}$, respectively.

The standard matrices in the augmented system's equation of motion (1) for this new model are

$$
\begin{gathered}
\mathbf{M}=\left[\begin{array}{ccc}
M+m_{1}+m_{2} & -m_{1}-m_{2} & -m_{2} \\
-m_{1}-m_{2} & m_{1}+m_{2} & m_{2} \\
-m_{2} & m_{2} & m_{2}
\end{array}\right], \\
\mathbf{K}=\left[\begin{array}{ccc}
0 & 0 & 0 \\
0 & \frac{g\left(m_{1}+m_{2}\right)}{L_{1}} & 0 \\
0 & 0 & \frac{g m_{2}}{L_{2}}
\end{array}\right],
\end{gathered}
$$



Figure 4: Histograms of (a) cart trajectory and (b) forces for the fully actuated gantry crane.


FIgure 5: Modified three-DOF gantry crane model.

$$
\mathbf{B}^{\prime}=\left[\begin{array}{ccc}
1 & 0 & 0  \tag{60}\\
0 & \frac{1}{L_{1}} & -\frac{1}{L_{1}} \\
0 & 0 & \frac{1}{L_{2}}
\end{array}\right]
$$

following numerical values are adopted: $M=m_{1}=m_{2}=$ $1 \mathrm{~kg}, L_{1}=L_{2}=1 \mathrm{~m}, a=-5 \mathrm{~m}$, and $g=9.8 \mathrm{~m} / \mathrm{s}^{2}$.

The set of $n_{m}=3$ equations of motion (1) are uncoupled in modal space, with matrices of ordered frequencies $\boldsymbol{\Omega}$ and mode shapes $\varphi$ normalized according to (7a) and (7b), taking the following forms:

The augmented system consists of DOF vector $\mathbf{q}=$ $\left[\begin{array}{lll}x & L_{1} \theta_{1} & L_{2} \\ \theta_{2}\end{array}\right]^{T}$ and force vector $F_{a}^{\prime}=\left[\begin{array}{lll}F_{a} & T_{a} & T_{d}\end{array}\right]^{T}$. The

$$
\boldsymbol{\Omega}=\left[\begin{array}{ccc}
0 & 0 & 0  \tag{61}\\
0 & 12.4 & 0 \\
0 & 0 & 46.4
\end{array}\right], \quad \boldsymbol{\Phi}=\left[\begin{array}{ccc}
.577 & .577 & .577 \\
0 & .366 & 1.37 \\
0 & 1 & -1
\end{array}\right] .
$$

As before, the first mode represents the rigid body mode of motion $\left(\omega_{1}^{2}=0\right)$, while the second and third modes, with the squared frequencies $\omega_{2}^{2}=12.43(\mathrm{rad} / \mathrm{s})^{2}$ and $\omega_{3}^{2}=$ $46.37(\mathrm{rad} / \mathrm{s})^{2}$, represent the swinging modes of the rotating link-masses.

Augmented force vector $F_{a}^{\prime}$ is related to modal control vector $\mathbf{U}$ through the inverse of transformation (6b) which is partitioned according to (9) to give

$$
\widehat{\mathbf{B}}^{-1} \mathbf{U}=\left[\begin{array}{cc|c}
1.73 & 0 & 0  \tag{62}\\
-1.73 & 1.37 & .366 \\
\hline-.577 & .789 & -.211
\end{array}\right]\left[\begin{array}{c}
u_{1} \\
u_{2} \\
\hline u_{3}
\end{array}\right]=\left[\begin{array}{c}
F_{a} \\
T_{a} \\
\hline T_{d}
\end{array}\right] .
$$

The constraint equation $\left(n_{r}=1\right)$ is obtained from the bottom row of (62) $\left(T_{d}=0\right)$ and normalized into the following form:

$$
\mathbf{A} \mathbf{U}=\left[\begin{array}{lll}
1 & -1.37 & .366
\end{array}\right]\left[\begin{array}{l}
u_{1}  \tag{63}\\
u_{2} \\
u_{3}
\end{array}\right]=0
$$

The $n_{a}=2$ actuation forces may be obtained directly from the top two rows of (62) in terms of all modal controls, but according to (11) these forces may be expressed in terms of two independent modal controls (chosen as $u_{1}$ and $u_{2}$ ) in the following form:

$$
\overline{\mathbf{B}} \mathbf{U}_{\mathbf{a}}=\left[\begin{array}{cc}
1.73 & 0  \tag{64}\\
-2.73 & 2.73
\end{array}\right]\left[\begin{array}{l}
u_{1} \\
u_{2}
\end{array}\right]=\left[\begin{array}{l}
F_{a} \\
T_{a}
\end{array}\right] .
$$

Selecting a performance index of form (12) gives three ( $n_{m}=$ 3 ) optimality equations in the form (18) that, with the constraint equation (63), may be written according to (21) in the following form:

$$
\mathbf{E}_{\mathbf{p}} \mathbf{Y}=\left[\begin{array}{ccc|c}
E_{1} & 0 & 0 & \widehat{E}_{11}  \tag{65}\\
0 & E_{2} & 0 & \widehat{E}_{21} \\
0 & 0 & E_{3} & \widehat{E}_{31} \\
\hline \widetilde{E}_{11} & \widetilde{E}_{21} & \widetilde{E}_{31} & 0
\end{array}\right]\left[\begin{array}{c}
\eta_{1} \\
\eta_{2} \\
\eta_{3} \\
\hline v
\end{array}\right]=0
$$

where

$$
\begin{gather*}
E_{i}=\widehat{R}_{i i} D^{4}-\left(2 \widehat{R}_{i i} \omega_{i}^{2}-\widehat{Q}_{v i i}\right) D^{2}+\left(\widehat{R}_{i i} \omega_{i}^{4}+\widehat{Q}_{d i i}\right), \\
\widehat{E}_{i j}=A_{j i}\left(D^{2}+\omega_{i}^{2}\right) \quad(i=1, \ldots, 3, j=1) . \tag{66}
\end{gather*}
$$

The parameters $A_{j i}$ in the equation above are the $j$ th row and $i$ th column components of the constraint matrix $\mathbf{A}$ given by (63). The selected weightings for the performance index are $\widehat{Q}_{d 11}=6, \widehat{Q}_{d 22}=\widehat{Q}_{d 33}=50, \widehat{Q}_{v 11}=4, \widehat{Q}_{v 22}=\widehat{Q}_{v 33}=50$, and $R_{11}=R_{22}=R_{33}=1$.

The twelve $\left(4 n_{m}\right)$ roots of the characteristic equation (23) are obtained in the complex form (24) and are used to generate an assumed solution of form (25) for each unknown modal variable ( $n_{m}=3$ ) and Lagrange multiplier ( $n_{r}=1$ ). Half of these roots generate exponential growth functions that are eliminated by assuming their corresponding integration constants to be zero-valued. Then through the method of undetermined coefficients, four $\left(n_{m}+n_{r}\right)$ sets of six linear algebraic equations are obtained. Replacing one set by the set
of six initial conditions, the unknown integration constants are obtained by solving the set of twenty-four equations (the number of equations is $2 n_{m}\left(n_{m}+n_{r}\right)$ ).

The boundary conditions for this problem are the same as those chosen for the original gantry crane, written in (36a), with the additional condition that the initial and final positions and velocities of the third modal variable $\eta_{3}$ are also zero. In other words, the manoeuvre requires a horizontal cart translation from a resting position at $x=a=-5 \mathrm{~m}$ with both links hanging vertically to the same resting position at the origin.

Figure 6 shows trajectories $x, \theta_{1}$, and $\theta_{2}$ of the three-DOF gantry crane as well as required actuation forces $F_{a}$ and $T_{a}$.

The manoeuvre, requiring a maximum force of 29 N and a maximum torque of 18 Nm , is effectively completed after $t_{\mathrm{eff}}^{3 \%}=4 \mathrm{~s}$. The maximum load swing angle of the first link is $0.11 \mathrm{rad}\left(6.3^{\circ}\right)$ and that of the second link is $0.48 \mathrm{rad}(27.5 \mathrm{deg})$.

## (E2) Open-Loop Control of Modified Three-DOF Gantry Crane

 (One Actuator). In order to show the case where one actuator controls three DOFs, the optimal manoeuvre for the modified gantry crane using only a single actuator-cart-driving force $F_{a}$-is investigated for an open-loop scheme. Both of the torque actuators $T_{a}$ and $T_{d}$ (Figure 5) are treated as dummy actuators and so the inverse transformation, while identical to (62), is repartitioned in the following form:$$
\widehat{\mathbf{B}}^{-1} \mathbf{U}=\left[\begin{array}{cc|c}
1.73 & 0 & 0  \tag{67}\\
\hline-1.73 & 1.37 & .366 \\
-.5774 & .789 & -.211
\end{array}\right]\left[\begin{array}{c}
u_{1} \\
\hline u_{2} \\
u_{3}
\end{array}\right]=\left[\begin{array}{c}
F_{a} \\
\hline T_{a} \\
T_{d}
\end{array}\right]=\left[\begin{array}{c}
F_{a} \\
\hline 0 \\
0
\end{array}\right] .
$$

The constraint equations $\left(n_{r}=2\right)$ are obtained from the two bottom rows of (67) and normalized into the following form:

$$
\mathbf{A} \mathbf{U}=\left[\begin{array}{ccc}
1 & -1 & 0  \tag{68}\\
0 & 1 & 1
\end{array}\right]\left[\begin{array}{l}
u_{1} \\
u_{2} \\
u_{3}
\end{array}\right]=0
$$

According to (11) the single cart-driving force may be expressed in terms of the independent modal control (chosen as $u_{1}$ ) in the following form:

$$
\begin{equation*}
\overline{\mathbf{B}} \mathbf{U}_{\mathbf{a}}=[1.73]\left[u_{1}\right]=\left[F_{a}\right] . \tag{69}
\end{equation*}
$$

Note that matrix $\overline{\mathbf{B}}$ is the same for all cases involving one actuator. Choosing a performance index in form (12) gives three $\left(n_{m}=3\right)$ optimality equations in form (18) that, with constraint equation (68), may be written according to (21) in the following form:

$$
\mathbf{E}_{\mathbf{p}} \mathbf{Y}=\left[\begin{array}{ccc|cc}
E_{1} & 0 & 0 & \widehat{E}_{11} \widehat{E}_{12}  \tag{70}\\
0 & E_{2} & 0 & \widehat{E}_{21} & \widehat{E}_{22} \\
0 & 0 & E_{3} & \widehat{E}_{31} & \widehat{E}_{32} \\
\hline \widehat{E}_{11} & \widehat{E}_{21} & \widehat{E}_{31} & 0 & 0 \\
\widehat{E}_{12} & \widehat{E}_{22} & \widehat{E}_{32} & 0 & 0
\end{array}\right]\left[\begin{array}{c}
\eta_{1} \\
\eta_{2} \\
\eta_{3} \\
\frac{v_{1}}{v_{2}}
\end{array}\right]=0,
$$

where

$$
\begin{gather*}
E_{i}=\widehat{R}_{i i} D^{4}-\left(2 \widehat{R}_{i i} \omega_{i}^{2}-\widehat{Q}_{v i i}\right) D^{2}+\left(\widehat{R}_{i i} \omega_{i}^{4}+\widehat{Q}_{d i i}\right),  \tag{71}\\
\widehat{E}_{i j}=A_{j i}\left(D^{2}+\omega_{i}^{2}\right) \quad(i=1, \ldots, 3, j=2) .
\end{gather*}
$$



Figure 6: (a) Cart trajectory, (b) swing angles, and (c) force/torque for the modified gantry crane.


Figure 7: (a) Cart trajectory, (b) swing angles, and (c) force for modified gantry crane (open loop).

The assumed weightings are $R_{11}=R_{22}=R_{33}=1$ and $\widehat{Q}_{d 11}=\widehat{Q}_{d 22}=\widehat{Q}_{d 33}=\widehat{Q}_{v 11}=\widehat{Q}_{v 22}=\widehat{Q}_{v 33}=0$ (only the control effort is to be minimized). Consistent with the openloop control presented in case (A) the finite manoeuvre time is chosen to be $t_{f}=4 \mathrm{~s}$.

The solution procedure is similar to previous examples. Figure 7 shows trajectories $x, \theta_{1}$, and $\theta_{2}$ as well as the required cart-driving force $F_{a}$. The manoeuvre requires a peak force of 4.8 N and completes the task in exactly 4 s . The maximum load swing angle of the first link is $0.19 \mathrm{rad}\left(11^{\circ}\right)$ and that of the second link is $0.35 \mathrm{rad}\left(20^{\circ}\right)$.

## 4. Conclusions

The CMSOC methodology was presented as a means of solving linear underactuated (or fully actuated) control problems. The gantry crane problem was selected to illustrate in detail various operations required for different control methodologies. As demonstrated the method can be applied to open-loop control schemes as well as closed-loop (asymptotically convergent) control schemes. In the latter case the weightings of the performance index can be translated to the gains of the full-state feedback closed-loop controllers. The operations would be identical for any similar problems with larger numbers of modes and actuators.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

[1] I. Fantoni and R. Lozano, Non-Linear Control for Underactuated Mechanical Systems, Springer, 2002.
[2] L. Meirovitch and H. Baruh, "Control of self-adjoint distributed-parameter systems," Journal of Guidance, Control, and Dynamics, vol. 5, no. 1, pp. 60-66, 1982.
[3] B. L. Karihaloo and R. D. Parbery, "Optimal control of a dynamical system representing a gantry crane," Journal of Optimization Theory and Applications, vol. 36, no. 3, pp. 409417, 1982.
[4] M. W. Spong, "The swing up control problem for the acrobot," IEEE Control Systems Magazine, vol. 15, no. 1, pp. 49-55, 1995.
[5] R. A. Canfield and L. Meirovitch, "Integrated structural design and vibration suppression using independent modal space control," AIAA Journal, vol. 32, no. 10, pp. 2053-2060, 1994.
[6] R. Seifried, "Two approaches for feedforward control and optimal design of underactuated multibody systems," Multibody System Dynamics, vol. 27, no. 1, pp. 75-93, 2012.
[7] R. Seifried and W. Blajer, "Analysis of servo-constraint problems for underactuated multibody systems," Mechanical Sciences, vol. 4, pp. 113-129, 2013.
[8] W. Blajer and K. Kołodziejczyk, "Control of underactuated mechanical systems with servo-constraints," Nonlinear Dynamics, vol. 50, no. 4, pp. 781-791, 2007.
[9] W. Blajer and K. Kołodziejczyk, "A geometric approach to solving problems of control constraints: theory and a DAE framework," Multibody System Dynamics, vol. 11, no. 4, pp. 343364, 2004.
[10] M. Fliess, J. Lévine, P. Martin, and P. Rouchon, "Flatness and defect of non-linear systems: introductory theory and examples," International Journal of Control, vol. 61, no. 6, pp. 1327-1361, 1995.
[11] H. Sira-Ramirez and S. K. Agrawal, Differentially Flat Systems, Control Engineering Series, Marcel Dekker, 2004.
[12] K. -J. Bathe, "Modal superposition," in Finite Elements Procedures, chapter 9.3, Prentice Hall, Englewood Cliffs, NJ, USA, 1996.
[13] S. Woods and W. Szyszkowski, "Simulating active vibration attenuation in underactuated spatial structures," Journal of Guidance, Control, and Dynamics, vol. 32, no. 4, pp. 1246-1253, 2009.
[14] S. Woods and W. Szyszkowski, "Analysis and simulation of optimal vibration attenuation for underactuated mechanical systems," AIAA Journal, vol. 47, no. 12, pp. 2821-2835, 2009.
[15] D. E. Kirk, Optimal Control Theory, Dover, New York, NY, USA, 1998.
[16] D. G. Zill and M. R. Cullen, Advanced Engineering Mathematics, PWS Publishing Company, Boston, Mass, USA, 1992.
[17] E. M. Abdel-Rahman, A. H. Nayfeh, and Z. N. Masoud, "Dynamics and control of cranes: a review," Journal of Vibration and Control, vol. 9, no. 7, pp. 863-908, 2003.
[18] H.-H. Lee, "Modeling and control of a three-dimensional overhead crane," Journal of Dynamic Systems, Measurement and Control, vol. 120, no. 4, pp. 471-476, 1998.
[19] W. O. Connor and H. Habibi, "Gantry crane control of a double-pendulum, distributed-mass load, using mechanical wave concept," Mechanical Sciences, vol. 4, pp. 251-261, 2013.

## Research Article

# Exponential $L_{2}-L_{\infty}$ Filtering for a Class of Stochastic System with Mixed Delays and Nonlinear Perturbations 

Zhaohui Chen ${ }^{1,2}$ and Qi Huang ${ }^{1}$<br>${ }^{1}$ Sichuan Provincial key Lab of Power System Wide-Area Measurement and Control, University of Electronic Science and Technology of China, Chengdu 611731, China<br>${ }^{2}$ School of Mathematics and Physics, Chongqing University of Science and Technology, Chongqing 401331, China<br>Correspondence should be addressed to Qi Huang; hwong@uestc.edu.cn

Received 12 August 2013; Revised 21 November 2013; Accepted 28 November 2013
Academic Editor: Weihai Zhang
Copyright © 2013 Z. Chen and Q. Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The delay-dependent exponential $L_{2}-L_{\infty}$ performance analysis and filter design are investigated for stochastic systems with mixed delays and nonlinear perturbations. Based on the delay partitioning and integral partitioning technique, an improved delaydependent sufficient condition for the existence of the $L_{2}-L_{\infty}$ filter is established, by choosing an appropriate Lyapunov-Krasovskii functional and constructing a new integral inequality. The full-order filter design approaches are obtained in terms of linear matrix inequalities (LMIs). By solving the LMIs and using matrix decomposition, the desired filter gains can be obtained, which ensure that the filter error system is exponentially stable with a prescribed $L_{2}-L_{\infty}$ performance $\gamma$. Numerical examples are provided to illustrate the effectiveness and significant improvement of the proposed method.


## 1. Introduction

Time delays are quite often encountered in various practical engineering systems, and they are regarded as one of the main sources causing instability and degrading performance of control systems [1-3]. Over the past decades, numerous results and various approaches on delay systems have been reported in the literatures. Many researchers have focused on the stability analysis, stabilization, and filtering for time-delay systems; see [4-9] and the references therein. Time delays are usually classified into discrete delays and distributed delays. In the existing literatures, discrete time-delay system [10-12], distributed time-delay system [13, 14], and mixed (including both discrete and distributed time delays) system [15-17] are considered.

Since certain unavoidable stochastic perturbations are widely existing in many engineering systems, stochastic systems have gained considerable research attention over the past few years [18-20]. Stochastic dynamic modeling has come to play an important role in many fields of science and
engineering. In the past years, many researchers have focused on the problems of stability and stabilization of stochastic time-delay systems. For instance, robust stabilization for a class of large-scale stochastic systems was investigated in [21], delay-dependent stability results for stochastic systems were presented in [22-26], and $H_{\infty}$ state feedback control and $H_{\infty}$ dynamic output feedback control for uncertain stochastic time-delay systems were investigated in [27, 28], respectively.

In the field of stochastic dynamic system with time delays, the filtering problem, which is to estimate the unavailable state of variables of a given control system, is also an important issue. Kalman filtering scheme is a well-known effective way to deal with the filtering problem. However, it has some limitations in practical applications due to the fact that it assumes that the system and its disturbances are exactly known, that is, stationary Gaussian noised with known statistics. Under this view, recently, $H_{\infty}$ filtering, mixed $H_{2} / H_{\infty}$ filtering and $L_{2}-L_{\infty}$ filtering for stochastic time-delay systems have been widely studied [ $8,9,29-38]$. In $H_{2} / H_{\infty}$ filtering, and $L_{2}-L_{\infty}$ filtering problems, the external disturbances are
assumed to be bounded. In $H_{2} / H_{\infty}$ filtering problem, it requires that the filtering error systems satisfy not only a prescribed $H_{\infty}$ disturbance attenuation level but also the $H_{2}$ performance (minimum of the $\mathrm{H}_{2}$ norm of transfer function of the filter error systems), while in $L_{2}-L_{\infty}$ filtering problem, it requires that the filtering error systems satisfy a prescribed $L_{2}-L_{\infty}$ disturbance attenuation level. $H_{\infty}$ filtering and mixed $\mathrm{H}_{2} / H_{\infty}$ filtering problems of nonlinear stochastic systems are investigated in [30, 31]. In [32], a delay-independent robust $L_{2}-L_{\infty}$ filtering design approach for uncertain stochastic time-delay system is investigated. It is well known that the delay-independent results are generally more conservative than the delay-dependent ones. Authors in [33-35] developed delay-dependent filtering for stochastic time-delay systems. Authors in [36] proposed a delay-dependent $L_{2}-L_{\infty}$ filter design approach for stochastic time-delay systems, based on a delay partitioning technique presented in [37]. As the results showed, delay-partitioning can reduce conservatism to some extent. Authors in [38] investigated the problem of robust $L_{2}-L_{\infty}$ filtering for stochastic systems with both discrete and distributed delays. Although the filtering problems for stochastic systems with time delays have been well investigated in the aforementioned literatures, most of them are dealing with linear stochastic time-delay systems. To the authors' knowledge, the $L_{2}-L_{\infty}$ filtering problems of stochastic time-delay systems with nonlinear perturbation are still insufficient. This motivates the authors to deal with the $L_{2}-L_{\infty}$ filtering problem of a class of nonlinear stochastic time-delay systems.

This paper focuses on the problems of delay-dependent $L_{2}-L_{\infty}$ filtering for stochastic systems with mixed delays and nonlinear perturbations. By Lyapunov-Krasovskii approach based on the delay partitioning and integral partitioning technique, we first develop a delay-dependent sufficient condition for $L_{2}-L_{\infty}$ performance analysis. And then, an improved delay-dependent sufficient condition is obtained for the existence of desired filter in the form of linear matrix inequalities (LMIs). The $L_{2}-L_{\infty}$ performance analysis and filter design of linear stochastic system with mixed delays are also investigated. Finally, numerical examples are provided to show that the proposed method is effective and less conservative than some existing literatures.

Notations. Throughout this paper, $X>0(X<0)$ means that the matrix $X$ is positive definite (negative definite). $\mathbf{R}^{n}$ denotes the $n$-dimensional Euclidean space; $\mathbf{R}^{m \times n}$ is the set of all $m \times n$ real matrices; $\mathbf{L}_{2}[0, \infty)$ is the space of square-integrable vector functions over $[0, \infty)$. The superscript " $T$ " represents the transpose; "*" denotes the symmetric terms in a matrix; $\operatorname{diag}()$ denotes a block-diagonal matrix; $\lambda_{\max }()$ and $\lambda_{\text {min }}$ () denote the maximum eigenvalue and minimum eigenvalue, respectively. $\operatorname{sym}(X)=X+X^{T} ;|\cdot|$ denotes the Euclidean vector norm; $\|\cdot\|_{2}$ stands for the usual $\mathbf{L}_{2}[0, \infty)$ norm. $(\boldsymbol{\Omega}, \mathbf{F}, \mathbf{P})$ is a probability space with $\Omega$ the sample space, $\mathbf{F}$ the $\sigma$-algebra of subsets of $\boldsymbol{\Omega}$, and $\mathbf{P}$ the probability measure on $\mathbf{F} . \mathbf{E}\{\cdot\}$ denotes the expectation operator with respect to some probability measure $\mathbf{P} . \mathbf{0}$ and $\mathbf{I}$ represent zero matrix and identity matrix with appropriate dimensions, respectively, unless we say otherwise.

## 2. Problem Formulation

Consider the following stochastic systems with mixed delays and nonlinear perturbations:

$$
\begin{align*}
& d x(t) \\
& =\left[A x(t)+A_{1} x(t-h)\right. \\
& \left.+A_{2} \int_{t-d}^{t} x(s) d s+A_{3} f(x(t), x(t-h), t)+A_{v} v(t)\right] d t \\
& +g(x(t), x(t-h), t) d \omega(t), \\
& d y(t)=\left[C x(t)+C_{1} x(t-h)+C_{2} \int_{t-d}^{t} x(s) d s+C_{v} v(t)\right] d t \\
& \quad z(t)=L x(t), \\
& x(t)=\varphi(t), \quad \forall t \in[-\tau, 0] \tag{1}
\end{align*}
$$

where $x(t) \in \mathbf{R}^{n}$ is the state; $y(t) \in \mathbf{R}^{m}$ is the measured output; $z(t) \in \mathbf{R}^{p}$ is the signal to be estimated; $v(t) \in \mathbf{R}^{q}$ is the disturbance input which belongs to $\mathbf{L}_{2}[0, \infty)$, which is the space of square-integrable vector functions; $\omega(t)$ is a onedimensional Brownian motion defined on a complete probability space $(\boldsymbol{\Omega}, \mathbf{F}, \mathbf{P})$ satisfying $\mathbf{E}\{d \omega(t)\}=0$ and $\mathbf{E}\left\{d \omega^{2}(t)\right\}=$ $d t ; \varphi(t)$ is an initial function that is continuous on $[-\tau, 0]$ with $\tau=\max \{h, d\} . h$ and $d$ are discrete and distributed constant delays, respectively. $f(\cdot, \cdot, \cdot): \mathbf{R}^{n} \times \mathbf{R}^{n} \times \mathbf{R} \rightarrow \mathbf{R}^{n}$ is a nonlinear function, which satisfies

$$
\begin{equation*}
|f(x, y, t)|^{2} \leq\left|F_{1} x\right|^{2}+\left|F_{2} y\right|^{2}, \quad f(0,0,0)=0 \tag{2}
\end{equation*}
$$

where $F_{1} \in \mathbf{R}^{n \times n}$ and $F_{2} \in \mathbf{R}^{n \times n}$ are known constant matrices; $g(\cdot, \cdot, \cdot): \mathbf{R}^{n} \times \mathbf{R}^{n} \times \mathbf{R} \rightarrow \mathbf{R}^{n}$ is a nonlinear perturbance input function, satisfying

$$
\begin{equation*}
|g(x, y, t)|^{2} \leq\left|G_{1} x\right|^{2}+\left|G_{2} y\right|^{2}, \quad g(0,0,0)=0 \tag{3}
\end{equation*}
$$

where $G_{1} \in \mathbf{R}^{n \times n}$ and $G_{2} \in \mathbf{R}^{n \times n}$ are known constant matrices.
For system (1), we are interested in constructing the following full-order linear filter:

$$
\begin{gather*}
d x_{f}(t)=A_{f} x_{f}(t) d t+B_{f} d y(t)  \tag{4}\\
z_{f}(t)=C_{f} x_{f}(t)
\end{gather*}
$$

where $x_{f}(t) \in \mathbf{R}^{n}$ is the filter state; $A_{f}, B_{f}$, and $C_{f}$ are filter matrices to be determined.

Define

$$
\begin{equation*}
\xi^{T}(t)=\left[x^{T}(t), x_{f}^{T}(t)\right]^{T}, \quad e(t)=z(t)-z_{f}(t) \tag{5}
\end{equation*}
$$

Then, the filtering error system can be written as

$$
\begin{align*}
& d \xi(t)=\left[\bar{A} \xi(t)+\bar{A}_{1} H \xi(t-h)+\bar{A}_{2} H \int_{t-d}^{t} \xi(s) d s\right. \\
& \left.\quad+\bar{A}_{3} \bar{f}(\xi(t), \xi(t-h), t)+\bar{A}_{v} v(t)\right] d t  \tag{6}\\
& +\bar{g}(\xi(t), \xi(t-h), t) d \omega(t) \\
& e(t)=\bar{L} \xi(t)
\end{align*}
$$

where

$$
\begin{gather*}
\bar{A}=\left[\begin{array}{cc}
A & 0 \\
B_{f} C & A_{f}
\end{array}\right], \quad \bar{A}_{1}=\left[\begin{array}{c}
A_{1} \\
B_{f} C_{1}
\end{array}\right], \\
\bar{A}_{2}=\left[\begin{array}{c}
A_{2} \\
B_{f} C_{2}
\end{array}\right], \quad \bar{A}_{3}=\left[\begin{array}{cc}
A_{3} & 0 \\
0 & 0
\end{array}\right], \\
\bar{A}_{v}=\left[\begin{array}{c}
A_{v} \\
B_{f} C_{v}
\end{array}\right], \quad H=\left[\begin{array}{ll}
I_{n} & 0_{n}
\end{array}\right],  \tag{7}\\
\bar{f}(\xi(t), \xi(t-h), t)=\left[\begin{array}{c}
f(x(t), x(t-h), t) \\
0
\end{array}\right], \\
\bar{g}(\xi(t), \xi(t-h), t)=\left[\begin{array}{c}
g(x(t), x(t-h), t) \\
0
\end{array}\right], \\
\bar{L}=\left[\begin{array}{ll}
\left.-C_{f}\right] .
\end{array}\right.
\end{gather*}
$$

The objective of this paper is to design full-order $L_{2}-L_{\infty}$ filter (4) for the stochastic time-delay system (1) such that the filtering error system (6) satisfies the following two requirements:
(i) the filtering error system (6) with $v(t)=0$ is exponentially stable [39];
(ii) under the zero initial condition, the filtering error system (6) is stochastically asymptotically stable and achieves a prescribed $L_{2}-L_{\infty}$ attenuation level $\gamma$. The filtering error $e(t)$ satisfies

$$
\begin{equation*}
\|e(t)\|_{E_{\infty}}<\gamma\|v(t)\|_{2} \tag{8}
\end{equation*}
$$

with $\|e(t)\|_{E_{\infty}}=\sup _{t} \sqrt{\mathbf{E}\left\{|e(t)|^{2}\right\}},\|v(t)\|_{2}=\sqrt{\int_{0}^{\infty} v^{T}(t) v(t) d t}$ for any nonzero $v(t) \in \mathbf{L}_{2}[0, \infty]$.

Before presenting the main results of this paper, we introduce the following lemmas, which will be essential to our derivation.

Lemma 1 (see [40]). For a given symmetrical matrix $S=$ $\left(\begin{array}{cc}S_{11}^{1} \\ S_{12}^{T} & S_{22}\end{array}\right)$, where $S_{11}=S_{11}^{T}$, and $S_{22}=S_{22}^{T}$, the linear matrix inequality $S<0$ is equivalent to

$$
\begin{gathered}
S_{11}<0, \quad S_{22}-S_{12} S_{11}^{-1} S_{12}^{T}<0 \\
\text { or } \\
S_{22}<0, \quad S_{11}-S_{12} S_{22}^{-1} S_{12}^{T}<0
\end{gathered}
$$

Lemma 2 (see [1]). For any positive symmetric matrix $W \in$ $\mathbf{R}^{n \times n}$, scalars $\delta_{1}$ and $\delta_{2}$ satisfying $\delta_{1}<\delta_{2}$, a vector function $x:\left[\delta_{1}, \delta_{2}\right] \rightarrow \mathbf{R}^{n}$, one has

$$
\begin{align*}
& \int_{\delta_{1}}^{\delta_{2}} x^{T}(s) W x(s) d s \\
& \quad \geq \frac{1}{\left(\delta_{2}-\delta_{1}\right)}\left(\int_{\delta_{1}}^{\delta_{2}} x(s) d s\right)^{T} W\left(\int_{\delta_{1}}^{\delta_{2}} x(s) d s\right) \tag{10}
\end{align*}
$$

Lemma 3 (see [14]). For any positive symmetric matrix $W \in$ $\mathbf{R}^{n \times n}$, scalars $a$ and $b$ satisfying $a<b \leq 0$, a vector function $x:[a, b] \rightarrow \mathbf{R}^{n}$, one has

$$
\begin{align*}
& \int_{a}^{b} \int_{t+\lambda}^{t} x^{T}(s) W x(s) d s d \lambda \\
& \quad \geq \frac{2}{a^{2}-b^{2}}\left(\int_{a}^{b} \int_{t+\lambda}^{t} x(s) d s d \lambda\right)^{T} W\left(\int_{a}^{b} \int_{t+\lambda}^{t} x(s) d s d \lambda\right) \tag{11}
\end{align*}
$$

## 3. Filtering Performance Analysis

In this section, a new delay-dependent condition of the $L_{2}-L_{\infty}$ filtering performance analysis for system (1) will be presented. A Lyapunov-Krasovskii functional is constructed; based on the idea of delay partitioning and integral partitioning, the conservatism will be reduced. For the convenience of expression, assume that the filter matrices $\left(A_{f}, B_{f}\right.$, and $\left.C_{f}\right)$ are known.

Theorem 4. Consider the stochastic time-delay system (1). For given scalars $\gamma>0, h>0, d>0, \rho>0$, and $\varepsilon>0$ and integers $r_{1} \geq 1$ and $r_{2} \geq 1$, there exists a linear filter (4) such that the filtering error system (6) is stochastically asymptotically stable with a guaranteed $L_{2}-L_{\infty}$ performance $\gamma$, if there exist symmetrical positive definite matrices $P \in \mathbf{R}^{2 n \times 2 n}, Q_{i} \in \mathbf{R}^{n \times n}$, $R_{i} \in \mathbf{R}^{n \times n}\left(i=1,2, \ldots, r_{1}\right), W_{j} \in \mathbf{R}^{n \times n}$, and $Z_{j} \in \mathbf{R}^{n \times n}(j=$ $1,2, \ldots, r_{2}$ ) and matrix $M \in \mathbf{R}^{n \times n}$ satisfying

$$
\Phi=\left[\begin{array}{cccccc}
\Phi_{11} & \Phi_{12} & \Phi_{13} & P \bar{A}_{3} & P \bar{A}_{v} & \bar{A}^{T} H^{T} M  \tag{12}\\
* & \Phi_{22} & 0 & 0 & 0 & \Phi_{26} \\
* & * & \Phi_{33} & 0 & 0 & \Phi_{36} \\
* & * & * & -\varepsilon I & 0 & \bar{A}_{3}^{T} H^{T} M \\
* & * & * & * & -I & \bar{A}_{v}^{T} H^{T} M \\
* & * & * & * & * & \Phi_{66}
\end{array}\right]<0
$$

$$
\begin{equation*}
P \leq \rho I \tag{13}
\end{equation*}
$$

$$
\Gamma=\left[\begin{array}{cc}
P & \bar{L}^{T}  \tag{14}\\
* & \gamma^{2} I
\end{array}\right]>0
$$

where

$$
\begin{align*}
& \Phi_{11}=P \bar{A}+\bar{A}^{T} P+H^{T} Q_{1} H-H^{T} R_{1} H \\
& -\sum_{j=1}^{r_{2}} H^{T}\left(\frac{2}{2 j-1} Z_{j}\right) H \\
& +\left(\frac{d}{r_{2}}\right)^{2} \sum_{j=1}^{r_{2}} H^{T} W_{j} H+\rho H^{T} G_{1}^{T} G_{1} H+\varepsilon H^{T} F_{1}^{T} F_{1} H, \\
& \Phi_{12}=H^{T} R_{1} K+P \bar{A}_{1} K_{r_{1}}, \\
& \Phi_{13}=P \bar{A}_{2} K_{r_{2}}+H^{T} \overline{\mathrm{Z}}, \\
& \bar{Z}=\frac{2 m}{d}\left[\begin{array}{llll}
Z_{1} & \frac{1}{3} Z_{2} & \cdots & \frac{1}{2 r_{2}-1} Z_{r_{2}}
\end{array}\right], \\
& \Phi_{22}=\left[\begin{array}{ccccc}
Q_{2}-Q_{1} & R_{2} & \cdots & 0 & 0 \\
-R_{2}-R_{1} & Q_{3}-Q_{2} & \ldots & 0 & 0 \\
* & -R_{3}-R_{2} & & & \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
* & * & \cdots & Q_{r_{1}}-Q_{r_{1}-1} & R_{r_{1}} \\
* & * & \cdots & * & R_{r_{1}}-R_{r_{1}-1}
\end{array}\right] \text {, } \\
& \bar{\Phi}_{22}=-Q_{r_{1}}-R_{r_{1}}+\rho G_{2}^{T} G_{2}+\varepsilon F_{2}^{T} F_{2}, \\
& \Phi_{26}=K_{r_{1}}^{T} \bar{A}_{1}^{T} H^{T} M \text {, } \\
& \Phi_{33}=\operatorname{diag}\left(-W_{1}-\frac{2 m^{2}}{d^{2}} Z_{1},-W_{2}-\frac{2 m^{2}}{3 d^{2}} Z_{2}, \ldots,-W_{r_{2}}\right. \\
& \left.-\frac{2 m^{2}}{\left(2 r_{2}-1\right) d^{2}} Z_{r_{2}}\right), \\
& \Phi_{36}=K_{r_{2}}^{T} \bar{A}_{2}^{T} H^{T} M, \\
& \Phi_{66}=\left(\frac{h}{r_{1}}\right)^{2} \sum_{i=1}^{r_{1}} R_{i}+\left(\frac{d}{r_{2}}\right)^{2} \sum_{j=1}^{r_{2}} \frac{2 j-1}{2} Z_{j}-M-M^{T}, \\
& K=\left[\begin{array}{ll}
I_{n} & 0_{n \times\left(r_{1}-1\right) n}
\end{array}\right], \\
& K_{r_{1}}=\left[\begin{array}{ll}
0_{n \times\left(r_{1}-1\right) n} & I_{n}
\end{array}\right] \text {, } \\
& K_{r_{2}}=\left[\begin{array}{llll}
I & I & \cdots & I
\end{array}\right]_{n \times r_{2} n} . \tag{15}
\end{align*}
$$

Proof. First, show the asymptotic stability of system (6) with $v(t)=0$. For simplicity of notations, rewrite the filtering error system (6) as

$$
\begin{equation*}
d \xi(t)=u(t) d t+\pi(t) d \omega(t) \tag{16}
\end{equation*}
$$

where

$$
\begin{gathered}
u(t):=\bar{A} \xi(t)+\bar{A}_{1} H \xi(t-h)+\bar{A}_{2} H \int_{t-d}^{t} \xi(s) d s \\
+\bar{A}_{3} \bar{f}(\xi(t), \xi(t-h), t)+\bar{A}_{v} v(t) \\
\pi(t):=\bar{g}(\xi(t), \xi(t-h), t)
\end{gathered}
$$

Next, denote $\eta(t) d t=d \xi(t)$, and choose the following Lyapunov-Krasovskii functional:
$V\left(\xi_{t}, t\right)$
$=\xi^{T}(t) P \xi(t)+\sum_{i=1}^{r_{1}} \int_{t-\left(i / r_{1}\right) h}^{t-\left((i-1) / r_{1}\right) h} \xi^{T}(s) H^{T} Q_{i} H \xi(s) d s$ $+\sum_{i=1}^{r_{1}} \frac{h}{r_{1}} \int_{-\left(i / r_{1}\right) h}^{-\left((i-1) / r_{1}\right) h} \int_{t+\theta}^{t} \eta^{T}(s) H^{T} R_{i} H \eta(s) d s d \theta$ $+\sum_{j=1}^{r_{2}} \frac{d}{r_{2}} \int_{-\left(j / r_{2}\right) d}^{-\left((j-1) / r_{2}\right) d} \int_{t+\theta}^{t} \xi^{T}(s) H^{T} W_{j} H \xi(s) d s d \theta$ $+\sum_{j=1}^{r_{2}} \int_{-\left(j / r_{2}\right) d}^{-\left((j-1) / r_{2}\right) d} \int_{\theta}^{0} \int_{t+\beta}^{t} \eta^{T}(s) H^{T} Z_{j} H \eta(s) d s d \beta d \theta$.

Then, by Itô differential formula, the stochastic differential along the trajectories of system (6) is

$$
\begin{equation*}
d V\left(\xi_{t}, t\right)=\mathbf{L} V\left(\xi_{t}\right) d t+2 \xi(t) P \pi(t) d \omega(t) \tag{19}
\end{equation*}
$$

where

$$
\begin{align*}
\mathbf{L} V\left(\xi_{t}, t\right)= & 2 \xi^{T}(t) P u(t)+\operatorname{trace}\left(\pi^{T}(t) P \pi(t)\right) \\
& +\sum_{i=1}^{r_{1}} \xi^{T}\left(t-\frac{i-1}{r_{1}} h\right) H^{T} Q_{i} H \xi\left(t-\frac{i-1}{r_{1}} h\right) \\
& -\sum_{i=1}^{r_{1}} \xi^{T}\left(t-\frac{i}{r_{1}} h\right) H^{T} Q_{i} H \xi\left(t-\frac{i}{r_{1}} h\right) \\
& +\left(\frac{h}{r_{1}}\right)^{2} \sum_{i=1}^{r_{1}} \eta^{T}(t) H^{T} R_{i} H \eta(t) \\
& -\sum_{i=1}^{r_{1}} \frac{h}{r_{1}} \int_{t-\left(i / r_{1}\right) h}^{t-\left((i-1) / r_{1}\right) h} \eta^{T}(s) H^{T} R_{i} H \eta(s) d s \\
& +\left(\frac{d}{r_{2}}\right)^{2} \sum_{j=1}^{r_{2}} \xi^{T}(t) H^{T} W_{i} H \xi(t) \\
& -\sum_{j=1}^{r_{2}} \frac{d}{r_{2}} \int_{t-\left(j / r_{2}\right) d}^{t-\left((j-1) / r_{2}\right) d} \xi^{T}(s) H^{T} W_{j} H \xi(s) d s \\
& +\left(\frac{d}{r_{2}}\right)^{2} \sum_{j=1}^{r_{2}} \frac{2 j-1}{2} \eta^{T}(t) H^{T} Z_{j} H \eta(t) \\
& -\sum_{j=1}^{r_{2}} \int_{-\left(j / r_{2}\right) d}^{-\left((j-1) / r_{2}\right) d} \int_{t+\theta}^{t} \eta^{T}(t) H^{T} Z_{j} H \eta(t) d s d \theta . \tag{20}
\end{align*}
$$

By Lemma 2, we have

$$
\begin{align*}
& -\sum_{i=1}^{r_{1}} \frac{h}{r_{1}} \int_{t-\left(i / r_{1}\right) h}^{t-\left((i-1) / r_{1}\right) h} \eta^{T}(s) H^{T} R_{i} H \eta(s) d s \\
& \leq-\sum_{i=1}^{r_{1}}\left(\int_{t-\left(i / r_{1}\right) h}^{t-\left((i-1) / r_{1}\right) h} \eta(s) d s\right)^{T} \\
& \quad \times H^{T} R_{i} H\left(\int_{t-\left(i / r_{1}\right) h}^{t-\left((i-1) / r_{1}\right) h} \eta(s) d s\right)  \tag{23}\\
& =-\sum_{i=1}^{r_{1}}\left(\xi\left(t-\frac{i-1}{r_{1}} h\right)-\xi\left(t-\frac{i}{r_{1}} h\right)\right)^{T}  \tag{24}\\
& \quad \times H^{T} R_{i} H\left(\xi\left(t-\frac{i-1}{r_{1}} h\right)-\xi\left(t-\frac{i}{r_{1}} h\right)\right) \tag{21}
\end{align*}
$$

$$
\begin{align*}
& -\sum_{j=1}^{r_{2}} \frac{d}{r_{2}} \int_{t-\left(j / r_{2}\right) d}^{t-\left((j-1) / r_{2}\right) d} \xi^{T}(s) H^{T} W_{j} H \xi(s) d s  \tag{25}\\
& \leq-\sum_{j=1}^{r_{2}}\left(\int_{t-\left(j / r_{2}\right) d}^{t-\left((j-1) / r_{2}\right) d} \xi(s) d s\right)^{T}  \tag{22}\\
& \quad \quad \times H^{T} W_{j} H\left(\int_{t-\left(j / r_{2}\right) d}^{t-\left((j-1) / r_{2}\right) d} \xi(s) d s\right)
\end{align*}
$$

By Lemma 3, we have

$$
\begin{align*}
& -\sum_{j=1}^{r_{2}} \int_{-\left(j / r_{2}\right) d}^{-\left((j-1) / r_{2}\right) d} \int_{t+\theta}^{t} \eta^{T}(t) H^{T} Z_{j} H \eta(t) d s d \theta \\
& \quad \leq-\sum_{j=1}^{r_{2}} \frac{2}{2 j-1}\left(\frac{r_{2}}{d}\right)^{2}\left(\int_{-\left(j / r_{2}\right) d}^{-\left((j-1) / r_{2}\right) d} \int_{t+\theta}^{t} \eta(s) d s d \theta\right)^{T} \tag{27}
\end{align*}
$$

that

$$
\begin{aligned}
& \xi^{T}(t) \varepsilon H^{T} F_{1}^{T} F_{1} H \xi(t) \\
& \quad+\xi^{T}(t-h) \varepsilon H^{T} F_{2}^{T} F_{2} H \xi(t-h)-\varepsilon \bar{f}^{T} \bar{f} \geq 0
\end{aligned}
$$

$$
\begin{gathered}
\times H^{T} Z_{j} H\left(\int_{-\left(j / r_{2}\right) d}^{-\left((j-1) / r_{2}\right) d} \int_{t+\theta}^{t} \eta(s) d s d \theta\right) \\
=-\sum_{j=1}^{r_{2}} \frac{2}{2 j-1}\left(\frac{r_{2}}{d}\right)^{2}\left(\frac{d}{r_{2}} \xi(t)-\int_{t-\left(j / r_{2}\right) d}^{t-\left((j-1) / r_{2}\right) d} \xi(s) d s\right)^{T} \\
\times H^{T} Z_{j} H\left(\frac{d}{r_{2}} \xi(t)-\int_{t-\left(j / r_{2}\right) d}^{t-\left((j-1) / r_{2}\right) d} \xi(s) d s\right) .
\end{gathered}
$$

From (16), for any appropriately dimensioned matrix $M$, we have

$$
2 \eta^{T}(t) H^{T} M^{T} H[u(t) d t+\pi(t) d \omega(t)-\eta(t) d t]=0 .
$$

On the other hand, (2) implies that there exists $\varepsilon>0$ such
where we take $\bar{f}$ for $\bar{f}(x(t), x(t-h), t)$, for simplicity of notation.

Notice the fact of (3), and from (13), we have

$$
\begin{align*}
& \operatorname{trace}\left(\pi^{T}(t) P \pi(t)\right) \\
& \leq \xi^{T}(t) \rho H^{T} G_{1}^{T} G_{1} H \xi(t)+\xi^{T}(t-h) \rho H^{T} G_{2}^{T} G_{2} H \xi(t-h) . \tag{26}
\end{align*}
$$

Combine (20)-(26); then

$$
\mathbf{L} V\left(x_{t}, t\right) \leq \zeta^{T}(t) \bar{\Phi} \zeta(t)+2 \eta^{T}(t) H^{T} M^{T} H \pi(t) d \omega(t)
$$

where

$$
\begin{aligned}
& \zeta^{T}(t)=\left[\begin{array}{lll}
\xi_{p 1}^{T}(t) & \left.\xi_{p 2}^{T}(t) \bar{f}(t) \quad \eta^{T}(t) H^{T}\right], ~
\end{array}\right. \\
& \xi_{p 1}^{T}=\left[\xi^{T}(t) \xi^{T}\left(t-\frac{1}{r_{1}} h\right) H^{T} \quad \xi^{T}\left(t-\frac{2}{r_{1}} h\right) H^{T} \ldots \xi^{T}(t-h) H^{T}\right], \\
& \xi_{p 2}^{T}(t)=\left[\int_{t-\left(1 / r_{2}\right) d}^{t} \xi^{T}(s) H^{T} d s \int_{t-\left(2 / r_{2}\right) d}^{t-\left(1 / r_{2}\right) d} \xi^{T}(s) H^{T} d s \cdots \int_{t-d}^{t-\left(\left(r_{2}-1\right) / r_{2}\right) d} \xi^{T}(s) H^{T} d s\right], \\
& \bar{\Phi}=\left[\begin{array}{ccccc}
\Phi_{11} & \Phi_{12} & \Phi_{13} & P \bar{A}_{3} & \bar{A}^{T} H^{T} M \\
* & \Phi_{22} & 0 & 0 & \Phi_{26} \\
* & * & \Phi_{33} & 0 & \Phi_{36} \\
* & * & * & -\varepsilon I & \bar{A}_{3}^{T} H^{T} M \\
* & * & * & * & \Phi_{66}
\end{array}\right] .
\end{aligned}
$$

Thus,

$$
\begin{equation*}
\mathbf{E}\left\{\mathbf{L} V\left(\xi_{t}, t\right)\right\} \leq \mathbf{E}\left\{\zeta^{T}(t) \bar{\Phi} \zeta(t)\right\} . \tag{29}
\end{equation*}
$$

By Schur complement lemma, it is easy to show that $\Phi<0$ implies $\bar{\Phi}<0$. Combined with (29), these imply that, for any $\zeta(t) \neq 0$, we have $\mathbf{E}\left\{\mathbf{L} V\left(\xi_{t}, t\right)\right\}<0$.

By Dynkin's formula, there exists $\beta>0$, such that

$$
\begin{equation*}
e^{\beta t} \mathbf{E} V\left(\xi_{t}, t\right) \leq \mathbf{E} V\left(\xi_{0}, 0\right) \tag{30}
\end{equation*}
$$

Recalling the Lyapunov-Krasovskii functional in (18), notice the fact that there always exists $\kappa>0$ satisfying

$$
\begin{equation*}
|\eta(t)|^{2} \leq \kappa|\xi(t)|^{2} \tag{31}
\end{equation*}
$$

for any $-\tau \leq t \leq 0$ such that

$$
\begin{equation*}
\mathbf{E} V\left(\xi_{0}, 0\right) \leq \sum_{i=1}^{5} \alpha_{i} \sup _{-\tau \leq s \leq 0} \mathbf{E}|\xi(s)|^{2} \tag{32}
\end{equation*}
$$

where $\alpha_{1}=\lambda_{\max }(P), \alpha_{2}=h \max _{i}\left\{\left\|Q_{i}\right\|\right\}, \alpha_{3}=$ $\left(\kappa h^{3} / 2 r_{1}\right) \max _{i}\left\{\left\|R_{i}\right\|\right\}, \alpha_{4}=\left(d^{3} / 2 r_{2}\right) \max _{j}\left\{\left\|W_{j}\right\|\right\}$, and $\alpha_{5}=$ $\left(\kappa d^{3} / 6\right) \max _{j}\left\{\left\|Z_{j}\right\|\right\}$.

On the other hand, from (18)

$$
\begin{equation*}
\mathbf{E} V\left(\xi_{t}, t\right) \geq \lambda_{\min }(P) \mathbf{E}|\xi(t)|^{2} \tag{33}
\end{equation*}
$$

From (32) and (33), it can be easily obtained that

$$
\begin{equation*}
\mathbf{E}|\xi(t ; \widetilde{\varphi})|^{2} \leq \alpha e^{-\beta t} \sup _{-\tau \leq s \leq 0} \mathbf{E}|\xi(s)|^{2} \tag{34}
\end{equation*}
$$

where $\alpha=\sum_{i=1}^{5} \alpha_{i} / \lambda_{\text {min }}(P)$ and $\widetilde{\varphi}$ is the initial condition of filtering error system (6). Then by exponential stability definition of stochastic systems [39], the filtering error system (6) with $v(t)=0$ is exponentially stable in the sense of mean square.

Now, we will establish the $L_{2}-L_{\infty}$ performance for the filtering error system (6). To this end, we assume the zero initial condition $\zeta(t)=0$ for $t \in[-\tau, 0]$. Under the initial condition, it is easy to see that, for any $t>0$,

$$
\begin{equation*}
\mathbf{E}\left\{V\left(\xi_{t}, t\right)\right\}=\mathbf{E}\left\{\int_{0}^{t} \mathbf{L} V\left(\xi_{s}, s\right) d s\right\} \tag{35}
\end{equation*}
$$

Define

$$
\begin{equation*}
J(t)=\mathbf{E}\left\{V\left(\xi_{t}, t\right)\right\}-\int_{0}^{t} v^{T}(s) v(s) d s \tag{36}
\end{equation*}
$$

Then, for any nonzero $v(t) \in \mathbf{L}_{2}[0, \infty)$ and $t>0$, combined with (29), (35)-(36), we have

$$
\begin{align*}
J(t) & =\mathbf{E}\left\{\int_{0}^{t}\left[\mathbf{L} V\left(\xi_{s}, s\right)-v^{T}(s) v(s)\right] d s\right\}  \tag{37}\\
& \leq \mathbf{E}\left\{\int_{0}^{t} \mathcal{\vartheta}^{T}(s) \Phi \mathcal{\vartheta}(s) d s\right\},
\end{align*}
$$

where $\mathfrak{\vartheta}^{T}(t)=\left[\xi_{p 1}^{T}(t) \xi_{p 2}^{T}(t) \bar{f}^{T} v^{T}(t) \eta^{T}(t) H^{T}\right] . \Phi<0$ ensuring that $J(t) \leq 0$. Thus,

$$
\begin{equation*}
\mathbf{E}\left\{\xi^{T}(t) P \xi(t)\right\} \leq \mathbf{E}\left\{V\left(\xi_{t}, t\right)\right\} \leq \int_{0}^{t} v^{T}(s) v(s) d s \tag{38}
\end{equation*}
$$

Moreover, by Schur complement, (14) holds if and only if

$$
\begin{equation*}
\bar{L}^{T} \bar{L}<\gamma^{2} P \tag{39}
\end{equation*}
$$

It follows from (38) and (39) that

$$
\begin{align*}
\mathbf{E}\left\{|e(t)|^{2}\right\} & =\mathbf{E}\left\{\xi^{T}(t) \bar{L}^{T} \bar{L} \xi(t)\right\}<\gamma^{2} \mathbf{E}\left\{\xi^{T}(t) P \xi(t)\right\} \\
& \leq \gamma^{2} \mathbf{E}\left\{V\left(\xi_{t}, t\right)\right\} \leq \gamma^{2} \int_{0}^{t} v^{T}(s) v(s) d s \tag{40}
\end{align*}
$$

Therefore, if (12)-(14) hold, the filtering error system (6) is mean-square exponentially stable with a prescribed $L_{2}-L_{\infty}$ performance $\gamma$ under zero initial condition. This completes the proof.

In system (1), if $A_{3}=\mathbf{0}$ and $g(x(t), x(t-h), t)=B x(t)+$ $B_{1} x(t-h)+B_{2} \int_{t-d}^{t} x(s) d s+B_{v} v(t)$, then the linear stochastic system with mixed delays can be written as

$$
\begin{gather*}
d x(t)=\left[A x(t)+A_{1} x(t-h)\right. \\
\\
\left.+A_{2} \int_{t-d}^{t} x(s) d s+A_{v} v(t)\right] d t \\
+ \\
\quad\left[B x(t)+B_{1} x(t-h)\right. \\
d y(t)=\left[C x(t)+C_{1} x(t-h)+C_{2} \int_{t-d}^{t} x(s) d s+C_{v} v(t)\right] d t \\
\left.\quad z(s) d s+B_{v} v(t)\right] d \omega(t)  \tag{41}\\
\quad z(t)=L x(t),
\end{gather*}
$$

which is the same as the system in [38] with constant delays. Thus, following the similar lines in Theorem 4, a sufficient condition can be obtained guaranteeing that there exists a linear filter (4) such that the filtering error system is exponentially stable and achieves a prescribed $L_{2}-L_{\infty}$ performance $\gamma$.

Corollary 5. Consider the stochastic time-delay system (41). For given scalars $\gamma>0, h>0$, and $d>0$ and integers $r_{1} \geq 1$ and $r_{2} \geq 1$, there exists a linear filter (4) such that the corresponding filtering error system is exponentially stable with a guaranteed $L_{2}-L_{\infty}$ performance $\gamma$, if there exist symmetrical positive definite matrices $P \in \mathbf{R}^{2 n \times 2 n}, Q_{i} \in \mathbf{R}^{n \times n}, R_{i}^{T} \in$ $\mathbf{R}^{n \times n}\left(i=1,2, \ldots, r_{1}\right), W_{j} \in \mathbf{R}^{n \times n}$, and $Z_{j} \in \mathbf{R}^{n \times n}(j=$ $1,2, \ldots, r_{2}$ ) and matrix $M \in \mathbf{R}^{n \times n}$ satisfying (14) and

$$
\widetilde{\Phi}=\left[\begin{array}{ccccc}
\widetilde{\Phi}_{11} & \Phi_{12} & \Phi_{13} & P \bar{A}_{v} & \bar{A}^{T} H^{T} M  \tag{42}\\
* & \widetilde{\Phi}_{22} & 0 & 0 & \Phi_{26} \\
* & * & \Phi_{33} & 0 & \Phi_{36} \\
* & * & * & -I & \bar{A}_{v}^{T} H^{T} M \\
* & * & * & * & \Phi_{66}
\end{array}\right]+B_{\xi}^{T} P B_{\xi}<0
$$

where $\widetilde{\Phi}_{11}=\Phi_{11}-\rho H^{T} G_{1}^{T} G_{1} H-\varepsilon H^{T} F_{1}^{T} F_{1} H, \widetilde{\Phi}_{22}=\Phi_{22}-$ $K_{r_{1}}^{T}\left(\rho G_{2}^{T} G_{2}+\varepsilon F_{2}^{T} F_{2}\right) K_{r_{1}}$, and $B_{\xi}=\left[\begin{array}{llll}\bar{B} & \bar{B}_{1} K_{r_{1}} & \bar{B}_{2} K_{r_{2}} & \bar{B}_{v}\end{array}\right]$.

## 4. Filter Design

In this section, we will focus on the design of $L_{2}-L_{\infty}$ filter for stochastic system (1). Based on Theorem 4, a delay-dependent sufficient condition will be obtained in the forms of LMI, which ensures that the filtering error system (6) is stochastically asymptotically stable and achieves a prescribed $L_{2}-L_{\infty}$ performance $\gamma$.

Theorem 6. Consider the stochastic time-delay system (1). For given scalars $\gamma>0, h>0, d>0, \rho>0$, and $\varepsilon>0$ and integers $r_{1} \geq 1$ and $r_{2} \geq 1$, there exists a linear filter (4) such that the filtering error system (6) is stochastically asymptotically stable with a prescribed $L_{2}-L_{\infty}$ performance $\gamma$, if there exist symmetrical positive definite matrices $X \in \mathbf{R}^{n \times n}, Y \in \mathbf{R}^{n \times n}, Q_{i} \in \mathbf{R}^{n \times n}$, $R_{i} \in \mathbf{R}^{n \times n}\left(i=1,2, \ldots, r_{1}\right), W_{j} \in \mathbf{R}^{n \times n}$, and $Z_{j} \in \mathbf{R}^{n \times n}(j=$ $1,2, \ldots, r_{2}$ ) and matrices $M \in \mathbf{R}^{n \times n}, \widehat{A}_{f} \in \mathbf{R}^{n \times m}, \widehat{B}_{f} \in \mathbf{R}^{n \times n}$, and $\widehat{C}_{f} \in \mathbf{R}^{n \times p}$ satisfying

$$
\begin{align*}
& \Upsilon=\left[\begin{array}{cccccc}
\Upsilon_{11} & \Upsilon_{12} & \Upsilon_{13} & \Upsilon_{14} & \Upsilon_{15} & \Upsilon_{16} \\
* & \Phi_{22} & 0 & 0 & 0 & \Upsilon_{26} \\
* & * & \Phi_{33} & 0 & 0 & \Upsilon_{36} \\
* & * & * & -\varepsilon I & 0 & A_{3}^{T} M \\
* & * & * & * & -I & A_{v}^{T} M \\
* & * & * & * & * & \Phi_{66}
\end{array}\right]<0,  \tag{43}\\
& {\left[\begin{array}{cc}
X-\rho I & Y \\
Y & (1-\rho) Y
\end{array}\right]<0,}  \tag{44}\\
& \Pi_{2} \leq I,  \tag{45}\\
& \Lambda=\left[\begin{array}{ccc}
X & Y & L^{T} \\
* & Y & -\widehat{C}_{f}^{T} \\
* & * & \gamma^{2} I
\end{array}\right]>0, \tag{46}
\end{align*}
$$

where

$$
\left.\left.\begin{array}{c}
\Upsilon_{11}=\left[\begin{array}{cc}
\Upsilon_{11}^{1} & \widehat{A}_{f}+A^{T} Y+C^{T} \widehat{B}_{f}^{T} \\
* & \widehat{A}_{f}+\widehat{A}_{f}^{T}
\end{array}\right], \\
\Upsilon_{12}=\left[\begin{array}{c}
R_{1} K_{1}+\left(X A_{1}+\widehat{B}_{f} C_{1}\right) K_{r_{1}} \\
\left(Y A_{1}+\widehat{B}_{f} C_{1}\right) K_{r_{1}}
\end{array}\right], \\
\Upsilon_{13}=\left[\begin{array}{c}
\left(X A_{2}+\widehat{B}_{f} C_{2}\right) K_{r_{2}}+\bar{Z} \\
\left(Y A_{2}+\widehat{B}_{f} C_{2}\right) K_{r_{2}}
\end{array}\right], \\
\Upsilon_{11}^{1}=\operatorname{sym}\left(X A+\widehat{B}_{f} C\right)+Q_{1}-R_{1} \\
\Upsilon_{14}=\left[\begin{array}{c}
d \\
r_{2}
\end{array}\right)^{2} \sum_{j=1}^{r_{2}} W_{j}-\sum_{j=1}^{r_{2}} \frac{2}{2 j-1} Z_{j}+\rho G_{1}^{T} G_{1}+\varepsilon F_{1}^{T} F_{1}, \\
Y A_{3}
\end{array}\right], \begin{array}{c}
X A_{v}+\widehat{B}_{f} C_{v} \\
Y A_{v}+\widehat{B}_{f} C_{v} \tag{47}
\end{array}\right],
$$

In this case, the parameters of a desired filter in the form of (4) can be taken as

$$
\begin{gather*}
A_{f}=\Pi_{1}^{-1} \widehat{A}_{f} \Pi_{1}^{-T} \Pi_{2}, \quad B_{f}=\Pi_{1}^{-1} \widehat{B}_{f}  \tag{48}\\
C_{f}=\widehat{C}_{f} \Pi_{1}^{-T} \Pi_{2}
\end{gather*}
$$

where $\Pi_{1}$ and $\Pi_{2}$ are nonsingular matrices satisfying $0<\Pi_{2}=$ $\Pi_{2}^{T} \leq I, 0<Y=\Pi_{1} \Pi_{2}^{-1} \Pi_{1}^{T}$, and $X>Y>0$.

Proof. From (46), it can be seen that $\left[\begin{array}{c}X \\ Y \\ Y\end{array}\right]>0$, and $X>Y>$ 0 . For any positive definite and symmetric matrix $Y$, one can always find nonsingular matrix $\Pi_{1} \in \mathbf{R}^{n \times n}$ and $0<\Pi_{2}=$ $\Pi_{2}^{T} \in \mathbf{R}^{n \times n}$, such that $Y=\Pi_{1} \Pi_{2}^{-1} \Pi_{1}^{T}$.

Set

$$
P=\left[\begin{array}{cc}
X & \Pi_{1}  \tag{49}\\
\Pi_{1}^{T} & \Pi_{2}
\end{array}\right]
$$

Then $X-\Pi_{1} \Pi_{2}^{-1} \Pi_{1}^{T}=X-Y>0$, and $P>0$.
Define

$$
\Pi=\left[\begin{array}{cc}
I_{n} & 0_{n, n}  \tag{50}\\
0_{n, n} & \Pi_{2}^{-1} \Pi_{1}^{T}
\end{array}\right]
$$

Then

$$
\Pi^{-1}=\left[\begin{array}{cc}
I_{n} & 0_{n, n}  \tag{51}\\
0_{n, n} & \Pi_{1}^{-T} \Pi_{2}
\end{array}\right]
$$

Substitute $\widehat{A}_{f}=\Pi_{1} A_{f} \Pi_{2}^{-1} \Pi_{1}^{T}, \widehat{B}_{f}=\Pi_{1} B_{f}, \widehat{C}_{f}=$ $C_{f} \Pi_{2}^{-1} \Pi_{1}^{T}$, and $Y=\Pi_{1} \Pi_{2}^{-1} \Pi_{1}^{T}$ into (43), (44), and (46) and then pre- and postmultiply (43) by $\operatorname{diag}\left\{\Pi^{-T}, I_{r_{1} n}, I_{r_{2} n}, I_{n}, I_{n}\right.$, $\left.I_{n}\right\}$ and its transpose, respectively. Premultiply and postmultiply (46) by $\operatorname{diag}\left\{\Pi^{-T}, I_{p}\right\}$ and its transpose, respectively. Noticing that $P=\left[\begin{array}{cc}X & \Pi_{1} \\ \Pi_{1}^{T} & \Pi_{2}\end{array}\right]$, using Schur complement Lemma, one can obtain (12) and (14).

On the other hand, (44) implies

$$
\left[\begin{array}{ll}
X & Y  \tag{52}\\
Y & Y
\end{array}\right]<\rho\left[\begin{array}{ll}
I & 0 \\
0 & Y
\end{array}\right]
$$

Pre- and postmultiply (52) by $\Pi^{-T}$ and $\Pi^{-1}$, respectively. Notice that $Y=\Pi \Lambda^{-1} \Pi^{T}$, one can obtain

$$
\left[\begin{array}{cc}
X & \Pi_{1}  \tag{53}\\
\Pi_{1}^{T} & \Pi_{2}
\end{array}\right]<\rho\left[\begin{array}{cc}
I & 0 \\
0 & \Pi_{2}
\end{array}\right]
$$

By (45), it is easy to see that

$$
\left[\begin{array}{cc}
X & \Pi_{1}  \tag{54}\\
\Pi_{1}^{T} & \Pi_{2}
\end{array}\right]<\rho\left[\begin{array}{cc}
I & 0 \\
0 & I
\end{array}\right]
$$

So, (13) is satisfied.
Therefore, by Theorem 4, the suitable filter parameters can be constructed by (48), which ensures the filtering error system (6) to be stochastically asymptotically stable with $L_{2}-L_{\infty}$ performance $\gamma$. This completes the proof.

Remark 7. When deriving the results in Theorem 6 based on Theorem 4, considering dealing with the LMI (13), we give a method to avoid nonlinear terms emerging. Using Matlab LMI toolbox, one can solve linear matrix inequalities (43)(44) and (46). Then, by matrix diagonalization approach, one can easily find that diagonally positive matrix $\Pi_{2}$ and nonsingular matrix $\Pi_{1}$ satisfy $\Pi_{2}=\Pi_{1}^{T} Y^{-1} \Pi_{1}$. If the obtained matrix $\Pi_{2}$ does not satisfy (45), one can take $\Pi_{2}$ for $\Pi_{2} /$ $\max \left\{\operatorname{eig}\left(\Pi_{2}\right)\right\}$ and $\Pi_{1}$ for $\sqrt{\max \left\{\operatorname{eig}\left(\Pi_{2}\right)\right\}} \Pi_{1}$. Thus, the desired filter parameters can be obtained by (48).

Following the similar method in Theorem 6, one can obtain a result of filter design for linear stochastic time-delay system (41).

Corollary 8. Consider the stochastic time-delay system (41). For given scalars $\gamma>0, h>0$, and $d>0$ and integers $r_{1} \geq 1$ and $r_{2} \geq 1$, there exists a linear filter (4) such that the corresponding filtering error system is stochastically asymptotically stable with a prescribed $L_{2}-L_{\infty}$ performance $\gamma$, if there exist symmetrical positive definite $X \in \mathbf{R}^{n \times n}, Y \in \mathbf{R}^{n \times n}, Q_{i} \in \mathbf{R}^{n \times n}$, $R_{i} \in \mathbf{R}^{n \times n}\left(i=1,2, \ldots, r_{1}\right), W_{j} \in \mathbf{R}^{n \times n}$, and $Z_{j} \in \mathbf{R}^{n \times n}(j=$ $1,2, \ldots, r_{2}$ ) and matrices $M \in \mathbf{R}^{n \times n}, \widehat{A}_{f} \in \mathbf{R}^{n \times m}, \widehat{B}_{f} \in \mathbf{R}^{n \times n}$, and $\widehat{C}_{f} \in \mathbf{R}^{n \times p}$ satisfying (46) and

$$
\tilde{\Upsilon}=\left[\begin{array}{cccccc}
\widetilde{\Upsilon}_{11} & \Upsilon_{12} & \Upsilon_{13} & \Upsilon_{15} & \Upsilon_{16} & \Upsilon_{17}  \tag{55}\\
* & \widetilde{\Phi}_{22} & 0 & 0 & \Upsilon_{26} & \Upsilon_{27} \\
* & * & \Phi_{33} & 0 & \Upsilon_{36} & \Upsilon_{37} \\
* & * & * & -I & A_{v}^{T} M & \Upsilon_{47} \\
* & * & * & * & \Phi_{66} & 0 \\
* & * & * & * & * & \Upsilon_{77}
\end{array}\right]<0
$$

where

$$
\begin{gather*}
\widetilde{\Upsilon}_{11}=\left[\begin{array}{cc}
\widetilde{\Upsilon}_{11}^{1} & \widehat{A}_{f}+A^{T} Y+C^{T} \widehat{B}_{f}^{T} \\
* & \widehat{A}_{f}+\widehat{A}_{f}^{T}
\end{array}\right], \\
\widetilde{\Upsilon}_{11}^{1}=\Upsilon_{11}^{1}-\rho G_{1}^{T} G_{1}-\varepsilon F_{1}^{T} F_{1}, \quad \Upsilon_{17}=\left[\begin{array}{cc}
B^{T} X & B^{T} Y \\
0_{n, n} & 0_{n, n}
\end{array}\right], \\
\Upsilon_{27}=K_{r_{1}}^{T}\left[\begin{array}{ll}
B_{1}^{T} X & B_{1}^{T} Y
\end{array}\right], \quad \Upsilon_{37}=K_{r_{2}}^{T}\left[\begin{array}{cc}
B_{2}^{T} X & B_{2}^{T} Y
\end{array}\right], \\
\Upsilon_{47}=\left[\begin{array}{ll}
B_{v}^{T} X & B_{v}^{T} Y
\end{array}\right], \quad \Upsilon_{77}=\left[\begin{array}{cc}
-X & -Y \\
* & -Y
\end{array}\right] . \tag{56}
\end{gather*}
$$

Remark 9. The results presented in Theorem 6 and Corollary 8 can be easily extended to the systems with only discrete or distributed delays and also to the robust performance analysis for uncertain stochastic systems with mixed delays.

## 5. Numerical Examples

Example 1. Consider the stochastic time-delay system (1) with parameters

$$
\begin{array}{cc}
A=\left[\begin{array}{cc}
-1.5 & 0.5 \\
-1 & -3
\end{array}\right], & A_{1}=\left[\begin{array}{cc}
-0.8 & 0.2 \\
0.2 & -0.5
\end{array}\right], \\
A_{2}=\left[\begin{array}{cc}
0.2 & 0 \\
0 & 0.2
\end{array}\right], & A_{3}=\left[\begin{array}{cc}
0.1 & 0 \\
0 & 0.1
\end{array}\right]
\end{array}
$$

Table 1: The upper bound of $d_{\text {max }}$ for $h=1$ and $\gamma=0.2$.

| Methods | $d_{\max }$ |
| :--- | :---: |
| Theorem $6\left(r_{1}=1, r_{2}=1\right)$ | 7.481 |
| Theorem $6\left(r_{1}=2, r_{2}=2\right)$ | 8.190 |
| Theorem $6\left(r_{1}=3, r_{2}=3\right)$ | 8.317 |
| Theorem $6\left(r_{1}=5, r_{2}=5\right)$ | 8.379 |

$$
\begin{align*}
A_{v}=\left[\begin{array}{cc}
0.2 & 0 \\
0 & -0.2
\end{array}\right], & C=\left[\begin{array}{cc}
2 & -0.5 \\
-1.5 & 0.5
\end{array}\right], \\
C_{1}=\left[\begin{array}{cc}
0.15 & 0.1 \\
-0.1 & 0.1
\end{array}\right], & C_{2}=\left[\begin{array}{cc}
0.5 & -0.2 \\
0.6 & 0
\end{array}\right], \\
C_{v}=\left[\begin{array}{cc}
0.1 & 0.2 \\
0.1 & 0.03
\end{array}\right], & L=\left[\begin{array}{cc}
0.1 & -0.2 \\
0 & 0.1
\end{array}\right], \\
\varepsilon=1, & \rho=7 . \tag{57}
\end{align*}
$$

Moreover, for the nonlinear functions, we let $G_{1}=G_{2}=$ $0.1 I$ and $F_{1}=F_{2}=0.1 I$. Given $h=1$ and $\gamma=0.2$, from Theorem 6, one can obtain the upper bound of time delay $d$, which is listed in Table 1.

In the case of $r_{1}=2$ and $r_{2}=2$, the desired filter parameters can be obtained:

$$
\begin{gather*}
A_{f}=\left[\begin{array}{cc}
-6.1460 & 2.2644 \\
0.6168 & -4.5679
\end{array}\right], \quad B_{f}=\left[\begin{array}{cc}
-3.8749 & 1.8868 \\
0.7066 & -0.1266
\end{array}\right], \\
C_{f}=\left[\begin{array}{cc}
-0.0568 & 0.0052 \\
0.0202 & 0.0057
\end{array}\right] . \tag{58}
\end{gather*}
$$

Example 2. Consider the stochastic time-delay system (36)(39) with parameters

$$
\begin{array}{cc}
A=\left[\begin{array}{cc}
-1.5 & 0.5 \\
-1 & -3
\end{array}\right], & A_{1}=\left[\begin{array}{cc}
-0.8 & 0.2 \\
0.2 & -0.5
\end{array}\right], \\
A_{2}=\left[\begin{array}{cc}
0.2 & 0 \\
0 & 0.2
\end{array}\right], & A_{v}=\left[\begin{array}{cc}
0.2 & 0 \\
0 & -0.2
\end{array}\right], \\
B=\left[\begin{array}{cc}
-0.8 & 0.2 \\
0.5 & -0.5
\end{array}\right], & B_{1}=\left[\begin{array}{cc}
0.5 & 0.5 \\
0.2 & 0.3
\end{array}\right], \\
B_{2}=\left[\begin{array}{cc}
0.2 & 0 \\
0 & 0.2
\end{array}\right], & B_{v}=\left[\begin{array}{cc}
-0.2 & 0 \\
0 & 0.5
\end{array}\right],  \tag{59}\\
C=\left[\begin{array}{cc}
2 & -0.5 \\
-1.5 & 0.5
\end{array}\right], & C_{1}=\left[\begin{array}{cc}
0.15 & 0.1 \\
-0.1 & 0.1
\end{array}\right], \\
C_{2}=\left[\begin{array}{cc}
0.5 & -0.2 \\
0.6 & 0
\end{array}\right], & C_{v}=\left[\begin{array}{cc}
0.1 & 0.2 \\
0.1 & 0.03
\end{array}\right], \\
& L=\left[\begin{array}{cc}
0.1 & -0.2 \\
0 & 0.1
\end{array}\right] .
\end{array}
$$

Table 2: The upper bound of $h_{\text {max }}$ for $d=1$ and $\gamma=0.2$.

| Methods | $h_{\max }$ |
| :--- | :---: |
| $[38]$ | 1.725 |
| Corollary $8\left(r_{1}=1, r_{2}=1\right)$ | 3.755 |
| Corollary $8\left(r_{1}=2, r_{2}=1\right)$ | 5.054 |
| Corollary $8\left(r_{1}=2, r_{2}=2\right)$ | 5.111 |
| Corollary $8\left(r_{1}=3, r_{2}=3\right)$ | 5.688 |
| Corollary $8\left(r_{1}=5, r_{2}=5\right)$ | 5.920 |

Given $d=1$ and $\gamma=0.2$, from Corollary 8, one can obtain the upper bound of time delay $h$. Table 2 lists the results of Corollary 8 and [38] with constant delays. It is easy to see that the proposed filter design method in this paper is less conservative than [38].

From Corollary 8, in the case of $r_{1}=2$ and $r_{2}=2$, the desired filter parameters can be obtained:

$$
\begin{gather*}
A_{f}=\left[\begin{array}{cc}
-0.0000 & -4.1570 \\
-0.0000 & -2.5527
\end{array}\right], \quad B_{f}=\left[\begin{array}{cc}
10.1507 & -11.1236 \\
6.0149 & -6.1484
\end{array}\right], \\
C_{f}=\left[\begin{array}{cc}
0.0000 & 0.0065 \\
-0.0000 & -0.0027
\end{array}\right] . \tag{60}
\end{gather*}
$$

Remark 3. It can be seen from the results that the conservatism can be reduced with the increase of partition integers. However, it is necessary to point out that the less conservatism is at the cost of a higher computational complexity.

## 6. Conclusions

In this paper, a new approach has been developed to investigate the problems of delay-dependent $L_{2}-L_{\infty}$ filter design for stochastic system with mixed delays and nonlinear perturbations. Based on the idea of delay partitioning and integral partitioning, using Lyapunov-Krasovskii functional approach, a delay-dependent sufficient condition has been established that ensures the filtering error system is exponentially stable with $L_{2}-L_{\infty}$ performance $\gamma$. By solving the LMIs, one can get the desired filter gain matrices. The results also depend on the partition integers with the increase of partition integers, the conservatism can be decreased. Finally, numerical examples are presented to demonstrate the effectiveness of the proposed approach.

## Acknowledgments

The project is supported by Nature Science Foundation of China (NSFC, Grant no. 51277022). The authors would also like to thank the editor and the anonymous reviewers for their valuable comments and suggestions which helped to improve the quality of this paper.

## References

[1] K. Gu, V. L. Kharitonov, and J. Chen, Stability of Time-Delay Systems, Birkhäauser, Boston, Mass, USA, 2003.
[2] E. Fridman and U. Shaked, "A descriptor system approach to $H_{\infty}$ control of linear time-delay systems," IEEE Transactions on Automatic Control, vol. 47, no. 2, pp. 253-270, 2002.
[3] F. Gouaisbaut and D. Peaucelle, "A note on stability of time delay systems," in Proceedings of the IFAC Symposium on Robust Control Design, Toulouse, France, 2006.
[4] X. M. Zhang and Q. L. Han, "A delay decomposition approach to delay-dependent stability for linear systems with time-varying delays," International Journal of Robust and Nonlinear Control, vol. 19, no. 17, pp. 1922-1930, 2009.
[5] T. L. Hsien and C. H. Lee, "Robust stability of discrete bilinear uncertain time-delay systems," Circuits, Systems, and Signal Processing, vol. 30, no. 6, pp. 1417-1443, 2011.
[6] P. Shi, R. K. Agarwal, E. K. Boukas, and Y. Shi, "Optimal guaranteed cost control of uncertain discrete time-delay systems," Journal of Computational and Applied Mathematics, vol. 157, no. 2, pp. 435-451, 2003.
[7] S. Ma and E. K. Boukas, "Stability and robust stabilisation for uncertain discrete stochastic hybrid singular systems with time delay," IET Control Theory \& Applications, vol. 3, no. 9, pp. 12171225, 2009.
[8] L. Wu and D. W. C. Ho, "Reduced-order $L_{2}-L_{\infty}$ filtering for a class of nonlinear switched stochastic systems," IET Control Theory \& Applications, vol. 3, no. 5, pp. 493-508, 2009.
[9] H. Gao and C. Wang, "Robust $L_{2}-L_{\infty}$ filtering for uncertain systems with multiple time-varying state delays," IEEE Transactions on Circuits and Systems I, vol. 50, no. 4, pp. 594-599, 2003.
[10] C. Lin, Q. G. Wang, and T. H. Lee, "A less conservative robust stability test for linear uncertain time-delay systems," IEEE Transactions on Automatic Control, vol. 51, no. 1, pp. 87-91, 2006.
[11] H. Zhang and Z. Liu, "Stability analysis for linear delayed systems via an optimally dividing delay interval approach," Automatica, vol. 47, no. 9, pp. 2126-2129, 2011.
[12] Q. L. Han, "A discrete delay decomposition approach to stability of linear retarded and neutral systems," Automatica, vol. 45, no. 2, pp. 517-524, 2009.
[13] B. Zhang and Y. Li, "Exponential $L_{2}-L_{\infty}$ filtering for distributed delay systems with Markovian jumping parameters," Signal Processing, vol. 93, no. 1, pp. 206-216, 2013.
[14] Z. Feng and J. Lam, "Integral partitioning approach to robust stabilization for uncertain distributed time-delay systems," International Journal of Robust and Nonlinear Control, vol. 22, no. 6, pp. 676-689, 2012.
[15] Y. Liu, Z. Wang, and X. Liu, "Robust $H_{\infty}$ control for a class of nonlinear stochastic systems with mixed time delay", International Journal of Robust and Nonlinear Control, vol. 17, no. 16, pp. 1525-1551, 2007.
[16] Z. Wu and W. Zhou, "Delay-dependent robust stabilization for uncertain singular systems with discrete and distributed delays," Journal of Control Theory and Applications, vol. 6, no. 2, pp. 171176, 2008.
[17] J. Wu, G. Lu, S. Wo, and X. Xiao, "Exponential stability and stabilization for nonlinear descriptor systems with discrete and distributed delays," International Journal of Robust and Nonlinear Control, vol. 23, no. 12, pp. 1393-1404, 2013.
[18] M. S. Mahmoud, A. Y. Al-Rayyah, and Y. Xia, "Quantised feedback stabilization of interconnected discrete-delay systems,"

IET Control Theory \& Applications, vol. 5, no. 6, pp. 795-802, 2011.
[19] L. Wu, X. Su, P. Shi, and J. Qiu, "A new approach to stability analysis and stabilization of discrete-time T-S fuzzy time-varying delay systems," IEEE Transactions on Systems, Man, and Cybernetics $B$, vol. 41, no. 1, pp. 273-286, 2011.
[20] Y. E. Wang, X. M. Sun, and J. Zhao, "Stabilization of a class of switched stochastic systems with time delays under asynchronous switching," Circuits, Systems, and Signal Processing, vol. 32, no. 1, pp. 347-360, 2013.
[21] S. Xie and L. Xie, "Stabilization of a class of uncertain large-scale stochastic systems with time delays," Automatica, vol. 36, no. 1, pp. 161-167, 2000.
[22] D. Yue and S. Won, "Delay-dependent robust stability of stochastic systems with time delay and nonlinear uncertainties," Electronics Letters, vol. 37, no. 15, pp. 992-993, 2001.
[23] D. Yue and Q. L. Han, "Delay-dependent exponential stability of stochastic systems with time-varying delay, nonlinearity, and Markovian switching," IEEE Transactions on Automatic Control, vol. 50, no. 2, pp. 217-222, 2005.
[24] S. Xu, J. Lam, X. Mao, and Y. Zou, "A new LMI condition for delay-dependent robust stability of stochastic time-delay systems," Asian Journal of Control, vol. 7, no. 4, pp. 419-423, 2005.
[25] R. Yang, P. Shi, and H. Gao, "New delay-dependent stability criterion for stochastic systems with time delays," IET Control Theory \& Applications, vol. 2, no. 11, pp. 966-973, 2008.
[26] P. Cheng, F. Deng, and Y. Peng, "Delay-dependent exponential stability of impulsive stochastic systems with time-varying delay," Journal of Systems Engineering and Electronics, vol. 22, no. 5, pp. 799-809, 2011.
[27] S. Xu and T. Chen, "Robust $H_{\infty}$ control for uncertain stochastic systems with state delay," IEEE Transactions on Automatic Control, vol. 47, no. 12, pp. 2089-2094, 2002.
[28] S. Xu and T. Chen, " $H_{\infty}$ output feedback control for uncertain stochastic systems with time-varying delays," Automatica, vol. 40, no. 12, pp. 2091-2098, 2004.
[29] E. K. Boukas and Z. K. Liu, "Robust $H_{\infty}$ filtering for polytopic uncertain time-delay systems with Markov jumps," Computers and Electrical Engineering, vol. 28, no. 3, pp. 171-193, 2002.
[30] W. Zhang, B. S. Chen, and C. S. Tseng, "Robust $H_{\infty}$ filtering for nonlinear stochastic systems," IEEE Transactions on Signal Processing, vol. 53, no. 2, pp. 589-598, 2005.
[31] W. Zhang, B. S. Chen, L. Sheng, and M. Gao, "Robust $H_{2} / H_{\infty}$ filter design for a class of nonlinear stochastic systems with statedependent noise," Mathematical Problems in Engineering, vol. 2012, Article ID 750841, 16 pages, 2012.
[32] H. Gao, J. Lam, and C. Wang, "Robust energy-to-peak filter design for stochastic time-delay systems," Systems \& Control Letters, vol. 55, no. 2, pp. 101-111, 2006.
[33] Z. Wang, F. Yang, D. W. C. Ho, and X. Liu, "Robust $H_{\infty}$ filtering for stochastic time-delay systems with missing measurements," IEEE Transactions on Signal Processing, vol. 54, no. 7, pp. 25792587, 2006.
[34] J. Xia, S. Xu, and B. Song, "Delay-dependent $L_{2}-L_{\infty}$ filter design for stochastic time-delay systems," Systems \& Control Letters, vol. 56, no. 9-10, pp. 579-587, 2007.
[35] Y. Chen, A. Xue, and S. Zhou, "New delay-dependent $L_{2}-L_{\infty}$ filter design for stochastic time-delay systems," Signal Processing, vol. 89, no. 6, pp. 974-980, 2009.
[36] R. Yang, H. Gao, and P. Shi, "Delay-dependent $L_{2}-L_{\infty}$ filter design for stochastic time-delay systems," IET Control Theory \& Applications, vol. 5, no. 1, pp. 1-8, 2011.
[37] R. Yang, H. Gao, P. Shi, and L. Zhang, "Delay-dependent energy-to-peak filter design for stochastic systems with time delay: a delay partitioning approach," in Proceedings of the 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference (CDC/CCC '09), pp. 5472-5477, Shanghai, China, December 2009.
[38] L. Li and Y. Jia, "Robust $L_{2}-L_{\infty}$ filtering for stochastic systems with discrete and distributed time-varying delays," Asian Journal of Control, vol. 14, no. 4, pp. 1047-1058, 2012.
[39] Z. Wang, Y. Liu, and X. Liu, "Exponential stabilization of a class of stochastic system with Markovian jump parameters and mode-dependent mixed time-delays," IEEE Transactions on Automatic Control, vol. 55, no. 7, pp. 1656-1662, 2010.
[40] K. Gu, "An integral inequality in the stability problem of timedelay systems," in Proceedings of the 39th IEEE Confernce on Decision and Control, pp. 2805-2810, Sydney, Australia, December 2000 .

## Research Article

# Improving the Asymptotic Properties of Discrete System Zeros in Fractional-Order Hold Case 

Cheng Zeng, ${ }^{1,2}$ Shan Liang, ${ }^{1,3}$ and Yingying Su ${ }^{1,4}$<br>${ }^{1}$ College of Automation, Chongqing University, Chongqing 400044, China<br>${ }^{2}$ Department of Mathematics, Guiyang University, Guiyang 550005, China<br>${ }^{3}$ Key Laboratory of Dependable Service Computing in Cyber Physical Society Ministry of Education, Chongqing University, Chongqing 400044, China<br>${ }^{4}$ Department of Electric and Electronic Information Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

Correspondence should be addressed to Shan Liang; lightsun@cqu.edu.cn
Received 17 July 2013; Revised 11 October 2013; Accepted 21 October 2013
Academic Editor: Baocang Ding
Copyright © 2013 Cheng Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Remarkable improvements in the asymptotic properties of discrete system zeros may be achieved by properly adjusted fractionalorder hold (FROH) circuit. This paper analyzes asymptotic properties of the limiting zeros, as the sampling period $T$ tends to zero, of the sampled-data models on the basis of the normal form representation of the continuous-time systems with FROH. Moreover, when the relative degree of the continuous-time system is equal to one or two, an approximate expression of the limiting zeros for the sampled-data system with FROH is also given as power series with respect to a sampling period up to the third-order term. And, further, the corresponding stability conditions of the sampling zeros are discussed for fast sampling rates. The ideas of the paper here provide a more accurate approximation for asymptotic zeros, and certain known achievements on asymptotic behavior of limiting zeros are shown to be particular cases of the results presented.


## 1. Introduction

Zeros, along with poles, are fundamental characteristics of linear time-invariant systems, and the stability of zeros is one of the most important issues in the model matching and adaptive control problems. When a continuous-time system is discretized by the use of a sampler and a hold, the mapping between the discrete-time poles and their continuous-time counterparts is very simple; namely, stability of poles is reserved. There is unfortunately no simple transformation between the discrete-time zeros and their continuous-time ones because the zeros of discrete-time systems depend on sampling period $T$ [1]. More precisely, it is generally impossible to tranform a continuous-time system with zeros in the left-half plane to a discrete-time system with zeros inside the unit circle. That is to say, the stability of zeros is not necessarily preserved except in special cases. Therefore, one of the special cases (i.e., the limiting case) is that the sampling
period $T$ tends to zero which has attracted considerable attention from the engineering point of view.

Perhaps the first attempt to study discrete system zeros was given by Åström and coworkers [1], who describe the asymptotic behavior of the discrete-time zeros for fast sampling rate when the original continuous-time plant is discretized with zero-order hold (ZOH). In this case, the discretized zeros are further called limiting zeros which are composed of the intrinsic zeros and sampling zeros [2]. The former ones have counterparts in the underlying continuoustime system and go to unity [3] while the latter ones, which have no continuous-time counterparts and are generated in the sampling process, go toward roots of a certain polynomial $[4,5]$ determined by relative degree of the continuous-time system.

In much of discussion about the properties of discretetime zeros, ZOH has been mainly employed as a hold circuit since it is used most commonly in practice [1, 3, 6-10].

Taking into account the fact that the type of hold circuit used critically influences the position of zeros, it is an interesting problem to investigate the zeros in the case of various holds. Hagiwara et al. [4] have carried out a comparative study and demonstrated that a first-order hold $(\mathrm{FOH})$ provides no advantage over ZOH as far as the stability of zeros of the resulting discrete-time systems is concerned. Passino and Antsaklis [11] have considered the fractional-order hold ( FROH ) as an alternative to the ZOH and shown that it can locate the zeros of discrete-time system inside the unit circle by some examples while ZOH fails to do so. In the very motivating work by Ishitobi [12], the properties of limiting zeros with FROH have been analyzed, and the corresponding pulse-transfer function has been also derived.

Moreover, Ishitobi has definitely presented the relationship between the relative degree and discretized zeros behavior when the continuous-time systems have the relative degree up to five for sufficiently small sampling periods. Further, Bàrcena et al. [13, 14] and Liang et al. [15, 16] have extended Ishitobi's results [12] from different angles and methods by investigating the limiting zeros in the case of a FROH. In addition, the limiting FROH zeros [12] have been also extended by Blachuta [17], who describes the accuracy of the asymptotic results for both the intrinsic and sampling zeros in terms of Bernoulli numbers and parameters of the continuous-time transfer function for sufficiently small sampling periods $T$.

In FROH case, the intrinsic zeros are located inside (resp., outside) the unit circle for small sampling periods when the corresponding continuous-time zeros lie strictly in the lefthalf plane (resp., right). For sampling zeros, at least one of the zeros lies strictly outside the unit circle if the relative degree of a continuous-time transfer function is greater than or equal to three $[12,18]$. This fact indicates that even though all the zeros of such a continuous-time system are stable, the corresponding discrete-time system has at least one unstable zero in the limiting case as the sampling period tends to zero. Thus, attention is here focused on continuous-time systems with relative degree less than or equal to two. More specifically, the corresponding discrete-time plants have one or two sampling zero(s) in the case of a FROH when the relative degree of a continuous-time transfer function is one or two. However, in these cases, the sampling zeros are located just on the unit circle, that is, in the marginal case of the stability. More importantly, it is a valuable research topic to find the criteria which guarantee that stable discretized zeros are obtained. Thus, the asymptotic behavior of the sampling zeros is an interesting issue as we explore the stability properties of the sampling zeros by analyzing the asymptotic properties as the sampling period tends to zero.

The objective of this paper is to analyze the improved asymptotic properties of the limiting zeros for discrete-time models by using a new kind of method. More precisely, we give an approximate expression of limiting zeros for the sampled-data system on the basis of the normal form representation of continuous-time system with FROH as power series with respect to a sampling period up to the third-order term when the relative degree of the continuous-time system is one or two. Our results include also the finding of how close
limiting zeros are to the actual intrinsic and sampling ones, irrespectively of whether they are stable or not. The approach used could be referred to as an extension of that of [12, 17, 18], and one of the principal contribution in this paper, in particular, would consequently propose an analytical method to obtain the FROH zeros as stable as possible, or with improved asymptotic properties even when unstable, for a given continuous-time plant. Finally, we further discuss the stability of the sampling zeros for sufficiently small sampling periods, and some interesting examples are given to validate the main results.

## 2. Sampled-Data Models with FROH

Consider an $n$th continuous-time system with relative degree one or two described by a transfer function

$$
\begin{equation*}
G(s)=K \frac{N(s)}{D(s)}, \quad K \neq 0 \tag{1}
\end{equation*}
$$

where

$$
\begin{array}{r}
N(s)=s^{m}+b_{m-1} s^{m-1}+b_{m-2} s^{m-2}+\cdots+b_{0} \\
m=n-1 \quad \text { or } n-2,  \tag{2}\\
D(s)=s^{n}+a_{n-1} s^{n-1}+a_{n-2} s^{n-2}+\cdots+a_{0}
\end{array}
$$

The paper treats systems with relative degree one or two because at least one of the limiting zeros is unstable when the relative degree is greater than or equal to three though it is slightly a limitation.
2.1. Case of Relative Degree One ( $m=n-1$ ). The normal form of (1) with the relative degree one, $m=n-1$ is represented with an input $u$ and an output $y[19,20]$ as

$$
\begin{gather*}
\dot{\xi}=-d \xi+K u-\omega \\
\dot{\boldsymbol{\eta}}=P \boldsymbol{\eta}+\mathbf{q} \xi  \tag{3}\\
y=\xi
\end{gather*}
$$

where

$$
\begin{gather*}
\boldsymbol{\eta}=\left[\begin{array}{lll}
\boldsymbol{\eta}_{1} & \cdots & \boldsymbol{\eta}_{n-1}
\end{array}\right]^{T}, \\
\omega=\mathbf{c}^{T} \boldsymbol{\eta}, \quad \mathbf{c}=\left[\begin{array}{lll}
r_{0} & r_{1} & \cdots
\end{array} r_{n-2}\right]^{T}, \\
P=\left[\begin{array}{cccc}
0 & 1 & & O \\
& & \ddots & \\
-b_{0} & \cdots & & -b_{n-2}
\end{array}\right], \quad \mathbf{q}=\left[\begin{array}{c}
0 \\
\vdots \\
0 \\
1
\end{array}\right], \tag{4}
\end{gather*}
$$

and the scalars $d$ and $r_{i}(i=0, \ldots, n-2)$ are obtained from

$$
\begin{gather*}
D(s)=Q(s) N(s)+R(s) \\
Q(s)=s+d  \tag{5}\\
R(s)=r_{n-2} s^{n-2}+r_{n-3} s^{n-3}+\cdots+r_{0}
\end{gather*}
$$



Figure 1: The signal reconstruction of a fractional-order hold with $\beta=-0.5$.
where

$$
\begin{gather*}
d=a_{n-1}-b_{n-2} \\
r_{i}=a_{i}-b_{i-1}-b_{i} d, \quad i=0, \ldots, n-2 \tag{6}
\end{gather*}
$$

When the FROH signal reconstruction method is considered, the input is described by

$$
\begin{array}{r}
v(t)=u(k T)+\beta\left[\frac{u(k T)-u((k-1) T)}{T}\right](t-k T),  \tag{7}\\
k T \leq t<(k+1) T, \quad k=0,1, \ldots,
\end{array}
$$

where $\beta$ is a real design parameter and $T$ is a sampling period [11, 12, 18]. It is obvious that FROH is reduced to ZOH for $\beta=0$ while it becomes the FOH for $\beta=1$. The signal reconstruction of a FROH with $\beta=-0.5$ is shown in Figure 1.

Suppose $u(t)=v(t)$, and when a FROH is applied, we have

$$
\begin{equation*}
\dot{u}(t)=\beta\left[\frac{u(k T)-u((k-1) T)}{T}\right], \quad \ddot{u}(t)=\cdots=0 . \tag{8}
\end{equation*}
$$

Furthermore, (3) leads to the derivatives of the output

$$
\begin{gather*}
\dot{y}=-d \xi+K u-\mathbf{c}^{T} \boldsymbol{\eta},  \tag{9}\\
\ddot{y}=\left(d^{2}-\mathbf{c}^{T} \mathbf{q}\right) \xi-d K u+\left(d \mathbf{c}^{T}-\mathbf{c}^{T} P\right) \boldsymbol{\eta}+K \dot{u},  \tag{10}\\
y^{(3)}=\left(-d^{3}+2 d \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}\right) \xi+\left(d^{2}-\mathbf{c}^{T} \mathbf{q}\right) K u \\
 \tag{11}\\
+\left(d \mathbf{c}^{T} P-\mathbf{c}^{T} P^{2}-d^{2} \mathbf{c}^{T}+\mathbf{c}^{T} \mathbf{q} \mathbf{c}^{T}\right) \boldsymbol{\eta}-d K \dot{u} \\
y^{(4)}=\left\{d^{4}-3 d^{2} \mathbf{c}^{T} \mathbf{q}+2 d \mathbf{c}^{T} P \mathbf{q}-\mathbf{c}^{T} P^{2} \mathbf{q}+\left(\mathbf{c}^{T} \mathbf{q}\right)^{2}\right\} \xi \\
+\left(-d^{3}+2 d \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}\right) K u \\
+  \tag{12}\\
\\
\\
\\
\quad\left\{d^{3} \mathbf{c}^{T}-2 d \mathbf{c}^{T} \mathbf{q} \mathbf{c}^{T}+\mathbf{c}^{T} P \mathbf{q} \mathbf{c}^{T}+d \mathbf{c}^{T} P^{2}\right. \\
\end{gather*}
$$

which are expressed by $\xi, \boldsymbol{\eta}$, and $K u$. Further, the derivatives of $\boldsymbol{\eta}$ are also represented by $\xi, \boldsymbol{\eta}$, and $K u$ as

$$
\begin{gather*}
\dot{\boldsymbol{\eta}}=P \boldsymbol{\eta}+\mathbf{q} \xi,  \tag{13}\\
\ddot{\boldsymbol{\eta}}=(P \mathbf{q}-\mathbf{q} d) \xi+\left(P^{2}-\mathbf{q} \mathbf{c}^{T}\right) \boldsymbol{\eta}+\mathbf{q} K u,  \tag{14}\\
\boldsymbol{\eta}^{(3)}=\left(P^{2} \mathbf{q}-\mathbf{q} c^{T} \mathbf{q}-P \mathbf{q} d+\mathbf{q} d^{2}\right) \xi+(P \mathbf{q}-\mathbf{q} d) K u  \tag{15}\\
+\left(P^{3}-\mathbf{q} \mathbf{c}^{T} P-P \mathbf{q} c^{T}+\mathbf{q} c^{T} d\right) \boldsymbol{\eta}+\mathbf{q} K \dot{u} .
\end{gather*}
$$

Hence, by substituting (9)-(15) into the right-hand side of

$$
\begin{align*}
& y_{k+1}=\sum_{i=0}^{\infty} \frac{T^{i}}{i!} y_{k}^{(i)},  \tag{16}\\
& \boldsymbol{\eta}_{k+1}=\sum_{i=0}^{\infty} \frac{T^{i}}{i!} \boldsymbol{\eta}_{k}^{(i)}
\end{align*}
$$

and defining the state variables $x_{k}=\left[y_{k}, \boldsymbol{\eta}_{k}^{T}\right]^{T}$, where the subscript $k$ denotes $t=k T$, the discrete-time state equations are definitely obtained. It is easy to show that zeros of a discrete-time system for a transfer function (1) are derived from (16).

Now, by applying the explicit expressions of $y_{k}, \dot{y}_{k}$, $\ldots, y_{k}^{(4)}$ and $\boldsymbol{\eta}_{k}, \ldots, \boldsymbol{\eta}_{k}^{(3)}$, the zeros of (16) are analyzed as follows:

$$
\begin{aligned}
y_{k+1}= & \sum_{i=0}^{4} \frac{T^{i}}{i!} y_{k}^{(i)}+O\left(T^{5}\right) \\
= & \left(1-d T+\frac{d^{2}-r_{n-2}}{2} T^{2}+\frac{2 d r_{n-2}-d^{3}-\mathbf{c}^{T} P \mathbf{q}}{6} T^{3}\right. \\
& \left.+\frac{d^{4}-3 d^{2} r_{n-2}+2 d \mathbf{c}^{T} P \mathbf{q} r_{n-2}^{2}-\mathbf{c}^{T} P^{2} \mathbf{q}}{24} T^{4}\right) y_{k} \\
+ & \left\{\left(1+\frac{\beta}{2}\right) T-\left(\frac{d}{2}+\frac{d \beta}{6}\right) T^{2}\right. \\
& +\frac{4 d^{2}-4 r_{n-2}-r_{n-2} \beta+d^{2} \beta}{24} T^{3} \\
& \left.+\frac{2 d r_{n-2}-d^{3}-\mathbf{c}^{T} P \mathbf{q}}{24} T^{4}\right\} K u_{k} \\
+ & \left\{-\frac{\beta}{2} T+\frac{d \beta}{6} T^{2}-\frac{\left(d^{2}-r_{n-2}\right) \beta}{24} T^{3}\right\} K u_{k-1} \\
+ & \left(-\mathbf{c}^{T} T+\frac{d \mathbf{c}^{T}-\mathbf{c}^{T} P}{2} T^{2}\right. \\
& +\frac{d \mathbf{c}^{T} P-\mathbf{c}^{T} P^{2}+\left(r_{n-2}-d^{2}\right) \mathbf{c}^{T}}{6} T^{3}
\end{aligned}
$$

$$
\begin{align*}
&+\left(\left(\left(d^{3}-2 d r_{n-2}\right) \mathbf{c}^{T}+\mathbf{c}^{T} P \mathbf{q} c^{T}+d \mathbf{c}^{T} P^{2}\right.\right. \\
&\left.-\mathbf{c}^{T} P^{3}-d^{2} \mathbf{c}^{T} P+r_{n-2} \mathbf{c}^{T} P\right) \\
&\left.\left.\times(24)^{-1}\right) T^{4}\right) \boldsymbol{\eta}_{k}+O\left(T^{5}\right),  \tag{21}\\
& \boldsymbol{\eta}_{k+1}=\sum_{i=0}^{3} \frac{T^{i}}{i!} \boldsymbol{\eta}_{k}^{(i)}+O\left(T^{4}\right)  \tag{17}\\
&=( \mathbf{q} T+\frac{P \mathbf{q}-\mathbf{q} d}{2} T^{2}  \tag{22}\\
&\left.+\frac{P^{2} \mathbf{q}-r_{n-2} \mathbf{q}+\mathbf{q} d^{2}-P \mathbf{q} d}{6} T^{3}\right) y_{k} \\
&+\left\{\left(\frac{\mathbf{q}}{2}+\frac{\mathbf{q} \beta}{6}\right) T^{2} \frac{P \mathbf{q}-\mathbf{q} d}{6} T^{3}\right\} K u_{k}-\frac{\mathbf{q} \beta}{6} T^{2} K u_{k-1} \\
&+\left(I+P T+\frac{P^{2}-\mathbf{q c} c^{T}}{2} T^{2}\right.  \tag{23}\\
&\left.+\frac{P^{3}-\mathbf{q} c^{T} P-P \mathbf{q c} c^{T}+d \mathbf{q} c^{T}}{6} T^{3}\right) \boldsymbol{\eta}_{k}+O\left(T^{4}\right) . \tag{24}
\end{align*}
$$

and the scalars $d_{i}(i=0,1)$ and $\mathbf{c}_{i}(i=0, \ldots, n-3)$ are obtained from

$$
\begin{gathered}
D(s)=Q(s) N(s)+R(s) \\
Q(s)=s^{2}+d_{1} s+d_{0}
\end{gathered}
$$

$$
R(s)=\mathbf{c}_{n-3} s^{n-3}+\cdots+\mathbf{c}_{0}
$$

where

$$
\begin{gathered}
d_{0}=a_{n-2}-b_{n-4}-b_{n-3} d_{1} \\
d_{1}=a_{n-1}-b_{n-3} \\
\mathbf{c}_{i}=a_{i}-b_{i-2}-b_{i-1} d_{1}-b_{i} d_{0}, \quad i=0, \ldots, n-3
\end{gathered}
$$

When a FROH is used, the normal form (19) yields the derivatives of the output

Further, the derivatives of $\boldsymbol{\eta}$ are also represented as

$$
\begin{gather*}
\dot{\boldsymbol{\eta}}=P \boldsymbol{\eta}+\mathbf{q} \xi_{1}  \tag{27}\\
\ddot{\boldsymbol{\eta}}=P \mathbf{q} \xi_{1}+\mathbf{q} \xi_{2}+P^{2} \boldsymbol{\eta}  \tag{28}\\
\boldsymbol{\eta}^{(3)}=\left(P^{2} \mathbf{q}-\mathbf{q} d_{0}\right) \xi_{1}+\left(P \mathbf{q}-\mathbf{q} d_{1}\right) \xi_{2}+\mathbf{q} K u  \tag{29}\\
+\left(P^{3}-\mathbf{q c} \mathbf{c}^{T}\right) \boldsymbol{\eta}
\end{gather*}
$$

$$
\begin{align*}
& \ddot{y}=K u-d_{0} \xi_{1}-d_{1} \xi_{2}-\mathbf{c}^{T} \boldsymbol{\eta}, \\
& y^{(3)}=\left(d_{0} d_{1}-\mathbf{c}^{T} \mathbf{q}\right) \xi_{1}+\left(d_{1}^{2}-d_{0}\right) \xi_{2}-d_{1} K u \\
& +\left(d_{1} c^{T}-c^{T} P\right) \boldsymbol{\eta}+K \dot{u}, \\
& y^{(4)}=\left(d_{0}^{2}-d_{0} d_{1}^{2}+d_{1} \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}\right) \xi_{1} \\
& +\left(-\mathbf{c}^{T} \mathbf{q}+2 d_{0} d_{1}-d_{1}^{3}\right) \xi_{2}  \tag{25}\\
& +\left(d_{1}^{2}-d_{0}\right) K u-d_{1} K \dot{u} \\
& +\left\{-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T}+d_{1} \mathbf{c}^{T} P-\mathbf{c}^{T} P^{2}\right\} \boldsymbol{\eta}, \\
& y^{(5)}=\left(d_{0} d_{1}^{3}-d_{1}^{2} \mathbf{c}^{T} \mathbf{q}+2 \mathbf{c}^{T} \mathbf{q} d_{0}-2 d_{0}^{2} d_{1}\right. \\
& \left.-\mathbf{c}^{T} P^{2} \mathbf{q}+d_{1} \mathbf{c}^{T} P \mathbf{q}\right) \xi_{1} \\
& +\left(d_{0}^{2}-3 d_{0} d_{1}^{2}+2 d_{1} \mathbf{c}^{T} \mathbf{q}\right. \\
& \left.-\mathbf{c}^{T} P \mathbf{q}+d_{1}^{4}\right) \xi_{2}  \tag{26}\\
& +\left(-d_{1}^{3}+2 d_{0} d_{1}-\mathbf{c}^{T} \mathbf{q}\right) K u \\
& +\left\{\mathbf{c}^{T} \mathbf{q} \mathbf{c}^{T}-2 d_{0} d_{1} \mathbf{c}^{T}+d_{1}^{3} \mathbf{c}^{T}\right. \\
& -\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T} P+d_{1} \mathbf{c}^{T} P^{2} \\
& \left.-\mathbf{c}^{T} P^{3}\right\} \boldsymbol{\eta}+\left(d_{1}^{2}-d_{0}\right) K \dot{u} .
\end{align*}
$$

$$
\begin{align*}
\boldsymbol{\eta}^{(4)}= & \left(-P \mathbf{q} d_{0}+\mathbf{q} d_{0} d_{1}+P^{3} \mathbf{q}-\mathbf{q} \mathbf{c}^{T} \mathbf{q}\right) \xi_{1} \\
& +\left(-\mathbf{q} d_{0}+P^{2} \mathbf{q}-P \mathbf{q} d_{1}+\mathbf{q} d_{1}^{2}\right) \xi_{2}  \tag{30}\\
& +\left(P \mathbf{q}-\mathbf{q} d_{1}\right) K u+\mathbf{q K u} \\
& +\left(-P \mathbf{u} \mathbf{c}^{T}+\mathbf{q} d_{1} \mathbf{c}^{T}+P^{4}-\mathbf{q} \mathbf{c}^{T} P\right) \boldsymbol{\eta} .
\end{align*}
$$

Hence, by substituting (23)-(30) into the right-hand side of (16) and

$$
\begin{equation*}
\dot{y}_{k+1}=\sum_{i=0}^{\infty} \frac{T^{i}}{i!} y_{k}^{(i+1)} \tag{31}
\end{equation*}
$$

and defining the state variables $x_{k}=\left[y_{k}, \dot{y}_{k}, \boldsymbol{\eta}_{k}^{T}\right]^{T}$, the discrete-time state equations are obtained.

Now, by using the explicit expressions of $y_{k}, y_{k}^{\prime}, \ldots, y_{k}^{(5)}$ and $\boldsymbol{\eta}_{k}, \ldots, \boldsymbol{\eta}_{k}^{(4)}$, the zeros of the discrete-time system (16) and (31) are analyzed as follows:

$$
\begin{aligned}
y_{k+1}= & \sum_{i=0}^{5} \\
=(1 & T^{i} y_{k}^{(i)}+O\left(T^{6}\right) \\
& \frac{d_{0}}{2} T^{2}+\frac{d_{0} d_{1}-\mathbf{c}^{T} \mathbf{q}_{2}}{6} T^{3} \\
& +\frac{d_{0}^{2}-d_{0} d_{1}^{2}+d_{1} \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}}{24} T^{4} \\
& -\left(\left(2 \mathbf{c}^{T} \mathbf{q} d_{0}-2 d_{0}^{2} d_{1}+d_{0} d_{1}^{3}-d_{1}^{2} \mathbf{c}^{T} \mathbf{q}\right.\right. \\
& \left.\left.\times(120)^{-1}\right) T^{5}\right) y_{k} \\
+ & T-\frac{d_{1}}{2} T^{2}+\frac{d_{1}^{2}-d_{0}}{6} T^{3}+\frac{2 d_{0} d_{1}-d_{1}^{3}-\mathbf{c}^{T} \mathbf{q}}{24} T^{4} \\
& \left.+\frac{\left.d_{0}^{2}-3 d_{0} d_{1}^{2}+2 d_{1} \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}+d_{1}^{4} T^{5}\right) \dot{y}_{k}}{120} \mathbf{q}\right) \\
+ & \left\{\left(\frac{1}{2}+\frac{\beta}{6}\right) T^{2}-\frac{4 d_{1}+d_{1} \beta}{24} T^{3}\right. \\
& +\frac{5 d_{1}^{2}-5 d_{0}+\left(d_{1}^{2}-d_{0}\right) \beta}{120} T^{4} \\
+ & \left.+\frac{2 d_{0} d_{1}-\mathbf{c}^{T} \mathbf{q}-d_{1}^{3}}{120} T^{5}\right\} K u_{k} \\
& \left.\frac{\beta}{6} T^{2}+\frac{d_{1} \beta}{24} T^{3}-\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{120} T^{4}\right\} K u_{k-1}
\end{aligned}
$$

$$
\begin{aligned}
& +\left\{-\frac{c^{T}}{2} T^{2}+\frac{d_{1} c^{T}-\mathbf{c}^{T} P}{6} T^{3}\right. \\
& +\frac{d_{1} \mathbf{c}^{T} P-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T}-\mathbf{c}^{T} P^{2}}{24} T^{4} \\
& +\left(\left(\mathbf{c}^{T} \mathbf{q} \mathbf{c}^{T}-2 d_{0} d_{1} \mathbf{c}^{T}+d_{1}^{3} \mathbf{c}^{T} d_{1} \mathbf{c}^{T} P^{2}\right.\right. \\
& \left.-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T} P-\mathbf{c}^{T} P^{3}\right) \\
& \left.\left.\times(120)^{-1}\right) T^{5}\right\} \boldsymbol{\eta}_{k}+O\left(T^{6}\right), \\
& \dot{y}_{k+1}=\sum_{i=0}^{4} \frac{T^{i}}{i!} y_{k}^{(i+1)}+O\left(T^{5}\right) \\
& =\left(-d_{0} T+\frac{d_{0} d_{1}-\mathbf{c}^{T} \mathbf{q}}{2} T^{2}\right. \\
& +\frac{d_{0}^{2}-d_{0} d_{1}^{2}+d_{1} \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}}{6} T^{3} \\
& +\left(\left(\mathbf{c}^{T} P^{2} \mathbf{q}-d_{1} \mathbf{c}^{T} P \mathbf{q}-2 \mathbf{c}^{T} \mathbf{q} d_{0}\right.\right. \\
& \left.+2 d_{0}^{2} d_{1}-d_{0} d_{1}^{3}+d_{1}^{2} \mathbf{c}^{T} \mathbf{q}\right) \\
& \left.\left.\times(24)^{-1}\right) T^{4}\right) y_{k} \\
& +\left(1-d_{1} T+\frac{d_{1}^{2}-d_{0}}{2} T^{2}+\frac{2 d_{0} d_{1}-d_{1}^{3}-\mathbf{c}^{T} \mathbf{q}}{6} T^{3}\right. \\
& \left.+\frac{d_{0}^{2}-3 d_{0} d_{1}^{2}+2 d_{1} \mathbf{c}^{T} \mathbf{q}-\mathbf{c}^{T} P \mathbf{q}+d_{1}^{4}}{24} T^{4}\right) \dot{y}_{k} \\
& +\left\{\left(1+\frac{\beta}{2}\right) T-\frac{3 d_{1}+d_{1} \beta}{6} T^{2}\right. \\
& +\frac{4 d_{1}^{2}-4 d_{0}+\left(d_{1}^{2}-d_{0}\right) \beta}{24} T^{3} \\
& \left.+\frac{2 d_{0} d_{1}-\mathbf{c}^{T} \mathbf{q}-d_{1}^{3}}{24} T^{4}\right\} K u_{k} \\
& +\left\{-\frac{\beta}{2} T+\frac{d_{1} \beta}{6} T^{2}\right. \\
& \left.-\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{24} T^{3}\right\} K u_{k-1} \\
& +\left\{-\mathbf{c}^{T} T+\frac{d_{1} \mathbf{c}^{T}-\mathbf{c}^{T} P}{2} T^{2}\right. \\
& +\frac{d_{1} \mathbf{c}^{T} P-\mathbf{c}^{T} P^{2}-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T}}{6} T^{3}
\end{aligned}
$$

$$
\begin{align*}
& +\left(\left(\mathbf{c}^{T} \mathbf{q} \mathbf{c}^{T}-2 d_{0} d_{1} \mathbf{c}^{T}+d_{1}^{3} \mathbf{c}^{T}\right.\right. \\
& \left.-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T} P-\mathbf{c}^{T} P^{3}+d_{1} \mathbf{c}^{T} P^{2}\right) \\
& \left.\left.\times(24)^{-1}\right) T^{4}\right\} \boldsymbol{\eta}_{k}+O\left(T^{5}\right) \text {, }  \tag{33}\\
& \boldsymbol{\eta}_{k+1}=\sum_{i=0}^{4} \frac{T^{i}}{i!} \boldsymbol{\eta}_{k}^{(i)}+O\left(T^{5}\right) \\
& =\left(\mathbf{q} T+\frac{P \mathbf{q}}{2} T^{2}+\frac{P^{2} \mathbf{q}-\mathbf{q} d_{0}}{6} T^{3}\right. \\
& \left.+\frac{P^{3} \mathbf{q}-\mathbf{q} c^{T} \mathbf{q}-P \mathbf{q} d_{0}+\mathbf{q} d_{0} d_{1}}{24} T^{4}\right) y_{k} \\
& +\left(\frac{\mathbf{q}}{2} T^{2}+\frac{P \mathbf{q}-\mathbf{q} d_{1}}{6} T^{3}+\frac{P^{2} \mathbf{q}-\mathbf{q} d_{0}}{24} T^{4}\right. \\
& \left.+\frac{-P \mathbf{q} d_{1}+\mathbf{q} d_{1}^{2}}{24} T^{4}\right) \dot{y}_{k}-\frac{\mathbf{q} \beta}{24} T^{3} K u_{k-1} \\
& +\left\{\left(\frac{\mathbf{q}}{6}+\frac{\mathbf{q} \beta}{24}\right) T^{3}+\frac{P \mathbf{q}-\mathbf{q} d_{1}}{24} T^{4}\right\} K u_{k} \\
& +\left(I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}-\mathbf{q c}^{T}}{6} T^{3}\right. \\
& \left.+\frac{P^{4}-P \mathbf{q} c^{T}-\mathbf{q c}^{T} P-P \mathbf{q c}^{T}}{24} T^{4}\right) \boldsymbol{\eta}_{k}+O\left(T^{5}\right) . \tag{34}
\end{align*}
$$

Similarly, the reason why the explicit expressions of $y_{k}, \dot{y}_{k}$, $\ldots, y_{k}^{(5)}$ and $\boldsymbol{\eta}_{k}, \ldots, \boldsymbol{\eta}_{k}^{(4)}$ are used is to obtain the approximate expansion of the limiting zeros for the discrete-time system with the order $T^{3}$ when the relative degree of con-tinuous-time systems is two.

## 3. Main Results

In the following, a more accurate approximate model of the sampled-data system is considered by neglecting the higher order terms, and the approximate expression of the limiting zeros is further calculated in this section.
3.1. Case of Relative Degree One $(m=n-1)$. An approximate expression of limiting zeros for the discrete-time system is derived from (17) and (18). The first result is given by the following theorem.

Theorem 1. The zeros of a discrete-time system corresponding to the continuous-time transfer function (1) with FROH are given for $T \ll 1$ approximately by the roots of

$$
\begin{aligned}
& \left\{\left[-1-\frac{\beta}{2}+\frac{3 d+d \beta}{6} T\right.\right. \\
& \quad-\frac{4\left(d^{2}-r_{n-2}\right)+\left(d^{2}-r_{n-2}\right) \beta}{24} T^{2}
\end{aligned}
$$

$\left.-\frac{2 d r_{n-2}-d^{3}-\mathbf{c}^{T} P \mathbf{q}-5 \mathbf{c}^{T} P \mathbf{q} \beta+2 \mathbf{c}^{T} \mathbf{q} d \beta}{24} T^{3}\right] z$

$$
\begin{align*}
& +\frac{\beta}{2}-\frac{d \beta}{6} T+\frac{\left(d^{2}-r_{n-2}\right) \beta}{24} T^{2} \\
& \left.-\frac{5 \mathbf{c}^{T} P \mathbf{q} \beta-2 \mathbf{c}^{T} \mathbf{q} d \beta}{24} T^{3}\right\} \\
& \times\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{2 P^{3}+\mathbf{q c}^{T} P}{12} T^{3}\right|=0 . \tag{35}
\end{align*}
$$

Proof. The limiting zeros of the discrete-time system (16) are equivalent to zeros in (17) and (18), which are given by substituting $y_{k}=y_{k+1}=0$ into (17) and (18) as follows:

$$
M_{1}\left[\begin{array}{c}
K U_{k-1}  \tag{36}\\
H \\
K U_{k}
\end{array}\right]=\mathbf{0}_{n}
$$

where $U_{k-1}, H$, and $U_{k}$ are the $z$-transforms of $u_{k-1}, \boldsymbol{\eta}_{k}$, and $u_{k}$, respectively, and the matrix $M_{1}$ is defined by

$$
M_{1}=\left[\begin{array}{ccc}
m_{11} & \mathbf{m}_{12}^{T} & m_{13}  \tag{37}\\
-z & \mathbf{0}^{T} & 1 \\
\mathbf{m}_{31} & M_{32} & \mathbf{m}_{33}
\end{array}\right],
$$

with

$$
\begin{aligned}
m_{11}= & T \bar{m}_{11}+O\left(T^{4}\right) \\
\mathbf{m}_{12}^{T}= & T \overline{\mathbf{m}}_{12}^{T}+O\left(T^{5}\right), \\
m_{13}= & T \bar{m}_{13}+O\left(T^{5}\right), \\
\bar{m}_{11}= & -\frac{\beta}{2}+\frac{d \beta}{6} T-\frac{\left(d^{2}-r_{n-2}\right) \beta}{24} T^{2}, \\
\overline{\mathbf{m}}_{12}^{T}= & -\mathbf{c}^{T}+\frac{d \mathbf{c}^{T}-\mathbf{c}^{T} P}{2} T+\frac{d \mathbf{c}^{T} P-\mathbf{c}^{T} P^{2}}{6} T^{2} \\
& +\frac{\left(r_{n-2}-d^{2}\right) \mathbf{c}^{T}}{6} T^{2}+\frac{\left(d^{3}-2 d r_{n-2}\right) \mathbf{c}^{T}}{24} T^{3} \\
& +\frac{\mathbf{c}^{T} P \mathbf{q} \mathbf{c}^{T}+d \mathbf{c}^{T} P^{2}-\mathbf{c}^{T} P^{3}}{24} T^{3} \\
& +\frac{-d^{2} \mathbf{c}^{T} P+r_{n-2} \mathbf{c}^{T} P}{24} T^{3}, \\
\bar{m}_{13}= & \left(1+\frac{\beta}{2}\right)-\left(\frac{d}{2}+\frac{d \beta}{6}\right) T \\
& +\left\{\frac{d^{2}-r_{n-2}}{6}+\frac{\left(d^{2}-r_{n-2}\right) \beta}{24}\right\} T^{2} \\
& +\frac{-d^{3}+2 d r_{n-2}-\mathbf{c}^{T} P \mathbf{q}^{3} T^{3},}{24}
\end{aligned}
$$

$$
\begin{align*}
\mathbf{m}_{31}= & -\frac{\mathbf{q} \beta}{6} T^{2}+O\left(T^{4}\right) \\
M_{32}= & (-z+1) I+P T+\frac{P^{2}-\mathbf{q c} c^{T}}{2} T^{2} \\
& +\frac{d \mathbf{q c} c^{T}-P \mathbf{q} c^{T}+P^{3}-\mathbf{q} \mathbf{c}^{T} P}{6} T^{3}+O\left(T^{4}\right) \\
\mathbf{m}_{33}= & \left(\frac{\mathbf{q}}{2}+\frac{\mathbf{q} \beta}{6}\right) T^{2}+\frac{P \mathbf{q}-\mathbf{q} d}{6} T^{3}+O\left(T^{4}\right) \tag{38}
\end{align*}
$$

Thus, the zeros are derived from

$$
\begin{equation*}
\left|M_{1}\right|=0 \tag{39}
\end{equation*}
$$

From the relationship

$$
M=M_{1}\left[\begin{array}{lll}
1 & 0 & 0  \tag{40}\\
0 & 1 & 0 \\
1 & 0 & 1
\end{array}\right]
$$

it is obvious that the condition $|M|=0$ is equivalent to $\left|M_{1}\right|=0$.

Expanding the result along the second row leads to the following equation:

$$
\begin{align*}
|M|= & -(-z+1)\left|\begin{array}{ll}
\mathbf{m}_{12}^{T} & m_{13} \\
M_{32} & \mathbf{m}_{33}
\end{array}\right| \\
& -\left|\begin{array}{ll}
m_{11}+m_{13} & \mathbf{m}_{12}^{T} \\
\mathbf{m}_{31}+\mathbf{m}_{33} & M_{32}
\end{array}\right|  \tag{41}\\
& =T\left[-z\left|A_{1}\right|+(z-1)\left|A_{2}\right|\right]=0,
\end{align*}
$$

where

$$
\begin{aligned}
\left|A_{1}\right|= & {\left[1-\frac{d}{2} T+\frac{d^{2}-r_{n-2}}{6} T^{2}\right.} \\
& \left.+\frac{2 d r_{n-2}-d^{3}-\mathbf{c}^{T} p \mathbf{q}}{24} T^{3}\right] \\
& \times\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{2 P^{3}+\mathbf{q c}^{T} P}{12} T^{3}\right| \\
= & \Delta_{1} \times \Delta_{2}, \\
\left|A_{2}\right|= & {\left[-\frac{\beta}{2}+\frac{d \beta}{6} T-\frac{\left(d^{2}-r_{n-2}\right) \beta}{24} T^{2}\right.} \\
& \left.+\left(\frac{5 \mathbf{c}^{T} p \mathbf{q} \beta-2 \mathbf{c}^{T} \mathbf{q} d \beta}{24}\right) T^{3}\right] \\
& \times\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{2 P^{3}+\mathbf{q} \mathbf{c}^{T} P}{12} T^{3}\right| \\
= & \bar{\Delta}_{1} \times \Delta_{2} .
\end{aligned}
$$

Then,

$$
\begin{equation*}
|M|=\left[\left(\bar{\Delta}_{1}-\Delta_{1}\right) z-\bar{\Delta}_{1}\right] \times \Delta_{2} \tag{43}
\end{equation*}
$$

Hence, the approximate values of limiting zeros of the dis-crete-time system are obtained as the roots of (35).

Remark 2. Equation (35) implies that an approximation of the sampling zero is expressed as

$$
\begin{align*}
& {\left[-1-\frac{\beta}{2}+\left(\frac{d}{2}+\frac{d \beta}{6}\right) T\right.} \\
& \quad-\left(\frac{d^{2}-r_{n-2}}{6}+\frac{\left(d^{2}-r_{n-2}\right) \beta}{24}\right) T^{2} \\
& \left.+\frac{2 d r_{n-2}-d^{3}-\mathbf{c}^{T} P \mathbf{q}-5 \mathbf{c}^{T} P \mathbf{q} \beta+2 \mathbf{c}^{T} \mathbf{q} d \beta}{24} T^{3}\right] z  \tag{44}\\
& \quad+\frac{\beta}{2}-\frac{d \beta}{6} T+\frac{\left(d^{2}-r_{n-2}\right) \beta}{24} T^{2} \\
& \quad-\frac{5 \mathbf{c}^{T} P \mathbf{q} \beta-2 \mathbf{c}^{T} \mathbf{q} d \beta}{24} T^{3}=0
\end{align*}
$$

and the approximate values of the intrinsic zeros are derived from

$$
\begin{equation*}
\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{2 P^{3}+\mathbf{q c}^{T} P}{12} T^{3}\right|=0 . \tag{45}
\end{equation*}
$$

Remark 3. Theorem 1 is applicable to also the case of multiple zeros of the continuous-time system (1) with FROH and further gives approximate values with higher order of accuracy than those of the previous result [17].

Remark 4. An insightful observation in Theorem 1 is that it has a form of a correction to the asymptotic result of Ishitobi $[12,18]$ in the form of a power term of $T$. Similarly, the following result (Theorem 6) both the intrinsic zeros and sampling zeros is also clarified in a more precise manner than Ishitobi's result $[12,17,18]$ when the relative degree of continuous-time systems is two.

Remark 5. On the basis of the approach in [21], it is immediate to derive the asymptotic condition of the limiting zeros in the case of a FROH with relative degree one:

$$
\begin{aligned}
\left|M_{1}\right| \approx & \left\{\left(1+\frac{\beta}{2}\right)-\left(\frac{d}{2}+\frac{d \beta}{6}\right) T\right. \\
& +\frac{4 d^{2}-4 r_{n-2}+\left(d^{2}-r_{n-2}\right) \beta}{24} T^{2} \\
& \left.+\frac{2 d r_{n-2}-d^{3}-c^{T} P \mathbf{q}}{24} T^{3}\right\} \\
\times & {\left[-z+1-\frac{2}{2+\beta}+\frac{d \beta}{3(2+\beta)^{2}} T\right.} \\
& +\frac{\beta\left(6 r_{n-2}+3 r_{n-2} \beta-\beta d^{2}\right)}{18(2+\beta)^{3}} T^{2}
\end{aligned}
$$

$$
\begin{gather*}
+\left(\left(\beta \left(1014 d \beta r_{n-2}-348 d^{3}\right.\right.\right. \\
\left.+348 d r_{n-2}\right) \\
+\beta\left(1188 d \beta^{2} r_{n-2}-1026 d^{3} \beta\right. \\
\left.-1193 d^{3} \beta^{2}\right) \\
+\beta\left(36 \mathbf{c}^{T} P \mathbf{q} \beta+36 \mathbf{c}^{T} P \mathbf{q}\right. \\
\left.\left.+9 \mathbf{c}^{T} P \mathbf{q} \beta^{2}\right)\right) \\
\left.\left.\times\left(108(2+\beta)^{4}\right)^{-1}\right) T^{3}\right] \\
\times\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{2 P^{3}+\mathbf{q c}^{T} P}{12} T^{3}\right| . \tag{46}
\end{gather*}
$$

When the relative degree of continuous-time systems is one and the continuous-time input is generated by a FROH , further research is needed to establish connections between (46) and (35) of Theorem 1 in this paper, wherein the idea (35) has more decent effect than the literature [21] in terms of techniques in studying the discrete system zeros.
3.2. Case of Relative Degree Two ( $m=n-2$ ). Next, we present asymptotic properties of limiting zeros of discrete-time control system in the case of a FROH as power series with respect to a sampling period up to the third-order term when the relative degree of the continuous-time system is two. An approximate expression, in fact, of zeros of a discrete-time system is derived from (32)-(34), and the other results of this paper are given by the following Theorem.
Theorem 6. The zeros of a discrete-time system for the continuous-time transfer function (1) with $F R O H$ are given for $T \ll 1$ approximately by the roots of

$$
\begin{aligned}
& \left\{\left[\frac{1}{2}+\frac{\beta}{6}-\left(\frac{4 d_{1}+d_{1} \beta}{24}\right) T+\frac{5 d_{1}^{2}-5 d_{0}}{120} T^{2}\right.\right. \\
& \left.\quad+\frac{\beta d_{1}^{2}-\beta d_{0}}{120} T^{2}+\frac{2 d_{0} d_{1}-\mathbf{c}_{n-3}-d_{1}^{3}}{120} T^{3}\right] z^{2} \\
& +\left[\frac{3+\beta}{6}-\frac{2 d_{1}+d_{1} \beta}{6} T\right. \\
& \quad+\frac{15 d_{1}^{2}-5 d_{0}+8 \beta d_{1}^{2}-3 \beta d_{0}}{120} T^{2} \\
& \quad+\left(\frac{d_{0} d_{1}}{40}+\frac{\mathbf{c}_{n-3}}{120}-\frac{d_{1}^{3}}{30}+\frac{d_{0} d_{1} \beta}{180}\right. \\
& \left.\left.\quad+\frac{\mathbf{c}_{n-3} \beta}{72}-\frac{d_{1}^{3} \beta}{80}\right) T^{3}\right] z
\end{aligned}
$$

$$
\begin{align*}
& -\frac{\beta}{3}+\frac{5 d_{1} \beta}{24} T+\frac{4 d_{0} \beta-9 d_{1}^{2} \beta}{120} T^{2} \\
& \left.+\frac{9 d_{1}^{3}-4 d_{0} d_{1} \beta-10 \mathbf{c}_{n-3} \beta}{720} T^{3}\right\} \\
& \times\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3}\right|=0 \tag{47}
\end{align*}
$$

Proof. Zeros of the discrete-time system (16) and (31), equivalent to (32)-(34), are given by substituting $y_{k}=y_{k+1}=0$ into (32)-(34) as follows:

$$
M_{2}\left[\begin{array}{c}
Y_{d}  \tag{48}\\
K U_{k-1} \\
H \\
K U_{k}
\end{array}\right]=\mathbf{0}_{n}
$$

where $Y_{d}$ is the $z$-transforms of $\dot{y}_{k}$ and the matrix $M_{2}$ is defined by

$$
M_{2}=\left[\begin{array}{cccc}
m_{11} & m_{12} & \mathbf{m}_{13}^{T} & m_{14}  \tag{49}\\
m_{21} & m_{22} & \mathbf{m}_{23}^{T} & m_{24} \\
0 & -z & \mathbf{0}^{T} & 1 \\
\mathbf{m}_{41} & \mathbf{m}_{42} & M_{43} & \mathbf{m}_{44}
\end{array}\right]
$$

with

$$
\begin{aligned}
m_{11}= & T \bar{m}_{11}+O\left(T^{5}\right) \\
m_{12}= & T \bar{m}_{12}+O\left(T^{5}\right) \\
\mathbf{m}_{13}^{T}= & T \overline{\mathbf{m}}_{13}^{T}+O\left(T^{5}\right), \\
m_{14}= & T \bar{m}_{14}+O\left(T^{5}\right) \\
\bar{m}_{11}= & 1-\frac{d_{1}}{2} T+\frac{d_{1}^{2}-d_{0}}{6} T^{2} \\
& +\frac{-d_{1}^{3}+2 d_{0} d_{1}-\mathbf{c}_{n-3}}{24} T^{3} \\
\bar{m}_{12}= & -\frac{\beta}{6} T+\frac{d_{1} \beta}{24} T^{2}-\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{120} T^{3} \\
\overline{\mathbf{m}}_{13}^{T}= & -\frac{\mathbf{c}^{T}}{2} T+\frac{d_{1} \mathbf{c}^{T}-\mathbf{c}^{T} P}{6} T^{2} \\
& +\frac{-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T}+d_{1} \mathbf{c}^{T} P-\mathbf{c}^{T} P^{2}}{24} T^{3}, \\
\bar{m}_{14}= & \left(\frac{1}{2}+\frac{\beta}{6}\right) T+\left(-\frac{d_{1}}{6}-\frac{d_{1} \beta}{24}\right) T^{2} \\
& +\left[\frac{d_{1}^{2}-d_{0}}{24}+\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{120}\right] T^{3},
\end{aligned}
$$

$$
\begin{align*}
m_{21}= & -z+1-d_{1} T+\frac{d_{1}^{2}-d_{0}}{2} T^{2} \\
& +\frac{-d_{1}^{3}+2 d_{0} d_{1}-\mathbf{c}_{n-3}}{6} T^{3}+O\left(T^{4}\right), \\
m_{22}= & -\frac{\beta}{2} T+\frac{d_{1} \beta}{6} T^{2}-\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{24} T^{3}+O\left(T^{4}\right), \\
\mathbf{m}_{23}^{T}= & -\mathbf{c}^{T} T+\frac{d_{1} \mathbf{c}^{T}-\mathbf{c}^{T} P}{2} T^{2} \\
& +\frac{d_{1} \mathbf{c}^{T} P-\left(d_{1}^{2}-d_{0}\right) \mathbf{c}^{T}-\mathbf{c}^{T} P^{2}}{6} T^{3}+O\left(T^{4}\right), \\
m_{24}= & \left(1+\frac{\beta}{2}\right) T+\left(-\frac{d_{1}}{2}-\frac{d_{1} \beta}{6}\right) T^{2} \\
& +\left(\frac{d_{1}^{2}-d_{0}}{6}+\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{24}\right) T^{3}+O\left(T^{4}\right), \\
\mathbf{m}_{41}= & \frac{\mathbf{q}}{2} T^{2}+\frac{P \mathbf{q}-\mathbf{q} d_{1}}{6} T^{3}+O\left(T^{4}\right), \\
\mathbf{m}_{42}= & -\frac{\mathbf{q} \beta}{24} T^{3}+O\left(T^{4}\right), \\
M_{43}= & (-z+1) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}-\mathbf{q c} c^{T}}{6} T^{3} \\
& +O\left(T^{4}\right) \\
\mathbf{m}_{44}= & \left(\frac{\mathbf{q}}{6}+\frac{\mathbf{q} \beta}{24}\right) T^{3}+O\left(T^{4}\right) . \tag{50}
\end{align*}
$$

Thus, the zeros are derived from

$$
\begin{equation*}
\left|M_{2}\right|=0 \tag{51}
\end{equation*}
$$

From the relationship

$$
\bar{M}=M_{2}\left[\begin{array}{llll}
1 & 0 & 0 & 0  \tag{52}\\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1
\end{array}\right]
$$

it is obvious that the condition $|\bar{M}|=0$ is equivalent to $\left|M_{2}\right|=0$.

Expanding the result along the third row leads to the following equation:

$$
\begin{aligned}
|\bar{M}|= & -(-z+1)\left|\begin{array}{lll}
m_{11} & \mathbf{m}_{13}^{T} & m_{14} \\
m_{21} & \mathbf{m}_{23}^{T} & m_{24} \\
\mathbf{m}_{41} & M_{43} & \mathbf{m}_{44}
\end{array}\right| \\
& -\left|\begin{array}{lll}
m_{11} & m_{12}+m_{14} & \mathbf{m}_{13}^{T} \\
m_{21} & m_{22}+m_{24} & \mathbf{m}_{23}^{T} \\
\mathbf{m}_{41} & \mathbf{m}_{42}+\mathbf{m}_{44} & M_{43}
\end{array}\right| \\
& =T\left[-z\left|\bar{A}_{1}\right|+(-z+1)\left|\bar{A}_{2}\right|\right]=0,
\end{aligned}
$$

where

$$
\begin{align*}
&\left|\bar{A}_{1}\right|= {\left[\frac{1}{2}-\frac{d_{1}}{6} T+\frac{d_{1}^{2}-d_{0}}{24} T^{2}\right.} \\
&\left.+\frac{-\mathbf{c}_{n-3}+2 d_{0} d_{1}-d_{1}^{3}}{120} T^{3}\right] \\
& \times\left[-z-1+\frac{d_{1}}{3} T-\frac{d_{1}^{2}}{18} T^{2}\right.  \tag{54}\\
&\left.+\frac{d_{1}^{3}+3 d_{0} d_{1}-9 \mathbf{c}_{n-3}}{270} T^{3}\right] \\
&= \Delta_{1} \times \Delta_{2} \times \Delta_{3}, \\
&\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3}\right| \\
&\left|\bar{A}_{2}\right|=\left(\frac{\beta}{6} T-\frac{d_{1} \beta}{24} T^{2}+\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{120} T^{3}\right) \\
& \times\left[-z-2+\frac{3 d_{1}}{4} T+\left(-\frac{13 d_{1}^{2}}{80}+\frac{d_{0}}{10}\right) T^{2}\right. \\
&\left.+\left(\frac{7 d_{0} d_{1}}{240}-\frac{3 d_{1}^{3}}{960}-\frac{\mathbf{c}_{n-3}}{12}\right) T^{3}\right]  \tag{55}\\
& \times\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3}\right| \\
&= \bar{\Delta}_{1} \times \bar{\Delta}_{2} \times \Delta_{3} .
\end{align*}
$$

Equations (54) and (55) will be calculated in the appendix. Then,

$$
\begin{equation*}
|M|=\left[\left(-\bar{\Delta}_{1} \times \bar{\Delta}_{2}-\Delta_{1} \times \Delta_{2}\right) z+\bar{\Delta}_{1} \times \bar{\Delta}_{2}\right] \times \Delta_{3} \tag{56}
\end{equation*}
$$

Hence, the approximate values of the zeros of the discretetime system are obtained as the roots of (47).

Remark 7. Equation (47) implies that the approximations of the sampling zeros are expressed as

$$
\begin{aligned}
& {\left[\frac{1}{2}+\frac{\beta}{6}-\left(\frac{d_{1}}{6}+\frac{d_{1} \beta}{24}\right) T+\frac{5 d_{1}^{2}-5 d_{0}+\beta d_{1}^{2}-\beta d_{0}}{120} T^{2}\right.} \\
& \left.\quad+\frac{2 d_{0} d_{1}-\mathbf{c}_{n-3}-d_{1}^{3}}{120} T^{3}\right] z^{2} \\
& \quad+\left[\frac{3+\beta}{6}-\frac{2 d_{1}+d_{1} \beta}{6} T\right. \\
& \quad+\frac{15 d_{1}^{2}-5 d_{0}+8 \beta d_{1}^{2}-3 \beta d_{0}}{120} T^{2}
\end{aligned}
$$

$$
\begin{gather*}
+\left(\frac{d_{0} d_{1}}{40}+\frac{\mathbf{c}_{n-3}}{120}-\frac{d_{1}^{3}}{30}+\frac{d_{0} d_{1} \beta}{180}\right. \\
\left.\left.+\frac{\mathbf{c}_{n-3} \beta}{72}-\frac{d_{1}^{3} \beta}{80}\right) T^{3}\right] z \\
-\frac{\beta}{3}+\frac{5 d_{1} \beta}{24} T+\frac{4 d_{0} \beta-9 d_{1}^{2} \beta}{120} T^{2} \\
+\frac{9 d_{1}^{3}-4 d_{0} d_{1} \beta-10 \mathbf{c}_{n-3} \beta}{720} T^{3}=0 \tag{57}
\end{gather*}
$$

and the approximate values of the intrinsic zeros are derived from

$$
\begin{equation*}
\left|(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3}\right|=0 \tag{58}
\end{equation*}
$$

Remark 8. When FROH is implemented in practice, an approximate fractional-order hold (AFROH) using ZOH would be convenient practical solution. The basic idea of AFROH is that, at each sampling interval, the output of FROH is approximated by staircase waveforms that can be generated by ZOH [15, 22] (see Figure 2). Therefore, an asymptotic expression of the limiting zeros in AFROH case is derived similarly.

In the particular case when the sampling period tends to zero, it is immediate to obtain the following Corollary although a similar result is also obtained by Ishitobi [12, 18].

## Corollary 9. One has the following cases.

Case $a$. Assume that the relative degree of a continuous-time system is one. If $-1<\beta$ (resp., $\beta<-1$ ), then the sampling zero of the sampled-data model is stable (resp., unstable) in the case of a FROH when the sampling period tends to zero.

Case b. Assume that the relative degree of a continuous-time system is two. If $-1<\beta<0$ (resp., $\beta \leq-1$ or $\beta>0$ ), then the sampling zeros of the sampled-data model are stable (resp., unstable) in the case of a FROH when the sampling period tends to zero.

Proof. One has the following cases.
Case $a$. For $n-m=1$, we have from (35)

$$
\begin{equation*}
A_{1}(z ; \beta)=\left(-1-\frac{\beta}{2}\right) z+\frac{\beta}{2} \tag{59}
\end{equation*}
$$

Simple straightforward calculation will verify that the root of $A_{1}(z ; \beta)=0$ is stable if $-1<\beta$ and is unstable if $\beta \leq$ -1 . In addition, when $\beta=-1$, there remains the possibility


Figure 2: Output of the approximate fractional-order hold with $N=$ 2 and $\beta=-0.5$.
that the sampling zero corresponding to $A_{1}(z ; \beta)$, which approaches $z=-1$, lies inside the open unit disc as $T \rightarrow 0$.

Case $b$. For $n-m=2$, the polynomial $A_{2}(z ; \beta)$ is represented from (47) as

$$
\begin{equation*}
A_{2}(z ; \beta)=\left(\frac{1}{2}+\frac{\beta}{6}\right) z^{2}+\left(\frac{1}{2}+\frac{\beta}{6}\right) z-\frac{\beta}{3} . \tag{60}
\end{equation*}
$$

When we perform the bilinear transformation $z=(\omega+$ $1) /(\omega-1)$ on the above equation, the polynomial is written as

$$
\begin{equation*}
A_{2}(z ; \beta)=3 \omega^{2}+3(1+\beta) \omega-\beta=0 \tag{61}
\end{equation*}
$$

It is clear that the two roots of (61) lie in the open left half of $\omega$-plane if $-1<\beta<0$, and at least one of them stays in the closed right-half plane if $\beta \leq-1$ or $\beta \geq 0$. In particular, only one of the sampling zeros approaches -1 at $\beta=0$. Namely, the stability of the sampling zeros is marginal, that is, in the case of a $\mathrm{ZOH}[4,23]$.

Remark 10. When the FROH signal reconstruction device is used, the parameter $\beta$ which is the device adjustable gain (generalised gain) is the major factor that decides the stability properties of sampling zeros of sampled-data systems with FROH. In other words, the appropriate $\beta$ is determined to obtain the FROH that provides sampling zeros as stable as possible, or with improved stability properties even when being unstable, for a given continuous-time plant.

Remark 11. If the relative degree of a continuous-time transfer function is two and the sum of the zeros is less than or equal to the sum of the poles, the limiting zeros of the sampled system with FROH of $-1<\beta<0$ stay definitely inside the unit circle while those with ZOH may lie outside or on the unit circle. Therefore, the FROH with $-1<\beta<0$ will produce all stable sampling zeros for a wider class of continuous-time plants than that of the ZOH .

## 4. Simulation Examples

This section presents three interesting examples to show the stability of sampling zeros with FROH by improved

Table 1: Intrinsic zeros of the sampled-data system with relative degree one.

| $T$ | Approximate values (35) | Exact values |
| :--- | :---: | :---: |
| 0.01 | 0.980311498, | 0.980310411, |
|  | 0.980082836 | 0.980084008 |
| 0.02 | 0.961230229, | 0.961222048, |
|  | 0.960324437 | 0.960333962 |
| 0.05 | 0.907398436, | 0.907292742, |
|  | 0.901893230 | 0.902049060 |
| 0.1 | 0.827729968, | 0.827130532, |
|  | 0.806603365 | 0.807949540 |
| 0.2 | 0.697266611, | 0.695230664, |
|  | 0.617400056 | 0.629936753 |

Table 2: Sampling zero of the sampled-data system with relative degree one and $\beta=-1 / 2$.

| $T$ | Approximate values (35) | Exact values |
| :--- | :---: | :---: |
| 0.01 | 0.329701686 | 0.329657909 |
| 0.02 | 0.326139088 | 0.325821314 |
| 0.05 | 0.316359822 | 0.314371422 |
| 0.1 | 0.302545156 | 0.294211129 |
| 0.2 | 0.283625731 | 0.248072696 |

TABLE 3: Sampling zero of the sampled-data system with relative degree one and $\beta=1$.

| $T$ | Approximate values (35) | Exact values |
| :--- | :---: | :---: |
| 0.01 | -0.331507563 | -0.331319933 |
| 0.02 | -0.329739729 | -0.329258964 |
| 0.05 | -0.324624858 | -0.322146285 |
| 0.1 | -0.317193888 | -0.307047943 |
| 0.2 | -0.306446658 | -0.255131263 |

asymptotic properties. It has also shown that the stability of zeros will be improved by using FROH instead of ZOH. Both kinds of zeros are calculated by applying MATLAB, and in the simulation figures (Figures 3, 4, 5, 6, 7, 8, and 9), the solid line and dotted line indicate the exact values and approximate values, respectively.

Example 1. Consider the following transfer function with the relative degree one [21]:

$$
\begin{equation*}
G(s)=\frac{(s+2)^{2}}{s(s+1)(s-2)} \tag{62}
\end{equation*}
$$

The approximate values (35) and the exact values of zeros of the sampled-data system for the transfer function (62) are shown in Tables 1-4 and corresponding figures, where the intrinsic zeros are shown in Table 1 and the sampling zero is respectively shown in Tables 2, 3, and 5 owing to the difference of the parameter $\beta$. Equation (35) gives good approximation also for the case of a continuous-time transfer function with FROH.

Table 4: Sampling zero of the sampled-data system with relative degree one.

| $T$ | Approximate values (46) <br> $\beta=-1 / 2$ | Approximate values (46) <br>  <br> 0.01 |
| :--- | :---: | :---: |
| 0.3297 | -0.3315 |  |
| 0.02 | 0.3262 | -0.3297 |
| 0.05 | 0.3164 | -0.3246 |
| 0.1 | 0.3025 | -0.3169 |
| 0.2 | 0.2841 | -0.3046 |

Table 5: Sampling zero of the sampled-data system with relative degree one and $\beta=-2$.

| $T$ | Approximate values (35) | Exact values |
| :--- | :---: | :---: |
| 0.01 | 121.047619 | 120.640978703 |
| 0.02 | 60.80588235 | 60.607244462 |
| 0.05 | 24.85125858 | 24.530135132 |
| 0.1 | 13.10222222 | 12.387624318 |
| 0.2 | 7.603157322 | 6.1116305864 |

When the continuous-time systems have relative degree one, a discrete-time system corresponding to a continuoustime transfer function (62) has two intrinsic zeros and one sampling zero in the case of a FROH. In particular, the values of the intrinsic zeros with FROH are approximately equal to those with ZOH owing to the parameter $\beta$ (see also Remark 13). Further, the stability of sampling zero with FROH depends on the parameter $\beta$. When $-1<\beta$ (resp., $\beta<-1$ ), the sampling zero of the sampled-data model is stable (resp., unstable) in the case of a FROH for small sampling periods (see Tables 2-4).

Case a $(\beta=-1 / 2)$. See Table 2 and Figure 5.
Case $b(\beta=1)$. See Table 3 and Figure 6.
Case c $(\beta=-2)$. See Table 5. From the foregoing analysis, it is obvious that the limiting zeros of the sampled-data system with FROH of $-1<\beta$ are located inside the unit circle. In addition, (46) gives good approximation and the sampling zero lies inside the unit circle for small sampling periods at $\beta=-1 / 2$ and $\beta=1$ (see Table 4). Furthermore, it can be seen from the corresponding Tables $2-4$ that (35) can offer a more accurate approximation than that of (46) in terms of the stable sampling zero of discrete-time model.

Example 2. Consider a transfer function with the relative degree two

$$
\begin{equation*}
G(s)=\frac{s+7}{(s+1)(s+2)(s+3)} \tag{63}
\end{equation*}
$$

On the basis of the results in [21, 24], the stability condition of sampling zeros with ZOH is dissatisfied since $d_{1}=$ $a_{2}-b_{0}=-1<0$. However, the stability of the sampling zeros will be preserved in the case of a FROH when $-1<\beta<0$. The approximate values (47) and the exact values of zeros of the sampled-data system for the transfer function (63) are


Figure 3: Intrinsic zero of sampled-data model with relative degree one.


Figure 4: Intrinsic zero of sampled-data model with relative degree one.

Table 6: Intrinsic zero of the sampled-data system with relative degree two.

| $T$ | Approximate values (47) | Exact values |
| :--- | :---: | :---: |
| 0.01 | 0.9324 | 0.932347819 |
| 0.02 | 0.8693 | 0.869400472 |
| 0.05 | 0.7041 | 0.704589951 |
| 0.1 | 0.4878 | 0.496541723 |
| 0.2 | 0.1227 | 0.244801539 |

shown in Tables 6-9, where the intrinsic zero is shown in Table 6 and the sampling zeros are shown in Tables 7, 8, and 9. Equation (47) gives good approximation and the sampling zeros lie inside the unit circle for small sampling periods with FROH, while ZOH fails to do so.


Figure 5: Sampling zero of sampled-data model with relative degree one and $\beta=-1 / 2$.


Figure 6: Sampling zero of sampled-data model with relative degree one and $\beta=1$.

Remark 12. From Examples 1 and 2, it can be obviously seen that FROH is reduced to FOH for $\beta=1$. The limiting zeros for sufficiently small $T$ in the case of a FOH are stable with relative degree one while it is unstable with relative degree two. Thus, a FOH provides no advantage over ZOH and FROH with the stability of the limiting zeros [4].

Remark 13. When the FROH signal reconstruction device is used, the parameter $\beta$, so called the device adjustable gain (generalised gain), is also a factor which affects the intrinsic zeros of sampled-data systems by numerically verifying in the case of a FROH. More precisely, it only affects the distribution of intrinsic zeros while the stability of intrinsic zeros is still preserved for different values of $\beta$. See also the literature by De la Sen [25], who has similar conclusion by applying different technique.

Next, we display the improvement of the asymptotic properties of discrete system zeros with FROH through an

Table 7: Sampling zeros of the sampled-data system with relative degree two and $\beta=-1 / 2$.

| $T$ | Approximate values (47) | Exact values |
| :--- | :---: | :---: |
| 0.01 | $-0.501195886-0.387172957 i$, | $-0.501218099-0.386984162 i$, |
| 0.02 | $-0.501195886+0.387172957 i$ | $-0.501218099+0.386984162 i$ |
|  | $-0.502623421-0.386611476 i$, | $-0.502596060-0.386768672 i$, |
| 0.05 | $-0.502623421+0.386611476 i$ | $-0.502596060+0.386768672 i$ |
|  | $-0.507102272-0.384595337 i$, | $-0.506565882-0.385411371 i$, |
| 0.1 | $-0.507102272+0.384595337 i$ | $-0.506565882+0.385411371 i$ |
|  | $-0.516933208-0.375577576 i$, | $-0.512704082-0.381808431 i$, |
| 0.2 | $-0.516933208+0.375577576 i$ | $-0.512704082+0.381808431 i$ |
|  | $-0.553360151-0.309177835 i$, | $-0.519378658-0.371599469 i$, |

Table 8: Sampling zeros of the sampled-data system with relative degree two and $\beta=1$.

| $T$ | Approximate values (47) | Exact values |
| :--- | :---: | :---: |
| 0.01 | $0.366015522,-1.370352295$ | $0.366029257,-1.370288147$ |
| 0.02 | $0.365962901,-1.374613162$ | $0.365977015,-1.374587969$ |
| 0.05 | $0.365635742,-1.386524631$ | $0.365876959,-1.385974102$ |
| 0.1 | $0.364115478,-1.403715066$ | $0.364513566,-1.401081359$ |
| 0.2 | $0.353215369,-1.422680342$ | $0.369015125,-1.407272983$ |



Figure 7: Intrinsic zero of sampled-data model with relative degree two.
example of an electronic circuit in the remainder of this section.

Example 3. Consider an electric circuit shown in Figure 10 [15], where $R_{i}(i=1, \ldots, 4)$ and $C_{j}(j=1, \ldots, 3)$ represent resistance and condenser, respectively.

The transfer function with the voltage $e_{i}(t)$ as an input and with the voltage $e_{0}(t)$ as an output is given by

$$
\begin{equation*}
G(s)=\frac{\sum_{k=0}^{1} b_{k} s^{k}}{\sum_{\ell=0}^{3} a_{\ell} s^{\ell}}=K \frac{s+\bar{b}_{0}}{s^{3}+\bar{a}_{2} s^{2}+\bar{a}_{1} s+\bar{a}_{0}} \tag{64}
\end{equation*}
$$



Figure 8: Sampling zero of sampled-data model with relative degree two and $\beta=-1 / 2$.
where

$$
\begin{equation*}
K=\frac{b_{1}}{a_{3}}, \quad \bar{b}_{0}=\frac{b_{0}}{b_{1}}, \quad \bar{a}_{0}=\frac{a_{0}}{a_{3}}, \quad \bar{a}_{1}=\frac{a_{1}}{a_{3}}, \quad \bar{a}_{2}=\frac{a_{2}}{a_{3}} . \tag{65}
\end{equation*}
$$

It is easy to see that the relative degree of transfer function (64) is two. Here, when the parameters are set as $R_{1}=R_{2}=$ $1[\mathrm{k} \Omega], R_{3}=5[\mathrm{k} \Omega], R_{4}=13 / 70[\mathrm{k} \Omega]$, and $C_{1}=C_{2}=$ $C_{3}=1[\mu \mathrm{~F}]$, the corresponding discrete-time system with ZOH has an unstable sampling zero for the sufficiently small sampling periods according to the $\bar{a}_{2}-\bar{b}_{0}<0$ [21, 24]. In fact, the absolute value of the sampling zero of the discretetime system with ZOH is 1.006418 for $T=0.001$. The magnitudes of limiting zeros of the corresponding discrete-time system with FROH are shown in Figure 11 for the sampling period $T=0.001$. All the limiting zeros stay inside the unit circle for $-1<\beta<0$ (see also Figure 11). The stability condition can be achieved by means of a suitable choice of the parameter $\beta$ of the improving asymptotic properties.

Remark 14. From Example 3, it has been shown that the limiting zeros of the sampled-data models with FROH can be located inside the stability region by analyzing the improved asymptotic properties while ZOH fails to do so. In addition,

TABLE 9: Sampling zeros of the sampled-data system with relative degree two and $\beta=-2$.

| $T$ | Approximate values (47) | Exact values |
| :--- | :---: | :---: |
| 0.01 | $-0.497611941+1.938185095 i$, | $-0.497685781+1.938246055 i$, |
| 0.02 | $-0.497611941-1.938185095 i$ | $-0.497685781-1.938246055 i$ |
|  | $-0.495838288+1.940321092 i$, | $-0.495813965+1.940339611 i$, |
| 0.05 | $-0.495838288-1.940321092 i$ | $-0.495813965-1.940339611 i$ |
|  | $-0.493521791+1.946037129 i$, | $-0.492000028+1.946432145 i$, |
| 0.1 | $-0.493521791-1.946037129 i$ | $-0.492000028-1.946432145 i$ |
|  | $-0.505847953+1.950474754 i$, | $-0.492189267+1.954872572 i$, |
| 0.2 | $-0.505847953-1.950474754 i$ | $-0.492189267-1.954872572 i$ |
|  | $-0.627817319+1.907010132 i$, | $-0.510399647+1.959417228 i$, |



Figure 9: Sampling zero of sampled-data model with relative degree two and $\beta=-1 / 2$.


Figure 10: The electric circuit plant.
the limiting zeros with AFROH are also stable in some cases due to the same advantages of the FROH and AFROH cases.

## 5. Conclusions

This paper has analyzed the improved asymptotic behavior of limiting zeros for the discrete-time system by using Taylor expansion and the FROH signal reconstruction device. When the normal form representation of continuous-time system with relative degree one or two is discretized, we have given an approximate expression of limiting zeros as power series


Figure 11: The magnitudes of zeros of the sampled-data models with FROH for $T=0.001 \mathrm{~s}$.
expansions with respect to a sampling period up to the thirdorder term. Furthermore, the stability of the sampling zeros is also discussed as the sampling period tends to zero. Finally, it has been shown that FROH provides advantage over ZOH with stability of the limiting zeros of sampled-data systems. The idea of this paper is a further extension of the previous results. For a future study, an extension of the approach to multivariable systems is left.

## Appendix

Calculation of (54) and (55). Denote

$$
\begin{gather*}
A_{1}=\left[\begin{array}{lll}
m_{11} & m_{12}+m_{14} & \mathbf{m}_{13}^{T} \\
m_{21} & m_{22}+m_{24} & \mathbf{m}_{23}^{T} \\
\mathbf{m}_{41} & \mathbf{m}_{42}+\mathbf{m}_{44} & M_{43}
\end{array}\right] \\
A_{2}=\left[\begin{array}{lll}
m_{11} & -m_{12} & \mathbf{m}_{13}^{T} \\
m_{21} & -m_{22} & \mathbf{m}_{23}^{T} \\
\mathbf{m}_{41} & -\mathbf{m}_{42} & M_{43}
\end{array}\right] \tag{A.1}
\end{gather*}
$$

It is immediate to obtain the value of $\left|A_{1}\right|$ from [21].
Here, consider a matrix $M_{\alpha}$ which is defined by neglecting the higher order terms $O(\cdot)$ with respect to $T$ in the
matrix $A_{2}$ since the interests lie in the case of small sampling periods $T$.

Multiplying $M_{\alpha}$ by

$$
L_{1}=\left[\begin{array}{cccc}
\frac{1}{T} & \ell_{1} & \mathbf{0}_{n-2}^{T} & \ell_{2}  \tag{A.2}\\
\mathbf{0}_{n}^{T} & I_{n} & ]^{T}, ., \text {, }, ~
\end{array}\right.
$$

where

$$
\begin{align*}
& \ell_{1}=-\frac{1}{m_{0}}\left(\frac{\beta}{2}-\frac{d_{1} \beta}{6} T+\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{24} T^{2}\right),  \tag{A.3}\\
& \ell_{2}=-\frac{\beta T^{2}}{24 m_{0}}, \quad m_{0}=\frac{\beta}{6}-\frac{d_{1} \beta}{24} T+\frac{\left(d_{1}^{2}-d_{0}\right) \beta}{120} T^{2}
\end{align*}
$$

from the left-hand side leads to

$$
L_{1} M_{\alpha}=\left[\begin{array}{ccc}
\bar{m}_{11} & -\bar{m}_{12} & \overline{\mathbf{m}}_{13}^{T}  \tag{A.4}\\
\bar{m}_{21} & 0 & \overline{\mathbf{m}}_{23}^{T} \\
\overline{\mathbf{m}}_{41} & \mathbf{0}_{n-1} & \bar{M}_{43}
\end{array}\right],
$$

where

$$
\begin{align*}
\bar{m}_{21}= & -z-2+\frac{3 d_{1}}{4} T+\frac{8 d_{0}-13 d_{1}^{2}}{80} T^{2} \\
& +\frac{28 d_{0} d_{1}-3 d_{1}^{3}-40 \mathbf{c}_{n-3}}{960} T^{3}, \\
\overline{\mathbf{m}}_{23}^{T}= & \frac{\mathbf{c}^{T}}{2} T-\frac{d_{1} \mathbf{c}^{T}}{8} T^{2}+\frac{9 d_{1}^{2} \mathbf{c}^{T}-4 d_{0} \mathbf{c}^{T}-20 \mathbf{c}^{T} P^{2}}{480} T^{3},  \tag{A.5}\\
\overline{\mathbf{m}}_{41}= & \frac{\mathbf{q}}{4} T^{2}+\frac{8 P \mathbf{q}-5 d_{1} \mathbf{q}}{48} T^{3}, \\
\bar{M}_{43}= & (1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3} .
\end{align*}
$$

Noting here that

$$
\begin{align*}
\delta= & \bar{m}_{21} \times \bar{M}_{43}-\overline{\mathbf{m}}_{23}^{T} \times \overline{\mathbf{m}}_{41} \\
= & {\left[-z-2+\frac{3 d_{1}}{4} T+\frac{8 d_{0}-13 d_{1}^{2}}{80} T^{2}\right.} \\
& \left.+\frac{28 d_{0} d_{1}-3 d_{1}^{3}-40 \mathbf{c}_{n-3}}{960} T^{3}\right] \\
& \times\left[(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3}\right]+\frac{\mathbf{q c} c^{T}(z-1)}{24} T^{3} \\
= & {\left[-z-2+\frac{3 d_{1}}{4} T+\left(-\frac{13 d_{1}^{2}}{80}+\frac{d_{0}}{10}\right) T^{2}\right.} \\
& \left.+\left(\frac{7 d_{0} d_{1}}{240}-\frac{3 d_{1}^{3}}{960}-\frac{\mathbf{c}_{n-3}}{12}\right) T^{3}\right] \\
& \times\left[(1-z) I+P T+\frac{P^{2}}{2} T^{2}+\frac{P^{3}}{6} T^{3}\right] . \tag{A.6}
\end{align*}
$$

Then,

$$
\begin{equation*}
\left|A_{2}\right|=-\bar{m}_{12} \delta . \tag{A.7}
\end{equation*}
$$

As a result, the calculation is completed.

## Acknowledgments

This work is partially supported by the National Basic Research Program of China ("973" Grant no. 2013CB328903), the National Natural Science Foundation of China (no. 60574003), the Natural Science Foundation Project of CQ CSTC (no. cstc2012jjA40026), and Project no. CDJXS12170006 supported by the Fundamental Research Funds for the Central Universities. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation of the paper.

## References

[1] K. J. Åström, P. Hagander, and J. Sternby, "Zeros of sampled systems," Automatica, vol. 20, no. 1, pp. 31-38, 1984.
[2] T. Hagiwara and M. Araki, "Properties of limiting zeros of sampled systems," Transactions of the Institute of Electrical Engineers of Japan, vol. 110-C, no. 4, pp. 235-244, 1990 (Japanese).
[3] T. Hagiwara, "Analytic study on the intrinsic zeros of sampleddata systems," IEEE Transactions on Automatic Control, vol. 41, no. 2, pp. 261-263, 1996.
[4] T. Hagiwara, T. Yuasa, and M. Araki, "Stability of the limiting zeros of sampled-data systems with zero- and first-order holds," International Journal of Control, vol. 58, no. 6, pp. 1325-1346, 1993.
[5] S. R. Weller, W. Moran, B. Ninness, and A. D. Pollington, "Sampling zeros and the Euler-Frobenius polynomials," IEEE Transactions on Automatic Control, vol. 46, no. 2, pp. 340-343, 2001.
[6] M. J. Błachuta, "On zeros of pulse transfer functions," IEEE Transactions on Automatic Control, vol. 44, no. 6, pp. C1229C1234, 1999.
[7] Y. Hayakawa, S. Hosoe, and M. Ito, "On the limiting zeros of sampled multivariable systems," Systems \& Control Letters, vol. 2, no. 5, pp. 292-300, 1983.
[8] S. R. Weller, "Limiting zeros of decouplable MIMO systems," IEEE Transactions on Automatic Control, vol. 44, no. 1, pp. 129134, 1999.
[9] M. Ishitobi, "A stability condition of zeros of sampled multivariable systems," IEEE Transactions on Automatic Control, vol. 45, no. 2, pp. 295-299, 2000.
[10] S. Liang, M. Ishitobi, W. R. Shi, and X. D. Xian, "On stability of the limiting zeros of discrete-time MIMO systems," Acta Automatica Sinica, vol. 33, no. 4, pp. 439-441, 2007 (Chinese).
[11] K. M. Passino and P. J. Antsaklis, "Inverse stable stampled lowpass systems," International Journal of Control, vol. 47, no. 6, pp. 1905-1913, 1988.
[12] M. Ishitobi, "Stability of zeros of sampled system with fractional order hold," IEE Proceedings, vol. 143, no. 3, pp. 296-300, 1996.
[13] R . Bàrcena, M . de la Sen, and I . Sagastabeitia, "Improving the stability properties of the zeros of sampled systems with fractional order hold," IEE Proceedings, vol. 147, no. 4, pp. 456-464, 2000.
[14] R. Bárcena, M. de la Sen, I. Sagastabeitia, and J. M. Collantes, "Discrete control for a computer hard disk by using a fractional order hold device," IEE Proceedings, vol. 148, no. 2, pp. 117-124, 2001.
[15] S. Liang, M. Ishitobi, and Q. Zhu, "Improvement of stability of zeros in discrete-time multivariable systems using fractionalorder hold," International Journal of Control, vol. 76, no. 17, pp. 1699-1711, 2003.
[16] S. Liang and M. Ishitobi, "The stability properties of the zeros of sampled models for time delay systems in fractional order hold case," Dynamics of Continuous, Discrete \& Impulsive Systems B, vol. 11, no. 3, pp. 299-312, 2004.
[17] M. J. Blachuta, "On fast sampling zeros of systems with frac-tional-order hold," in Proceedings of the American Control Conference, vol. 4, pp. 3229-3230, Arlington, Va, USA, June 2001.
[18] M. Ishitobi, "Properties of zeros of a discrete-time system with fractional order hold," in Proceedings of the 35th IEEE Conference on Decision and Control (CDC '96), pp. 4339-4344, Kobe, Japan, December 1996.
[19] A. Isidori, Nonlinear Control Systems: An Introduction, Springer, Berlin, Germany, 1995.
[20] H. Khalil, Nonlinear Systems, Prentice-Hall, Upper Saddle River, NJ, USA, 2002.
[21] M. Ishitobi, T. Koga, M. Nishi, and S. Kunimatsu, "Asymptotic properties of zeros of sampled-data systems," in Proceedings of the 49th IEEE Conference on Decision and Control (CDC '10), pp. 4952-4957, December 2010.
[22] M. Ishitobi and Q. Zhu, "Zeros of sampled systems with fractional order hold implemented by zero order hold," in Proceedings of the 1997 IEEE International Conference on Intelligent Processing Systems (ICIPS '97), pp. 698-702, Beijing, China, October 1997.
[23] T. Hagiwara, T. Yuasa, and M. Araki, "Limiting properties of the zeros of sampled-data systems with zero- and firstorder holds," in Proceedings of the 31th Conference on Decision and Control (CDC '92), pp. 1949-1954, Tuscon, Ariz, USA, 1992.
[24] S. Liang, M. Ishitobi, T. Koga, and S. Kunimatsu, "On zeros of sampled-data system with relative degree two," in Proceedings of the International Conference on Advanced Mechatronic Systems (ICAMechS '11), pp. 67-71, Zhengzhou, China, August 2011.
[25] M. de la Sen, R. Bárcena, and A. J. Garrido, "On the intrinsic limiting zeros as the sampling period tends to zero," IEEE Transactions on Circuits and Systems I, vol. 48, no. 7, pp. 898900, 2001.

# Modeling and Parameter Analysis of the OC3-Hywind Floating Wind Turbine with a Tuned Mass Damper in Nacelle 

Yulin Si, ${ }^{1}$ Hamid Reza Karimi, ${ }^{1}$ and Huijun Gao ${ }^{2}$<br>${ }^{1}$ Department of Engineering, Faculty of Engineering and Science, University of Agder, N4898 Grimstad, Norway<br>${ }^{2}$ Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Yulin Si; yulin.si@uia.no
Received 12 August 2013; Accepted 10 October 2013
Academic Editor: Dewei Li
Copyright © 2013 Yulin Si et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Floating wind turbine will suffer from more fatigue and ultimate loads compared with fixed-bottom installation due to its floating foundation, while structural control offers a possible solution for direct load reduction. This paper deals with the modelling and parameter tuning of a spar-type floating wind turbine with a tuned mass damper (TMD) installed in nacelle. First of all, a mathematical model for the platform surge-heave-pitch motion and TMD-nacelle interaction is established based on D'Alembert's principle. Both intrinsic dynamics and external hydro and mooring effects are captured in the model, while tower flexibility is also featured. Then, different parameter tuning methods are adopted to determine the TMD parameters for effective load reduction. Finally, fully coupled nonlinear wind turbine simulations with different designs are conducted in different wind and wave conditions. The results demonstrate that the design of TMD with small spring and damping coefficients will achieve much load reduction in the above rated condition. However, it will deteriorate system performance when the turbine is working in the below rated or parked situations. In contrast, the design with large spring and damping constants will produce moderate load reduction in all working conditions.


## 1. Introduction

With less space constraints and more consistent wind, offshore deep sea wind energy has attracted great worldwide attention in recent years. Wind turbines in deep water are usually installed at places where sea depth is between 60 m and 900 m ; thus, floating foundations are generally considered to be an economical and feasible way of deployment [1]. Based on decades of experience from offshore oil and gas industry, several different traditional floating platforms have been proposed to support large wind turbines in deep sea regions, including spar-buoy, tension leg, barge, and semisubmersible [2]. One of the most promising concepts is the spar-type supporting structure, based on which one Norwegian company Statoil has developed the world first experimental large floating offshore wind turbine in 2009.

Different from fixed-bottom wind turbines, the very first challenge for floating windmills is the wave and wind induced platform tilt motion, which will heavily increase the loads on turbine structure due to high inertial and gravitational
forces [3]. According to [4], when comparing a barge-type floating wind turbine with an onshore design, the sea-toland ratio of fatigue loads with respect to tower base bending moments has reached 7. The ratio is still over 1.5 for the OC3-Hywind spar, which may require extra reinforcement or advanced control technique to improve wind turbine reliability. Besides, soft foundation properties of floating wind turbines will lead to low natural frequency platform motion, so that commonly used blade pitch control strategy for fixedbottom wind turbines may cause negative damping of tower bending and even large platform resonant motion [5]. These problems have drawn a lot of attention from researchers on improving the system design and control strategy of floating wind turbines for load reduction.

One approach for vibration inhibition is to utilize structural vibration control devices. This method has been successfully applied in civil engineering structures [6], such as buildings and bridges, and thus is also expected to be a promising solution for extending the fatigue life of floating wind turbines. In [7], Murtagh et al. investigated the use of
a tuned mass damper (TMD) placed at the tower top for the vibration mitigation due to the along-wind forced vibration response of a simplified wind turbine. Following the same installation idea, Colwell and Basu explored the structural responses of fixed-bottom offshore wind turbines with tuned liquid column dampers (TLCD) to control the vibrations [8]. Moreover, Li et al. performed an experimental study on an offshore wind turbines with a ball vibration absorber fixed on top of the nacelle [9]. However, these discussions are about vibration mitigation of fixed-bottom wind turbines, while their dynamics are quite different from that of floating ones. Besides, these works are not based on the cutting edge highfidelity codes for wind turbine simulations, which cannot capture the comprehensive coupled nonlinear dynamics of wind turbines.

FAST (fatigue, aerodynamics, structures, and turbulence) is one of the state-of-the-art aero-hydro-servo-elastic wind turbine numerical simulators [10]. Based on FAST, Lackner and Rotea implemented a new simulation tool, called FASTSC, for passive, semiactive, and active structural control design of wind turbines [11]. Utilizing this code, Lackner and Rotea presented more realistic simulation results with a TMD installed in the nacelle of either a barge-type or a monopile supported wind turbine, and a simple parametric study was also performed to determine the optimal TMD parameters [11]. Further, it was shown that more load reduction could be achieved when introducing active structural control in their following works [12, 13]. In order to perform a more comprehensive parametric study of passive structural control design, the authors in $[14,15]$ established a 3-DOF dynamic model for different types of floating wind turbines based on first principles. This limited DOF model has greatly facilitated the parameter analysis and active control design, while the coupling between surge and pitch motion, however, was not captured, which can be ignored for the barge design but might be an important mode for other platforms, such as spar [16, 17].

Motivated by the above-mentioned problems and research potentials, this work focuses on modeling and parameter analysis of a passive structural control design for a spar-type floating wind turbine. The remainder of this paper is organized as follows. Section 2 introduces the OC3-Hywind floating wind turbine, and the coupled surge-heave-pitch dynamic model with a TMD installed in nacelle is established. Parameter estimation is also performed for model validation. In Section 3, different parameter tuning methods and performance indices are used for TMD parameter determination. Section 4 presents the nonlinear simulation results under different wind and wave conditions. Advantages and limitations of this design with different TMD parameters are also analyzed. At last, we draw conclusions in Section 5.

## 2. Dynamic Modelling

In cooperation with Statoil, Jonkman from NREL has specified a detailed OC3-Hywind spar-type floating wind turbine model, which is a combination of the data for the 5MW

Table 1: Properties of the OC3-Hywind model [16, 18].

| Item | Value |
| :--- | :--- |
| Rating | 5 MW |
| Rotor configuration | Upwind, 3 blades |
| Cut-in, rated, cut-out wind speed | $3 \mathrm{~m} / \mathrm{s}, 11.4 \mathrm{~m} / \mathrm{s}$, |
|  | $25 \mathrm{~m} / \mathrm{s}$ |
| Total draft below sea water level (SWL) | 120 m |
| Tower base above SWL | 10 m |
| Hub height above SWL | 90 m |
| Nacelle dimension (length, width, height) | $14.2 \mathrm{~m}, 2.3 \mathrm{~m}, 3.5 \mathrm{~m}$ |
| Platform diameter above taper | 6.5 m |
| Platform diameter below taper | 9.4 m |
| Rotor nacelle assembly (RNA) mass | $350,000 \mathrm{~kg}$ |
| Tower mass | $249,718 \mathrm{~kg}$ |
| Platform mass | $7,466,000 \mathrm{~kg}$ |
| Number of mooring lines | 3 |
| Depth to fairleads below SWL | 70 m |
| Baseline control in Region 3 | GSPI and constant |

baseline wind turbine from NREL and the Hywind floating platform from Statoil [16, 18]. Properties of the OC3-Hywind model are shown in Table 1. According to [16], in order to avoid resonant platform pitch motion, the conventional controller in Region 3 is modified into a combination of gain reduced gain-scheduled proportional-integral (GSPI) collective blade pitch control and constant torque control, which is used all through this work as the baseline.

The passive structural control strategy in this work is to install one TMD in the nacelle, which is assumed to move on an ideal nonfriction linear track along the fore-aft direction. The stiffness and damping parameters of TMD can be tuned, and they are regarded as constant in all simulations. In order to investigate these parameters, optimize system performance, or design an active controller, establishing one dynamic mathematical model is usually helpful. Figure 1 shows a diagram of the OC3-Hywind surge-heave-pitch motion with tower fore-aft bending and the TMD-nacelle interaction. Definition of each term in this figure can be found in Table 2. Before presenting the dynamic model, the following premises and assumptions need to be listed.
(1) OC3-Hywind is treated as a multibody dynamic system, and the motion of reference point $P$ is chosen for output analysis, which is in accordance with the definition in [16]. Rigid bodies in the model include the spar platform, tower, and rotor nacelle assembly (RNA). Dynamics in rotor, generator, and gearbox are not considered in this work.
(2) Based on the same assumption, the tower fore-aft flexibility is represented as that in [13], where the tower, for simplicity, is treated as a linear rigid rotating beam hinged at tower bottom. It is also assumed that the spring and damping coefficients of this hinge are constant.

Table 2: Term descriptions in the model of OC3-Hywind surge-heave-pitch motion.

| Terms | Descriptions |
| :--- | :--- |
| sg | DOF of platform surge motion |
| hv | DOF of platform have motion |
| $p$ | DOF of platform pitch motion |
| tmd | DOF of TMD motion |
| $t$ | DOF of tower fore-aft bending |
| $\theta_{i}$ | Rotation angle of DOF $i$ |
| $x_{i}$ | Displacement of DOF $i$ |
| $M_{i}^{j}$ | Generalized mass for DOF $i$ with regard to DOF $j$ |
| $I_{i}^{j}$ | Generalized inertia tensor for DOF $i$ with regard to |
| $F_{i}^{j}$ | DOF $j$ |
| $\tau_{i}^{j}$ | Generalized force for DOF $i$ due to effect or DOF $j$ |
| gr | Generalized torque for DOF $i$ due to effect or DOF $j$ |
| hdr | Gravitational effect |
| ctr | Hydro effect |
| moor | Centripetal effect |
| spr.damp | Mooring lines effect |
| $A_{i}^{j}$ | Spring and damping effect of TMD |
| $J_{u}^{X}$ | Generalized added mass for DOF $i$ with regard to |
| $L_{u}$ | Inertia tensor for $u$ with regard to point $X$ |
| $m_{u}$ | Length of part $u$ |
| ptfm | Mass of part $u$ |
| twr | Platform |
| rna | Tower |
| $d$ | Rotor nacelle assembly (RNA) |
| jot | Misalignment between RNA mass center and tower |
| $D_{i}^{j}$ | centerline |
| $K_{i}^{j}$ | Joint between platform and tower |
| $g$ | Equivalent damping coefficient for DOF $i$ with |
| regard to DOF $j$ |  |
| CG | Equivalent spring coefficient for DOF $i$ with regard |
|  | to DOF $j$ |
| Gravitational acceleration |  |
|  | Center of buoyancy |
| Gravity center of part $u$ |  |

(3) In total, the model has five DOFs, that is, platform surge, heave, pitch, tower fore-aft bending, and TMD motion. The other DOFs, such as rotor yaw motion and generator rotation, are not included.
(4) This model focuses on the system intrinsic coupled dynamics with hydro and mooring loads, while the loads from winds and incident waves have not yet been considered in the modelling process.

Based on the above descriptions, we treat the overall system dynamics as the motion of a rigid body with distributed mass particles in the surge-heave-pitch plane, which can be seen as the sum of a translation and a rotation about the axis passing through $P$ and perpendicular to this plane [19].


Figure 1: Diagram of the OC3-Hywind surge-pitch-heave motion with tower fore-aft flexibility and passive structural control.

According to D'Alembert's principle of inertial forces, the following static equilibrium equations for system translation and rotation about the reference point $P$

$$
\begin{gather*}
\mathbf{F}-\sum m_{i} \mathbf{a}_{i}=0 \\
\boldsymbol{\tau}-\sum \mathbf{r}_{i} \times m_{i} \mathbf{a}_{i}=0, \tag{1}
\end{gather*}
$$

hold. $\mathbf{F}$ and $\boldsymbol{\tau}$ denote vectors of external forces and moments about $P$, while $-\sum m_{i} \mathbf{a}_{i}$ and $-\sum \mathbf{r}_{i} \times m_{i} \mathbf{a}_{i}$ are vector sums of inertial forces and torques about $P . m_{i}$ is the mass of particle $i$, that is, platform, tower, RNA, and TMD, and $\mathbf{r}_{i}$ represents the position vector from $P$ to particle $i . \mathbf{a}_{i}$ is the acceleration vector for mass particle $i$, and it consists of the translational acceleration, normal, and tangential rotational acceleration components.

When considering the tower translation and rotation about tower bottom, the motion of tower fore-aft bending can be described as

$$
\begin{equation*}
\sum\left(\mathbf{r}_{i} \times m_{i} \mathbf{a}_{i}\right)+I_{t}^{t} \boldsymbol{\alpha}_{t}=\boldsymbol{\tau}_{t}^{g r}+\boldsymbol{\tau}_{t}^{p}, \tag{2}
\end{equation*}
$$

which is also based on D'Alembert's principle. $m_{i}$ denotes the mass of tower, RNA, and TMD. $I_{t}^{t}$ is the equivalent moment of inertia for tower and RNA about tower bottom, and $\boldsymbol{\alpha}_{t}$ denotes the angular acceleration vector of tower pitch motion. $\boldsymbol{\tau}_{t}^{p}$ is the torque vector due to the spring-damping


Figure 2: Diagram for calibration of nacelle rotation angle.
effect between tower and platform. To be consistent with the output of FAST simulator, the tower top displacement is also calculated, which is given by

$$
\begin{equation*}
x_{t}=\sin \left(\theta_{t}-\theta_{p}\right) l_{\mathrm{twr}} \tag{3}
\end{equation*}
$$

where $l_{\text {twr }}$ is the length of flexible tower. However, in the system validation process, one problem is found which is that there will exist huge misalignment between the responses of FAST-SC and established model when the spring and damping coefficients of TMD are in small scale. This is mainly due to the inaccuracy of nacelle rotation angle when flexible tower is modeled as a rigid rotating beam. When TMD has tiny spring and damping constants, its acceleration will be mainly contributed by gravity, so that inaccuracy of $\theta_{t}$ will lead to tremendous difference of TMD dynamics. Therefore, the nacelle rotation angle should be calibrated in order to produce more convincing dynamic responses. In FAST, the tower flexibility is depicted by the predefined mode shapes $\Phi$, where tower top rotation angle is determined by the product of tower top mode shape slope $\partial \Phi(h)_{\text {rna }} / \partial h$ and tower top displacement $x_{t}$. Following similar calculation procedure, the diagram for tower top rotation calibration is illustrated in Figure 2, and the calibrated nacelle rotation angle $\widetilde{\theta}_{t}$ satisfies

$$
\begin{equation*}
\widetilde{\theta}_{t}=\left.\frac{\partial \Phi(h)}{\partial h}\right|_{h=L_{\mathrm{rna}}} x_{t}+\theta_{p} \tag{4}
\end{equation*}
$$

Next, the hydrodynamic loads are characterized. When formulating the motion of object submerged in water, we must also consider the added-mass effect, resulting from its surrounding fluid [20]. It is summarized in [1] that the hydrodynamic loads mainly include contributions from hydrostatics (from water-plane area and buoyancy), radiation (from outgoing waves generated by platform motion), and diffraction (from incident waves). In accordance with this analysis, the hydrodynamic load calculation in this work follows a similar path. Firstly, hydrostatic load in this model consists of buoyancy force and restoring load resulting
from the effects of water-plane area and buoyancy, and the restoring force and moment are set to be constantly proportional to platform displacement and tilt angle which have been specified in [16]. Secondly, the radiation loads can be represented by nonlinear vicious drag, hydrodynamic radiation damping, and the above mentioned added-mass effects. Thirdly, incident wave loads are not considered here since wind turbine is supposed to be located in still water in design process.

Regarding the mooring system, FAST simulator uses a quasistatic model to calculate the load of an individual mooring line, which exhibits nonlinear behaviors due to both mooring dynamics and the asymmetry of the three-point mooring system. In the simulations of this work, the platform displacement and tilt angle are usually not in big scale where the mooring system load-displacement relationship does not show strong nonlinearities in surge and pitch modes, so we still choose the simple linear model to represent this effect.

In sum, except for added mass, the hydrodynamic loads and mooring effect are modeled as

$$
\begin{gather*}
F_{\mathrm{sg}}^{\mathrm{hdr} \cdot \mathrm{moor}}=-D_{\mathrm{sg}}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}-\widehat{D}_{\mathrm{sg}}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}^{2}-K_{\mathrm{sg}}^{\mathrm{sg}} x_{\mathrm{sg}}-D_{\mathrm{sg}}^{p} \dot{\theta}_{p}-K_{\mathrm{sg}}^{p} \theta_{p}, \\
F_{\mathrm{hv}}^{\mathrm{hdr} \cdot \mathrm{moorr}}=-D_{\mathrm{hv}}^{\mathrm{hv}} \dot{x}_{\mathrm{hv}}-K_{\mathrm{hv}}^{\mathrm{hv}} x_{\mathrm{hv}}-F_{\mathrm{moor}}^{0}+F_{\mathrm{buoy}}^{0} \\
-K_{\mathrm{hv}}^{p \cdot \mathrm{sg}}\left(x_{\mathrm{sg}}-L_{\mathrm{moor}} \sin \theta_{p}\right)^{2}, \\
\tau_{p}^{\mathrm{hdr} \cdot \mathrm{moor}}=-D_{p}^{p} \dot{\theta}_{p}-K_{p}^{p} \theta_{p}-D_{p}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}-\widehat{D}_{p}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}^{2}-K_{p}^{\mathrm{sg}} x_{\mathrm{sg}} . \tag{5}
\end{gather*}
$$

$D_{i}^{j}, \widehat{D}_{i}^{j}$, and $K_{i}^{j}$ denote equivalent damping and spring coefficients for DOF $i$ with regard to DOF $j$ for the calculation of hydro and mooring effects. $F_{\text {moor }}^{0}$ and $F_{\text {buoy }}^{0}$ represent initial mooring line and buoyancy forces when there isno platform displacement or rotation. It should be noted that the mooring load for platform heave motion shows strong nonlinear relationship with the surge and pitch modes; thus, it is not simplified.

Based on the above analysis and equations, the nonlinear dynamic model of OC3-Hywind surge-heave-pitch motion can be established in the following implicit form:

$$
\begin{align*}
& {\left[\begin{array}{ccccc}
M_{\mathrm{sg}}^{\mathrm{sg}} & 0 & I_{\mathrm{sg}}^{p} & M_{\mathrm{sg}}^{\mathrm{tmd}} & I_{\mathrm{sg}}^{t} \\
0 & M_{\mathrm{hv}}^{\mathrm{hv}} & I_{\mathrm{hv}}^{p} & M_{\mathrm{hbv}}^{\mathrm{md}} & I_{\mathrm{hv}}^{t} \\
M_{p}^{\mathrm{sg}} & M_{p p}^{\mathrm{hV}} & I_{p}^{p} & M_{p}^{\mathrm{ttd}} & 0 \\
M_{\mathrm{tmd}}^{\mathrm{sg}} & M_{\mathrm{tmd}}^{\mathrm{hv}} & I_{\mathrm{tmd}}^{p} & M_{\mathrm{tmd}}^{\mathrm{tmd}} & I_{\mathrm{tmd}}^{t} \\
M_{t}^{\mathrm{sg}} & M_{t}^{\mathrm{hvv}} & 0 & M_{t}^{\mathrm{tmd}} & I_{t}^{t}
\end{array}\right]\left[\begin{array}{c}
\ddot{x}_{\mathrm{sg}} \\
\ddot{x}_{\mathrm{hv}} \\
\ddot{\theta}_{p} \\
\ddot{x}_{\mathrm{tmd}} \\
\ddot{\theta}_{t}
\end{array}\right]} \\
& \quad=\left[\begin{array}{c}
F_{\mathrm{sg}}^{\mathrm{hdr} \cdot \mathrm{moor}}+F_{\mathrm{sg}}^{\mathrm{ctr}} \\
F_{\mathrm{hv}}^{\mathrm{gr}}+F_{\mathrm{hv}}^{\mathrm{hdr} \cdot \mathrm{moor}}+F_{\mathrm{hv}}^{\mathrm{ctr}} \\
\tau_{p}^{\mathrm{gr}}+\tau_{p}^{\mathrm{hdr} \cdot m o o r}+\tau_{p}^{\mathrm{ctr}} \\
F_{\mathrm{tmd}}^{\mathrm{gr}}+F_{\mathrm{tmd}}^{\mathrm{spr} \cdot d a m p} \\
\tau_{t}^{\mathrm{gr}}+\tau_{t}^{p}+\tau_{t}^{\mathrm{ctr}}
\end{array}\right] . \tag{6}
\end{align*}
$$

In this model, $\mathrm{sg}, \mathrm{hv}, p, \mathrm{tmd}$, and $t$ represent, respectively, the enabled 5 DOFs , that is, platform surge, heave, pitch motion about $P$, TMD translation, and tower rotation. On
the left side, $M_{i}^{j}$ and $I_{i}^{j}$ denote generalized mass and generalized inertial tensor for DOF $i$ with regard to DOF $j$. On the right side, gr, hdr, moor, ctr, spr, and damp describe gravitational, hydro, centripetal, spring, and damping effects in forces and moments. Expanded expressions of this model for TMD platform installation are presented in the appendix, and the detailed term descriptions are listed in Table 2.

The mass matrix on the left side of (6) exhibits the system inertial property, that is, mass and inertia tensor, and it also includes hydro added mass and acceleration coupling terms. The terms on the right side of (6) are external loads, which can be classified into several different effects. Gravitational forces and moments are the first type of loads, labeled as gr. The second effect, labeled as hdr-moor, is the hydrodynamic and mooring loading, which consists of hydrostatics, vicious drag, radiation damping, additional linear damping, and mooring effects. The third type, which is produced by D'Alembert's principle, is the centripetal forces and moments which originate from the rotation of platform, tower, and TMD about the reference point $P$, and they are labeled as ctr. Tower and platform interaction is the fourth effect captured in this equation, and the bending moment is described by a linear spring-damper between them. The final consideration is the spring and damping effect in TMD, so it is labeled as spr•damp.

After obtaining the OC3-Hywind dynamic model for its surge-heave-pitch motion in still water, parameter identification and validation should be performed to quantize the unknown parameters and verify the correctness of the proposed model. The parameter estimation is accomplished by minimizing the output difference between FAST-SC and the established model. Based on the estimation result, free decay response comparison for the OC3-Hywind surge-pitch-heave motion without TMD is illustrated in Figure 3, where two results coincide well with each other. Then, in order to further validate the established model, free decay response comparisons are performed again with TMD installed in nacelle. In practice, there exist space limitations for the nacelle, so the TMD displacement should be restricted into a certain range. According to the nacelle dimensions defined in [21], the TMD displacement range is determined as $\pm 7 \mathrm{~m}$ in this work. In FAST-SC, the TMD motion constraints were modelled as stops, where there would be spring stiffness and damping interaction between TMD and nacelle or platform when its displacement exceeds the user defined constraints. The stops effect in this work is characterized in the same way. Figure 4 illustrates the free decay response comparison results with TMD stops. As expected, the established model still manages to capture the system dynamics including TMD stop interactions. It is worth mentioning that the stops with various spring and damping coefficients could have quite different impacts on system dynamics, but further analysis of stop parameters is not within the scope of this paper.

Based on the above analysis, the proposed model has captured most of the intrinsic dynamics for OC3-Hywind surge-heave-pitch motion, including hydrodynamic and mooring loads, tower flexibility, and TMD-nacelle interaction. Next step is to tune TMD parameters for effective system load reduction.


Figure 3: Free decay response comparison between identified model and FAST-SC numerical simulation for surge-pitch-heave motion without TMD ( $5^{\circ}$ initial platform pitch).


Figure 4: Free decay response comparison with TMD and stops in nacelle.

## 3. Parameter Tuning

Optimal parameter tuning of the vibration absorber is an important design consideration in passive structural control problems. The design aim in this work is to find the optimal TMD coefficients for wind turbine load reduction. The parameters to be determined include TMD spring and damping coefficients. TMD mass is not parametrically studied in this work since it is usually determined by cost and heavier mass will more likely produce better performance. Specifically, in order to be consistent with [11], the mass is chosen to be $20,000 \mathrm{~kg}$, which takes about $3.33 \%$ of the weight for tower-RNA structure.

In fact, the most convincing solution here is to try all possible values of these parameters in FAST-SC. However, this global searching process will take tens of thousands of calls from FAST-SC, and it usually will take minutes to run it for only one time. Therefore, exhaustive search is almost impossible with ordinary computers, and appropriate optimization methods are needed. Based on the established model, in this section, three different methods are used for this parameter tuning problem.
3.1. Frequency and Damping Analysis. In engineering applications, the natural frequency of TMD is usually tuned to be near that of the target system; thus, it will effectively dissipate the undesirable system vibration energy. In order to systematically describe this phenomenon, Den Hartog [22] analyzed the response of undamped main system with TMD subjected to harmonic external forces and derived an explicit expression to determine the optimal TMD natural frequency and damping ratio for vibration inhibition. The optimal solution is given by

$$
\begin{equation*}
f_{\mathrm{tmd}}=\frac{f}{1+\mu}, \quad \xi_{\mathrm{tmd}}=\sqrt{\frac{3 \mu}{8(1+\mu)}}, \tag{7}
\end{equation*}
$$

where $\mu$ denotes the mass ratio $m_{\text {tmd }} / m$ and $f$ and $\xi$ are the natural frequency and damping ratio of target system. $f_{\text {tmd }}$ and $\xi_{\text {tmd }}$ represent the optimal natural frequency and damping ratio of TMD.

In order to adopt this method, eigenanalysis based on model linearization result is performed first to obtain system natural frequencies and damping ratios for the modes of interest.

The eigenanalysis result has been presented in [23], where natural frequencies of two most critical modes, that is, platform pitch mode and first tower fore-aft bending mode, are 0.4732 Hz and 0.0342 Hz , and their damping ratios are 0.0087 and 0.1418 .

However, in this analysis process, the nonlinearity of TMD stops due to space constraints is not considered, which has been shown to have strong influence on TMD load reduction effectiveness according to the following nonlinear FAST-SC simulation results. Therefore, a more thorough method should be proposed to find the best combination of these TMD parameters.

Table 3: Performance indices.

| Index | Description |
| :--- | :--- |
| $J_{1}=\sqrt{(1 / T) \int_{0}^{T}\left(x_{t t}-\tilde{x}_{t t}\right)^{2} d t}$ | Standard deviation of tower top <br> displacement under its <br> equilibrium point |
| $J_{2}=\max \left(x_{t t}\right)-\min \left(x_{t t}\right)$ | Range of tower top <br> displacement |



Figure 5: Surface plot subjected to performance index $J_{1}$ with TMD installed in nacelle.
3.2. Surface Plot. In the previous section, we have obtained a mathematical model describing OC3-Hywind surge-heavepitch motion, which manages to capture most of the system structural dynamics, hydro and mooring effects. More importantly, the time for solving this dynamic equation is less than 1s; thus, surface plotting, a global parameter searching method, becomes a possible solution to determine the optimal TMD parameters.

Next, we introduce the performance indices in Table 3 which are used in the optimization process. The tower top fore-aft deflection is the best indicator of tower bottom bending moments, and the author in [14] used standard deviation of tower top displacement as the performance index, which is also adopted in this work as the first performance index $J_{1}$. Secondly, we also care about load reduction effectiveness of the proposed method in extreme events; thus, the range of tower top displacement in the free decay test is treated as another evaluation index $J_{2}$.

Based on these indices, exhaustive search is performed where TMD spring and damping constants are regarded as two coefficients to be optimized. The parameter range and interval are chosen when both time consumption and accuracy are considered. The surface plots for different design criteria are illustrated in Figures 5 and 6, and the optimization results are listed in Table 4.

Although surface plotting could be regarded as a global optimization method, which produces a relatively comprehensive evaluation of the performance index with possible parameters, it is still computationally expensive, which will take hours or days to finish one optimization process. Also, there might exist better solution if the parameter interval is not small enough. Therefore, more intelligent and efficient optimization algorithms are demanded.


Figure 6: Surface plot subjected to performance index $J_{2}$ with TMD installed in nacelle.

Table 4: Parameter optimization result with TMD in nacelle ( $m_{\text {tmd }}$ $=20,000 \mathrm{~kg}$ ).

| Method | Performance index | $K_{\mathrm{tmd}}$ <br> $(\mathrm{N} / \mathrm{m})$ | $D_{\mathrm{tmd}}$ <br> $(\mathrm{N} \cdot \mathrm{s} / \mathrm{m})$ |
| :--- | :---: | :---: | :---: |
| Den Hartog <br> [22] | Tower bending mode <br> (Denl) | 165571 | 12661 |
| Den Hartog | Platform pitch mode <br> (Den2) | 865 | 915 |
| [22] | $J_{1}=0.0872 \mathrm{~m}$ | 0 | 3200 |
| Surface plot | $J_{2}=0.8389 \mathrm{~m}$ | 1200 | 800 |
| Surface plot | $J_{1}=0.0871 \mathrm{~m}$ | 0 | 3130 |
| GA | $J_{2}=0.7620 \mathrm{~m}$ | 164231 | 20889 |
| GA |  |  |  |

3.3. Genetic Algorithm. In the past few years, genetic algorithm has been widely applied in a broad spectrum of realworld systems [24-26]. This approach starts with randomly generated population, and individuals with better fitness will be selected as the basis of the next generation. The improved population will keep evolving after inheritance, mutation, selection, and crossover procedures until it meets the final requirement. As a global optimization method, genetic algorithm is based on stochastic variables and does not require the derivatives of object function, which brings the advantages of global evaluation and objective tolerance when compared with other gradient based local optimization methods. It usually helps to obtain a better result in optimization problems with nonsmooth objective functions and thus is suitable for the optimization problem in this work.

When implementing the algorithm, probability of the roulette wheel uniform crossover is chosen as 0.6 , and the mutation probability 0.01 is used. Minimum number of generations is set as 20 . Optimization results are shown in Table 4. It can be noticed that genetic algorithm gives a better result with respect to $J_{2}$ since the surface plotting has a limited searching range.

## 4. Simulation and Analysis

In this section, based on the optimization result, fully nonlinear simulations are performed in FAST-SC with all wind turbine DOFs enabled. Each test runs for 630 seconds, and

Table 5: Percentage of load reduction with different TMD tuning results compared with baseline.

| Case | Evaluation index | Den1 | Den2 | $J_{1}$ | $J_{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $10 \mathrm{~m} / \mathrm{s}$ | DEL TwrBsMyt | 6.35 | 0.66 | 0.52 | 6.00 |
|  | DEL TwrBsMxt | 32.18 | 14.2 | 11.44 | 28.37 |
|  | DEL RootMycl | 1.07 | -0.18 | 0.10 | 0.85 |
|  | DEL AnchiTen | 0.93 | 3.01 | 1.21 | 0.93 |
|  | 95th TwrBsMyt | -2.00 | -4.04 | -3.89 | -2.00 |
|  | 95th TwrBsMxt | 6.01 | 2.7 | 2.55 | 5.06 |
|  | 95th PtfmPitch | -2.08 | -1.96 | -2.08 | -2.08 |
|  | 95th PtfmRoll | -1.67 | 0.21 | 0.13 | -1.67 |
| $18 \mathrm{~m} / \mathrm{s}$ | DEL TwrBsMyt | 3.61 | 7.77 | 8.78 | 3.35 |
|  | DEL TwrBsMxt | 25.55 | 0.98 | -3.94 | 21.24 |
|  | DEL RootMycl | 1.07 | 4.99 | 5.93 | 1.14 |
|  | DEL AnchiTen | 1.15 | 0.32 | 0.32 | 1.14 |
|  | 95th TwrBsMyt | -3.15 | 5.02 | 6.48 | -3.25 |
|  | 95th TwrBsMxt | 7.90 | 4.70 | 1.69 | 7.10 |
|  | 95th PtfmPitch | -1.05 | 10.66 | 12.43 | -1.04 |
|  | 95th PtfmRoll | 6.55 | 15.54 | 14.32 | 6.58 |
|  | RMS GenPwr | -5.46 | 21.09 | 29.22 | -5.41 |
| $37 \mathrm{~m} / \mathrm{s}$ | DEL TwrBsMyt | 1.47 | -19.95 | -16.25 | 1.22 |
|  | DEL TwrBsMxt | 0.14 | 0.51 | 0.42 | 0.18 |
|  | DEL RootMycl | 1.80 | -45.71 | -28.34 | 2.03 |
|  | DEL AnchiTen | 1.33 | 1.83 | 0.96 | 0.78 |
|  | 95th TwrBsMyt | -0.78 | -4.88 | -2.33 | -0.77 |
|  | 95th TwrBsMxt | 0.41 | 0.40 | 0.25 | 0.47 |
|  | 95th PtfmPitch | 4.41 | 5.40 | 4.44 | 4.41 |
|  | 95th PtfmRoll | -0.30 | -0.63 | -0.57 | -0.30 |

the output data in the first 30 s are not recorded, waiting for generator torque and blade pitch motion to arrive at normal operation state. The modified generator torque and blade pitch controller from NREL will be used in the form of a dynamic link library for all tests [16].

The wind and wave conditions in the experiment are defined almost the same as in [12]. For wind condition, both the above and below rated wind speeds are considered, and mean value of turbulent wind is defined as $18 \mathrm{~m} / \mathrm{s}$ and $10 \mathrm{~m} / \mathrm{s}$ separately. The turbulent wind file is generated by TurbSim, where Kaimal spectra and the power law exponent of 0.14 are used according to the IEC61400-3 offshore wind turbine design standard. The normal turbulence intensity is set as $15 \%$ ( $18 \mathrm{~m} / \mathrm{s}$ case) and $18 \%$ ( $10 \mathrm{~m} / \mathrm{s}$ case). Random seed in this work is arbitrarily chosen as 231857312. In order to define the wave condition, JONSWAP spectrum is utilized to generate the stochastic wave inputs. The significant wave height is set as $2.3 \mathrm{~m}(10 \mathrm{~m} / \mathrm{s}$ case) and 3.7 m ( $18 \mathrm{~m} / \mathrm{s}$ case), and the peak spectral period is defined as 14 s . Besides, the parked situation is also considered assuming the turbine suffers extreme 50year storm, that is, $37 \mathrm{~m} / \mathrm{s}$ turbulent wind with power law exponent of 0.11 and $11 \%$ turbulence intensity. Wave height and period are defined as 13.8 m and 19 s .

Percentage of load reduction with different TMD parameter choice is shown in Table 5. In order to measure the fatigue and extreme loading, damage equivalent load (DEL) and the


Figure 7: FAST-SC simulation results with $18 \mathrm{~m} / \mathrm{s}$ turbulent wind and 3.7 m significant height wave.

95th percentile of fore-aft and side-side tower base bending moments (TwrBsMyt and TwrBsMxt) and flapwise bending moment at the first blade root (RootMycl) are calculated, together with the 95th percentile of platform pitch and roll rotation angle. In the above rated situation, the root mean square (RMS) of generated power is considered as another index.

It can be seen from results that the design of TMD with small spring and damping coefficients will achieve much load reduction in the above rated condition, where one simulation result is shown in Figure 7. However, it will deteriorate system performance when the turbine is working in the below rated or parked situations. In contrast, the design with large spring
and damping constants will produce moderate load reduction in all working conditions.

## 5. Conclusion

This work focuses on the modeling and parameter tuning of a passive structural control design for the OC3-Hywind floating wind turbine. Firstly, the coupled surge-heave-pitch dynamic model with a TMD installed in nacelle is established based on the D'Alembert's principle. Parameter estimation is also performed for model validation. Then, different parameter tuning methods and performance indices are used for TMD parameter determination. FAST-SC is used for fully coupled nonlinear simulation with various wind and wave conditions. The results show that the design of TMD with small spring and damping coefficients will achieve much load reduction in the above rated condition, but it will deteriorate system performance when the turbine is working in the below rated or parked situations. In contrast, the design with large spring and damping constants will produce moderate load reduction in all working conditions. Therefore, inappropriate TMD design will not contribute to wind turbine load reduction. Besides, only enabling TMD in certain range of wind speed might be a possible solution for this design. Further real experiments need to be conducted to verify this idea. Future work will also consider the situation when TMD is installed in the spar itself or other types of platforms.

## Appendix

Consider the following:

$$
\begin{gathered}
M_{\mathrm{sg}}^{\mathrm{sg}}=A_{\mathrm{sg}}^{\mathrm{sg}}+m_{\mathrm{ptfm}}+m_{\mathrm{twr}}+m_{\mathrm{rna}}+m_{\mathrm{tmd}}, \\
I_{\mathrm{sg}}^{p}=A_{\mathrm{sg}}^{p}+m_{\mathrm{twr}}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \cos \theta_{p} \\
-m_{\mathrm{ptfm}} L_{\mathrm{ptfm}} \cos \theta_{p} \\
M_{\mathrm{sg}}^{\mathrm{tmd}}=M_{\mathrm{tmd}}^{\mathrm{sg}}=m_{\mathrm{tmd}} \cos \left(\theta_{p}+\sin \left(\theta_{t}-\theta_{p}\right) L_{\mathrm{rna}} \dot{\Phi}_{\mathrm{rna}}\right), \\
I_{\mathrm{sg}}^{t}=m_{\mathrm{rna}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \cos \theta_{t}+m_{\mathrm{tmd}} L_{\mathrm{rna}} \cos \theta_{t}, \\
M_{\mathrm{hv}}^{\mathrm{hv}}=A_{\mathrm{hv}}^{\mathrm{hv}}+m_{\mathrm{ptfm}}+m_{\mathrm{twr}}+m_{\mathrm{rna}}+m_{\mathrm{tmd}}, \\
I_{\mathrm{hv}}^{p}=-m_{\mathrm{twr}}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \sin \theta_{p} \\
\quad+m_{\mathrm{ptfm}} L_{\mathrm{ptfm}} \sin \theta_{p}, \\
M_{\mathrm{hv}}^{\mathrm{tmd}}=M_{\mathrm{tmd}}^{\mathrm{hv}}=-m_{\mathrm{tmd}} \sin \left(\theta_{p}+\sin \left(\theta_{t}-\theta_{p}\right) L_{\mathrm{rna}} \dot{\Phi}_{\mathrm{rna}}\right), \\
I_{\mathrm{hv}}^{t}=- \\
m_{\mathrm{rna}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \sin \theta_{t}-m_{\mathrm{tmd}} L_{\mathrm{rna}} \sin \theta_{t}, \\
M_{p}^{\mathrm{sg}}=A_{p}^{\mathrm{sg}}+m_{\mathrm{rna}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \cos \theta_{t} \\
\\
+m_{\mathrm{twr}}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \cos \theta_{p} \\
\quad-m_{\mathrm{ptfm}} L_{\mathrm{ptfm}} \cos \theta_{p}+m_{\mathrm{tmd}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \cos \theta_{t},
\end{gathered}
$$

$$
\begin{aligned}
& M_{p}^{\mathrm{hv}}=-m_{\mathrm{rna}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \sin \theta_{t} \\
& -m_{\mathrm{twr}}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \sin \theta_{p} \\
& +m_{\mathrm{ptfm}} L_{\mathrm{ptfm}} \sin \theta_{p}-m_{\mathrm{tmd}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \sin \theta_{t}, \\
& I_{p}^{p}=A_{p}^{p}+J_{\mathrm{ptfm}}^{\mathrm{CG}_{\mathrm{ptf}}}+m_{\mathrm{ptfm}} L_{\mathrm{ptfm}}^{2}+J_{\mathrm{twr}}^{\mathrm{CG}_{\mathrm{twr}}} \\
& +m_{\mathrm{twr}}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right)^{2}+m_{\mathrm{rna}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right)^{2} \\
& +m_{\mathrm{tmd}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right)^{2} \text {, } \\
& I_{\mathrm{tmd}}^{p}=0, \quad M_{\mathrm{tmd}}^{\mathrm{tmd}}=m_{\mathrm{tmd}}, \\
& I_{\mathrm{tmd}}^{t}=M_{t}^{\mathrm{tmd}}=m_{\mathrm{tmd}} L_{\mathrm{rna}} \cos \left(\sin \left(\theta_{t}-\theta_{p}\right) L_{\mathrm{rna}} \dot{\Phi}_{\mathrm{rna}}\right), \\
& M_{t}^{\mathrm{sg}}=m_{\mathrm{rna}} L_{\mathrm{rna}} \cos \theta_{t}+m_{\mathrm{twr}} L_{\mathrm{twr}} \cos \theta_{p} \\
& +m_{\mathrm{tmd}} L_{\mathrm{rna}} \cos \theta_{t} \text {, } \\
& M_{t}^{\mathrm{hv}}=-m_{\mathrm{rna}} L_{\mathrm{rna}} \sin \theta_{t}-m_{\mathrm{twr}} L_{\mathrm{twr}} \sin \theta_{p} \\
& -m_{\mathrm{tmd}} L_{\mathrm{rna}} \sin \theta_{t} \text {, } \\
& \dot{\Phi}_{\mathrm{rna}}=\left.\frac{\partial \Phi(h)}{\partial h}\right|_{h=L_{\mathrm{rna}}}, \\
& F_{\mathrm{hv}}^{\mathrm{gr}}=-\left(m_{\mathrm{ptfm}}+m_{\mathrm{twr}}+m_{\mathrm{rna}}+m_{\mathrm{tmd}}\right) g, \\
& \tau_{p}^{\mathrm{gr}}=m_{\mathrm{rna}}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \sin \theta_{t} \\
& +m_{\mathrm{twr}} g\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \sin \theta_{p} \\
& -m_{\mathrm{ptfm}} g L_{\mathrm{ptfm}} \sin \theta_{p}-m_{\mathrm{rna}} g L_{d} \cos \theta_{t} \\
& +m_{\mathrm{tmd}} g x_{\mathrm{tmd}} \cos \theta_{t}+m_{\mathrm{tmd}} g\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \sin \theta_{t}, \\
& F_{\mathrm{tmd}}^{\mathrm{gr}}=m_{\mathrm{tmd}} g \sin \left(\theta_{p}+\sin \left(\theta_{t}-\theta_{p}\right) L_{\mathrm{rna}} \dot{\Phi}_{\mathrm{rna}}\right) \text {, } \\
& \tau_{t}^{\mathrm{gr}}=m_{\mathrm{twr}} g L_{\mathrm{twr}} \sin \theta_{p}+m_{\mathrm{rna}} g L_{\mathrm{rna}} \sin \theta_{t} \\
& -m_{\text {rna }} g L_{d} \cos \theta_{t} \\
& +m_{\mathrm{tmd}} g x_{\mathrm{tmd}} \cos \theta_{t}+m_{\mathrm{tmd}} g L_{\mathrm{rna}} \sin \theta_{t}, \\
& F_{\mathrm{sg}}^{\text {hdr.moor }}=-D_{\mathrm{sg}}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}-\widehat{D}_{\mathrm{sg}}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}^{2}-K_{\mathrm{sg}}^{\mathrm{sg}} x_{\mathrm{sg}}-D_{\mathrm{sg}}^{p} \dot{\theta}_{p}-K_{\mathrm{sg}}^{p} \theta_{p}, \\
& F_{\mathrm{hv}}^{\mathrm{hdr} . \text { moor }}=-D_{\mathrm{hv}}^{\mathrm{hv}} \dot{x}_{\mathrm{hv}}-K_{\mathrm{hv}}^{\mathrm{hv}} x_{\mathrm{hv}}-F_{\text {moor }}+F_{\text {buoy }} \\
& -K_{\mathrm{hv}}^{p . \mathrm{sg}}\left(x_{\mathrm{sg}}-L_{\text {moor }} \sin \theta_{p}\right)^{2}, \\
& \tau_{p}^{\text {hdr.moor }}=-D_{p}^{p} \dot{\theta}_{p}-K_{p}^{p} \theta_{p}-D_{\mathrm{p}}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}-\widehat{D}_{p}^{\mathrm{sg}} \dot{x}_{\mathrm{sg}}^{2}-K_{p}^{\mathrm{sg}} x_{\mathrm{sg}}, \\
& F_{\mathrm{sg}}^{\mathrm{ctr}}=m_{\mathrm{twr}} \dot{\theta}_{p}^{2}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \sin \theta_{p} \\
& +m_{\mathrm{rna}} \dot{\theta}_{t}^{2}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \sin \theta_{t} \\
& -m_{\mathrm{ptfm}} \dot{\theta}_{p}^{2} L_{\mathrm{ptfm}} \sin \theta_{p}
\end{aligned}
$$

$$
\begin{align*}
&+ m_{\mathrm{tmd}} \dot{\theta}_{t}\left(\dot{\theta}_{t}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right)-\dot{x}_{\mathrm{tmd}}\right) \sin \theta_{t} \\
& F_{\mathrm{hv}}^{\mathrm{ctr}}= m_{\mathrm{twr}} \dot{\theta}_{p}^{2}\left(L_{\mathrm{twr}}+L_{\mathrm{jot}}\right) \cos \theta_{p} \\
&+m_{\mathrm{rna}} \dot{\theta}_{t}^{2}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right) \cos \theta_{t} \\
&-m_{\mathrm{ptfm}} \dot{\theta}_{p}^{2} L_{\mathrm{ptfm}} \cos \theta_{p} \\
&+m_{\mathrm{tmd}} \dot{\theta}_{t}\left(\dot{\theta}_{t}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right)-\dot{x}_{\mathrm{tmd}}\right) \cos \theta_{t}, \\
& \tau_{p}^{\mathrm{ctr}}=\tau_{t}^{\mathrm{ctr}}=-m_{\mathrm{tmd}} \dot{\theta}_{t}\left(\dot{\theta}_{t}\left(L_{\mathrm{rna}}+L_{\mathrm{jot}}\right)-\dot{x}_{\mathrm{tmd}}\right) x_{\mathrm{tmd}} \\
& F_{\mathrm{tmd}}^{\mathrm{spr} \cdot \mathrm{damp}}=-D_{\mathrm{tmd}} \dot{x}_{\mathrm{tmd}}-K_{\mathrm{tmd}} x_{\mathrm{tmd}} \\
& \tau_{t}^{p}=D_{t}\left(\dot{\theta}_{t}-\dot{\theta}_{p}\right)+K_{t}\left(\theta_{t}-\theta_{p}\right) . \tag{A.1}
\end{align*}
$$

## Acknowledgments

This work has been (partially) funded by the Norwegian Centre for Offshore Wind Energy (NORCOWE) under Grant 193821/S60 from the Research Council of Norway (RCN). NORCOWE is a consortium with partners from industry and science, hosted by Christian Michelsen Research. The authors would like to give sincere thanks to Dr. Lackner in University of Massachusetts Amherst for his generosity on the FASTSC code sharing and Dr. Jonkman from National Renewable Energy Laboratory for his enthusiastic help and support with the specifications of OC3-Hywind model.

## References

[1] J. Jonkman, Dynamics modeling and loads analysis of an offshore floating wind turbine [Ph.D. thesis], Department of Aerospace Engineering Sciences, University of Colorado, Boulder, Colo, USA, 2007.
[2] J. M. Jonkman, "Dynamics of offshore floating wind turbinesmodel development and verification," Wind Energy, vol. 12, no. 5, pp. 459-492, 2009.
[3] S. Butterfield, W. Musial, J. Jonkman, P. Sclavounos, and L. Wayman, "Engineering challenges for floating offshore wind turbines," in Proceedings of the Copenhagen Offshore Wind 2005 Conference and Expedition, Copenhagen, Denmark, October 2005.
[4] J. Jonkman and D. Matha, "A quantitative comparison of the responses of three floating platform concepts," in Proceedings of the European Offshore Wind 2009 Conference and Exhibition, pp. 14-16, 2009.
[5] T. J. Larsen and T. D. Hanson, "A method to avoid negative damped low frequent tower vibrations for a floating, pitch controlled wind turbine," Journal of Physics, vol. 75, no. 1, Article ID 012073, 2007.
[6] S. Korkmaz, "A review of active structural control: challenges for engineering informatics," Computers and Structures, vol. 89, no. 23-24, pp. 2113-2132, 2011.
[7] P. J. Murtagh, A. Ghosh, B. Basu, and B. M. Broderick, "Passive control of wind turbine vibrations including blade/tower interaction and rotationally sampled turbulence," Wind Energy, vol. 11, no. 4, pp. 305-317, 2008.
[8] S. Colwell and B. Basu, "Tuned liquid column dampers in offshore wind turbines for structural control," Engineering Structures, vol. 31, no. 2, pp. 358-368, 2009.
[9] J. Li, Z. Zhang, and J. Chen, "Experimental study on vibration control of offshore wind turbines using a ball vibration absorber," Energy and Power Engineering, vol. 4, no. 3, pp. 153157, 2012.
[10] J. Jonkman and M. Buhl Jr., "Fast users guide," Tech. Rep. NREL/EL-500-29798, National Renewable Energy Laboratory, Golden, Colo, USA, 2005.
[11] M. A. Lackner and M. A. Rotea, "Passive structural control of offshore wind turbines," Wind Energy, vol. 14, no. 3, pp. 373-388, 2011.
[12] M. A. Lackner and M. A. Rotea, "Structural control of floating wind turbines," Mechatronics, vol. 21, no. 4, pp. 704-719, 2011.
[13] G. M. Stewart and M. A. Lackner, "The effect of actuator dynamics on active structural control of offshore wind turbines," Engineering Structures, vol. 33, no. 5, pp. 1807-1816, 2011.
[14] G. Stewart, Load reduction of floating wind turbines using tuned mass dampers [M.S. thesis], University of Massachusetts Amherst, Amherst, Mass, USA, 2012.
[15] G. M. Stewart and M. A. Lackner, "Offshore wind turbine load reduction employing optimal passive tuned mass damping systems," IEEE Transactions on Control Systems Technology, vol. 21, no. 4, pp. 1090-1104, 2013.
[16] J. Jonkman, "Definition of the floating system for phase iv of oc3," Tech. Rep. NREL/TP- 500-38060, National Renewable Energy Laboratory, Golden, Colo, USA, 2009.
[17] H. Namik and K. Stol, "Performance analysis of individual blade pitch control of offshore wind turbines on two floating platforms," Mechatronics, vol. 21, no. 4, pp. 691-703, 2011.
[18] J. Jonkman, S. Butterfield, W. Musial, and G. Scott, "Definition of a $5-\mathrm{mw}$ reference wind turbine for offshore system development," Tech. Rep. NREL/TP-500-47535, National Renewable Energy Laboratory, Golden, Colo, USA, 2009.
[19] S. Rao and R. Durgaiah, Engineering Mechanics, Universities Press, Andhra Pradesh, India, 2005.
[20] J. Newman, Marine Hydrodynamics, MIT press, Cambridge, Mass, USA, 1977.
[21] H. Kooijman, C. Lindenbug, D. Winkelaar, and E. Van der Hooft, "Dowec 6 mw predesign: aeroelastic modeling of the dowec 6 mw pre-design in phatas," Tech. Rep. DOWEC-F1W2-HJK-01-046/9, Energy research Centre of the Netherlands, Amsterdam, The Netherlands, 2003.
[22] J. P. Den Hartog, Mechanical Vibrations, Dover, New York, NY, USA, 1985.
[23] D. Matha, "Model development and loads analysis of an offshore wind turbine on a tension leg platform with a comparison to other floating turbine concepts," Tech. Rep. NREL/SR-500-45891, National Renewable Energy Laboratory (NREL), Golden, Colo, USA, 2010.
[24] H. Wang and H. Ohmori, "Elasto-plastic analysis based truss optimization using genetic algorithm," Engineering Structures, vol. 50, pp. 1-12, 2013.
[25] J. D. Poirier, S. S. Vel, and V. Caccese, "Multi-objective optimization of laser-welded steel sandwich panels for static loads using a genetic algorithm," Engineering Structures, vol. 49, pp. 508-524, 2013.
[26] F. B. Ramires, S. A. L. d. Andrade, P. C. G. d. S. Vellasco, and L. R. O. d. Lima, "Genetic algorithm optimization of composite and steel endplate semi-rigid joints," Engineering Structures, vol. 45, pp. 177-191, 2012.

Research Article

# Reference Function Based Spatiotemporal Fuzzy Logic Control Design Using Support Vector Regression Learning 

Xian-Xia Zhang, Ye Jiang, Shiwei Ma, and Bing Wang<br>Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronics and Automation, Shanghai University, Shanghai 200072, China<br>Correspondence should be addressed to Xian-Xia Zhang; xianxia_zh@shu.edu.cn

Received 25 August 2013; Accepted 1 November 2013
Academic Editor: Baocang Ding
Copyright © 2013 Xian-Xia Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper presents a reference function based 3D FLC design methodology using support vector regression (SVR) learning. The concept of reference function is introduced to 3D FLC for the generation of 3D membership functions (MF), which enhance the capability of the 3D FLC to cope with more kinds of MFs. The nonlinear mathematical expression of the reference function based 3D FLC is derived, and spatial fuzzy basis functions are defined. Via relating spatial fuzzy basis functions of a 3D FLC to kernel functions of an SVR, an equivalence relationship between a 3D FLC and an SVR is established. Therefore, a 3D FLC can be constructed using the learned results of an SVR. Furthermore, the universal approximation capability of the proposed 3D fuzzy system is proven in terms of the finite covering theorem. Finally, the proposed method is applied to a catalytic packed-bed reactor and simulation results have verified its effectiveness.


## 1. Introduction

Many industrial processes and systems are "distributed" in space [1] and are usually called spatially distributed systems. Recently, a novel three-dimensional fuzzy logic controller (3D FLC) [2] has been developed for the control of such systems. The 3D FLC employs a three-dimensional (3D) fuzzy set [2], which is composed of the traditional fuzzy set plus a third dimension for the spatial information, and carries out a 3 D rule inference engine; thus, it has the inherent capability to process spatiotemporal dynamic systems. The control strategy of the 3D FLC is similar to how human operators or experts control the temperature in a space domain. Actually, it is a kind of spatiotemporal fuzzy control system with the traditional model-free advantage.

Currently, most 3D FLC designs are based on expert knowledge [2-5], which requires that the human knowledge to the control solution must exist and be structured [6]. However, in many real-world applications, experts may have problems structuring the knowledge. Sometimes, although experts have the structured knowledge, they may sway between extreme cases: offering too much knowledge in the field of expertise or tending to hide their knowledge [6]. On
the contrast, data sets hidden with effective control rules are usually available. The motivation of this study is to design a 3D FLC using spatiotemporal data information.

To date, few literatures are found to be focused on databased 3D FLC design methods. In [7], a table look-up scheme was employed to design 3D FLC in terms of input-output pairs. In [8], a fuzzy c-means algorithm (FCM) and gradientdescent approach were used to design a data-based 3D FLC, where FCM was used to learn the initial 3D fuzzy rule base and then the gradient-descent approach was used to optimize the parameters of MFs. In [9, 10], a clustering and linear support vector regression based 3D FLC design method was proposed, where the nearest neighborhood clustering was used to construct the antecedent part of 3D fuzzy rules and a linear support vector regression (SVR) was used to learn the consequent parameters. These methods either yield lots of fuzzy control rules (e.g. in [7]) or require additional algorithm to reduce redundant 3D fuzzy sets or 3D fuzzy rules [810]. As a complementary, Zhang et al. proposed a data-based 3D FLC design method using SVR learning [11], where the learned support vectors and associated learning parameters are directly used to design antecedent part and consequent part of 3D fuzzy rules. The best advantage of this method is
that reasonable 3D fuzzy control rules are directly extracted and constructed by SVR learning. The limitation of the design is that Gaussian shape membership function (MF) is the only choice for MF design.

In this study, we focus on a reference function based 3D FLC design using SVR learning, which integrates the merits of SVR learning and flexible MF choice. Utilizing the concept of reference function, the 3D FLC can cope with more kinds of MFs, for example, Symmetric triangle, Gaussian, Cauchy, Laplace, Hyperbolic Secant, and Squared Sinc. A nonlinear mathematical description of a reference function based 3D FLC can be derived, and spatial fuzzy basis functions are defined. Via relating spatial fuzzy basis functions of a 3D FLC to kernel functions (KFs) of an SVR, an equivalence relationship between a 3D FLC and an SVR is established. Therefore, a 3D FLC can be constructed using the learned results of an SVR. In addition, the universal approximation capability of the proposed 3D fuzzy system is proven in terms of the finite covering theorem.

The paper is organized as follows. Preliminaries about the reference function, 3D MF generated by reference function, and the nonlinear mapping of reference function based 3D FLC are addressed in Section 2. In Section 3, the methodology and design scheme of the reference function based 3D FLC design using SVR learning are presented. Then, the finite covering theorem is used to prove that the 3D FLC is a universal approximator in Section 4. In Section 5, a catalytic packed-bed reactor is presented as an example to illustrate the proposed 3D FLC and validate its effectiveness. In Section 6, conclusions are given.

## 2. Preliminaries

### 2.1. Reference Function

Definition of Reference Function (see [12, 13]). A function $v$ : $R \rightarrow[0,1]$ is a reference function if and only if the following two conditions hold:
(1) $v(x)=v(-x)$,
(2) $v(0)=1$.

Many functions may be reference functions. For instance, Symmetric triangle, Gaussian, Cauchy, Laplace, Hyperbolic Secant, and Squared Sinc as listed in Table 1 are reference functions. The reference functions can be used to generate 3D MFs, which provide a way for 3D FLC to access more kinds of 3D MFs.

### 2.2. Reference Function Based 3D FLC

2.2.1. 3D MF Generated by Reference Function. A 3D MF is an extension of a traditional MF by adding a third coordinate for the spatial information. In detail, the 3D MF has three coordinates: one is for the universe of discourse of the variable, another one is for the spatial information, and the third one is for the membership degree. If finite sensors are used, the 3D MF can be considered as the assembly of the traditional 2D MFs at each sensing location. In this way,

Table 1: Reference functions.

| Classification | Mathematical expression |
| :--- | :---: |
| Symmetric triangle | $v(x)=\max (1-d\|x\|, 0), \quad d>0$ |
| Gaussian | $v(x)=e^{-d x^{2}}, \quad d>0$ |
| Cauchy | $v(x)=\frac{1}{1+d x^{2}}, \quad d>0$ |
| Laplace | $v(x)=e^{-d\|x\|}, \quad d>0$ |
| Hyperbolic Secant | $v(x)=\frac{2}{e^{d x}+e^{-d x}}, d>0$ |
| Squared Sinc | $v(x)=\frac{\sin ^{2}(d x)}{d^{2} x^{2}}, \quad d>0$ |



Figure 1: Gaussian MF distribution of spatial input variable $x_{1}(z)$ at sensing location $z=z_{1} \cdot \beta_{1}^{1}\left(x_{1}\left(z_{1}\right)\right), \beta_{1}^{2}\left(x_{1}\left(z_{1}\right)\right)$, and $\beta_{1}^{3}\left(x_{1}\left(z_{1}\right)\right)$ are generated by the Gaussian type reference function $\beta_{1}\left(x_{1}\left(z_{1}\right)\right)$.
we can generate a 3D MF by location transformation of a reference function at each sensing location.

For example, we have a spatial input variable $x_{1}(z)$ defined in a discrete spatial domain $Z=\left\{z_{1}, z_{2}, \ldots, z_{p}\right\}$. A 3D MF of $x_{1}(z)$ can be an assembly of the traditional 2D MFs at each sensing location $z=z_{j}(j=1,2, \ldots, p)$. The MF distribution of $x_{1}(z)$ at sensing location $z=z_{1}$ can be shown in Figure 1, where $\beta_{1}\left(x_{1}\left(z_{1}\right)\right)$ is a Gaussian type reference function; $\beta_{1}^{1}\left(x_{1}\left(z_{1}\right)\right), \beta_{1}^{2}\left(x_{1}\left(z_{1}\right)\right)$, and $\beta_{1}^{3}\left(x_{1}\left(z_{1}\right)\right)$ are generated by location transformation of $\beta_{1}\left(x_{1}\left(z_{1}\right)\right)$.
2.2.2. Reference Function Based 3D FLC as a Nonlinear Mapping. The basic structure of a 3D FLC is composed of 3D fuzzifier, 3D rule inference, and defuzzifier. Due to its unique 3D nature, some detailed operations of a 3D FLC are different from a traditional one for spatial information expression, processing, and compression. For their detailed operations, one can refer to [2]. Once each component of a reference function based 3D FLC is set, the nonlinear mathematical description of the 3D FLC can be derived (see Appendix A for a brief derivation). Assuming that we employ 3D singleton fuzzifier, 3D fuzzy rules as shown in (A.4) of Appendix A, "product" t-norm and "weighted aggregation" dimension reduction [3] in the 3D rule inference, singleton fuzzy sets for the output variable, and "linear" defuzzifier [14],
the reference function based 3D FLC can be mathematically expressed as

$$
\begin{align*}
u\left(x_{z}\right) & =b^{0}+\sum_{l=1}^{N} b^{l} \sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}^{l}\left(x_{i}\left(z_{j}\right)\right) \\
& =b^{0}+\sum_{l=1}^{N} b^{l} \sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right), \tag{1}
\end{align*}
$$

where $x_{i}\left(z_{j}\right)$ denotes the input of the $i$ th spatial input variable $x_{i}(z)$ from the sensing location $z=z_{j} ; \beta_{i}^{l}\left(x_{i}\left(z_{j}\right)\right)=$ $\beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right)$ denotes MF generated by the reference function $\beta_{i}\left(x_{i}\left(z_{j}\right)\right) ; \tau_{i j}^{l}$ denotes a location parameter, that is, the location transformation of the MF of $x_{i}\left(z_{j}\right)$ with respect to $\beta_{i}\left(x_{i}\left(z_{j}\right)\right) ; a_{j}$ denotes the spatial weight from the $j$ th sensing location; $b^{0}$ and $b^{l}$ are constants; $p$ denotes sensor number.

In (1), let

$$
\begin{equation*}
\Psi^{l}\left(x_{z}\right)=\sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}^{l}\left(x_{i}\left(z_{j}\right)\right) \tag{2}
\end{equation*}
$$

then (1) can be rewritten as

$$
\begin{equation*}
u\left(x_{z}\right)=b^{0}+\sum_{l=1}^{N} b^{l} \Psi^{l}\left(x_{z}\right) \tag{3}
\end{equation*}
$$

We define $\Psi^{l}\left(x_{z}\right)$ as a Spatial Fuzzy Basis Function (SFBF) [11]. Each SFBF corresponds to a 3D fuzzy rule, and all the SFBFs correspond to a 3D rule base. Mathematically, a 3D FLC is a linear combination of all the SFBFs. Furthermore, we rewrite (2) into (4)

$$
\begin{equation*}
\Psi^{l}\left(x_{z}\right)=\sum_{j=1}^{p} a_{j} \varphi^{l}\left(x\left(z_{j}\right)\right) \tag{4}
\end{equation*}
$$

where $\varphi^{l}\left(x\left(z_{j}\right)\right)=\prod_{i=1}^{s} \beta_{i}^{l}\left(x_{i}\left(z_{j}\right)\right)$.
From (4), we can find that, at each sensing location, there exists a traditional FBF [15] $\varphi^{l}\left(x\left(z_{j}\right)\right)$; in the whole space domain, multiple traditional FBFs are assembled by the spatial weights $a_{1}, \ldots, a_{p}$ into a $\operatorname{SFBF} \Psi^{l}\left(x_{z}\right)$. All the spatial information expression and processing as well as the fuzzy linguistic expression and rule inference are integrated into SFBFs.

Equation (1) (or (3)) shows that the reference function based 3D FLC is a nonlinear mapping from the input space $x_{z} \in \Omega \subset R^{p \times s}$ to the output space $u\left(x_{z}\right) \in U \subset R$. In particular, using (3) a reference function based 3D FLC can be represented by a three-layer network structure as show in Figure 2.

## 3. Reference Function Based 3D FLC Design Using SVR Learning

3.1. Design Methodology. The design methodology can be depicted by Figure 3. The SFBFs from a 3D FLC are input to


Figure 2: Three-layer network structure of a 3D FLC.
an SVR as the KFs, and the learned spatial support vectors as leading spatiotemporal data points from the SVR are imported for the design of a 3D fuzzy rule base. The design theory will involve two crucial issues. The first one is whether SFBFs from a 3D FLC can be used as KFs in an SVR. The second one is what the relationship between a 3D FLC and an SVR would be like on the basis of the first issue.
3.1.1. Spatial Fuzzy Basis Function as Mercer KF. When relating the SFBFs with the KFs in an SVR, for instance, SFBFs are regarded as KFs, the SVR and the 3D FLC will have the same network structures and then have the same mathematical expressions, which will be discussed in Section 3.1.2. Generally speaking, a function satisfying Mercer theorem can be used as a KF for an SVR [16]. In this study, we will prove that an SFBF is a Mercer KF.

In (3), we rewrite $\Psi^{l}\left(x_{z}\right)$ into $K\left(x_{z}, \tau^{l}\right)$, which can be further expressed as follows:

$$
\begin{gather*}
K\left(x_{z}, \tau^{l}\right)=K\left(x_{z}-\tau^{l}\right)=\sum_{j=1}^{p} a_{j} K^{j}\left(x\left(z_{j}\right), \tau_{j}^{l}\right), \\
K^{j}\left(x\left(z_{j}\right), \tau_{j}^{l}\right)=K^{j}\left(x\left(z_{j}\right)-\tau_{j}^{l}\right)=\prod_{i=1}^{s} \beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right), \tag{5}
\end{gather*}
$$

where $x_{z} \in R^{p \times s}$ is a spatial input, $\tau^{l} \in R^{p \times s}$ is the location transformation parameter of 3D MF in the $l$ th rule, and $K^{j}\left(x\left(z_{j}\right), \tau_{j}^{l}\right)$ and $K\left(x_{z}, \tau^{l}\right)$ are translation invariant KFs [13].

In terms of [13], if the reference functions are positive definite functions, then we do get a Mercer kernel. The reference functions as listed in Table 1 are positive definite functions. Using these reference functions to generate MF, from [13], we can conclude that $K^{j}\left(x\left(z_{j}\right), \tau_{j}^{l}\right)$ is a Mercer kernel. Since the linear combination of KFs is still a KF [16], we can derive that $K\left(x_{z}, \tau^{l}\right)$ is still a Mercer KF. Therefore, SFBFs are Mercer KFs, which can be used as KFs for SVR learning.


FIGURE 3: Design methodology of reference function based 3D FLC design using SVR learning.


Figure 4: Design scheme of reference function based 3D FLC using SVR learning.
3.1.2. Mathematical Equivalence of a Spatial SVR and a 3D FLC. Once the SFBFs from the 3D FLC are employed as the KFs for an SVR, an inherent equivalence relationship will be built between the SVR and the 3D FLC (comparing Figure 2 and Figure 10). By combining (1) and (B.7), we have the following mathematical expressions:

$$
\begin{align*}
u\left(x_{z}\right) & =b^{0}+\sum_{l=1}^{N} b^{l} \Psi^{l}\left(x_{z}\right) \\
& =b^{0}+\sum_{l=1}^{N} b^{l} \sum_{j=1}^{p} a_{j} \varphi^{l}\left(x\left(z_{j}\right)\right)  \tag{6}\\
& =b+\sum_{l=1}^{N}\left(\alpha_{l}^{*}-\alpha_{l}\right) K\left(x_{z}, x_{z}^{l}\right)
\end{align*}
$$

where $b^{0}=b, b^{l}=\left(\alpha_{l}^{*}-\alpha_{l}\right)$, and $\Psi^{l}\left(x_{z}\right)=K\left(x_{z}, \tau^{l}\right)=$ $K\left(x_{z}, x_{z}^{l}\right)$.

From (6), we can find that each spatial support vector $x_{z}^{l}$ and its associated learning parameter $\left(\alpha_{l}^{*}-\alpha_{l}\right)$ correspond to one 3D fuzzy rule, where $x_{z}^{l}$ is applied to set the center of the 3D MF of the 3D fuzzy set $\bar{C}_{i}^{l}(i=1, \ldots, s)$ in the $l$ th rule, that is, the location transformation of the 3D fuzzy set with respect to reference function $\beta\left(x_{z}, \tau^{l}\right)$, and $\left(\alpha_{l}^{*}-\alpha_{l}\right)$ is used
to set $b^{l}$ (the constant for the consequent set of the $l$ th rule in 3D FLC).
3.2. Design Scheme. The design of a reference function based 3D FLC consists of five parts: data collection, KF generation, SVR learning, 3D fuzzy rule construction, and 3D fuzzy controller integration, as shown in Figure 4.
(1) Data Collection. A set of spatiotemporal data will be collected. The data should contain effective control laws. Essentially, the reference function based 3D FLC design is a fuzzy modeling [17] that extracts fuzzy control rules from the spatiotemporal data.
(2) KF Generation. Before SVR learning, KFs should be properly designed. In this step, via properly selecting reference function, SFBFs from a 3D FLC will be formulated (as in (4)) to set KFs for SVR learning.
(3) SVR Learning. With proper KFs, the SVM learning algorithm directly executes the spatiotemporal data set and yields spatial support vectors $x_{z}^{1}, \ldots, x_{z}^{N}$ and their associated learning parameters $\alpha_{1}^{*}-\alpha_{1}, \ldots, \alpha_{N}^{*}-\alpha_{N}$.
(4) 3D Fuzzy Rule Construction. The spatial support vectors and their associated learning parameters, as leading control laws, are used to construct 3D fuzzy control rules. In detail,


Figure 5: Sketch of a catalytic packed-bed reactor.
the spatial support vector $x_{z}^{l}$ is employed to construct the antecedent part of the $l$ th rule; $\alpha_{l}^{*}-\alpha_{l}$ is employed to construct the consequent part of the $l$ th rule. The form of each 3D fuzzy rule is shown as below

$$
\begin{equation*}
\bar{R}^{l}: \text { if } x_{z} \text { is close to } x_{z}^{l} \text { then } u \text { is close to }\left(\alpha_{l}^{*}-\alpha_{l}\right) \tag{7}
\end{equation*}
$$

It is shown that the result of the SVM learning can be easily interpreted using structured linguistic knowledge. Finally, a 3D rule base with $N$ rules is established.
(5) 3D Fuzzy Controller Integration. Once the 3D rule base is established, a 3D FLC can be achieved by integrating other components including 3D fuzzifier, 3D rule inference, and defuzzifier. The detailed settings are given in Section 2.2.2. Finally, we obtain a complete 3D FLC, which can be used as a controller for a spatially distributed dynamic system.

## 4. Universal Approximation of Reference Function Based 3D FLC

The reference function based 3D FLC design method is used to construct a 3D FLC from spatiotemporal data hidden with effective control laws. In other words, the 3D FLC aims at approximating an unknown nonlinear control function. In this study, we use the finite covering theorem to prove that the 3D FLC is a universal approximator; that is, it can approximate continuous control functions to arbitrary accuracy.

The universal approximation capability of the reference function based 3D FLC can be described by Theorem 1.

Theorem 1. Let $\widetilde{g}\left(x_{z}\right): R^{p \times s} \rightarrow R$ be a continuous function defined on a compact $\Omega$. For each $\varepsilon>0$, there exists a reference function based 3D FLC $u\left(x_{z}\right)$ such that

$$
\begin{equation*}
\sup _{x_{z} \in \Omega}\left(\left|u\left(x_{z}\right)-\tilde{g}\left(x_{z}\right)\right|\right)<\varepsilon . \tag{8}
\end{equation*}
$$

From (A.4), it is shown that $\bar{R}^{0}$ is an universal rule, namely, for any spatial input $x_{z}, \bar{R}^{0}$ will be fired. In (A.4), the fired rule $\bar{R}^{0}$ will produce the constant $b^{0}$. Let

$$
\begin{align*}
F\left(x_{z}\right) & =u\left(x_{z}\right)-b^{0} \\
& =\sum_{l=1}^{N} b^{l} \sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right)  \tag{9}\\
& =\sum_{l=1}^{N} b^{l} K\left(x_{z}, \tau^{l}\right) .
\end{align*}
$$



Figure 6: Spatial-temporal data set: (a) spatial error $e(z)$; (b) spatial error in change $\Delta e(z)$; (c) incremental output $\Delta u$ ( $z$ : spatial dimension; $k$ : serial number of input-output data).
$F\left(x_{z}\right)$ can be regarded as a 3D FLC generated by rule base $\left\{\bar{R}^{1}, \bar{R}^{2}, \ldots, \bar{R}^{N}\right\}$. Then, Theorem 1 can be restated as Theorem 2 as follows.

Theorem 2. Under the condition of Theorem 1, let $g\left(x_{z}\right)$ : $R^{p \times s} \rightarrow R$ be a continuous function defined on a compact $\Omega$. For any constant $b^{0}$, one has $\widetilde{g}\left(x_{z}\right)=g\left(x_{z}\right)+b^{0}$. For each $\varepsilon>0$, there exists a reference function based 3D FLC F $\left(x_{z}\right)$ such that

$$
\begin{equation*}
\sup _{x_{z} \in \Omega}\left(\left|F\left(x_{z}\right)-g\left(x_{z}\right)\right|\right)<\varepsilon . \tag{10}
\end{equation*}
$$



Figure 7: Input data and support vectors from each sensing location. Circle (o): support vector; dot (+): input data.


Figure 8: The first four 3D fuzzy rules and their associated MFs $(b=-0.0752)$.

(a) Controlled by a reference function based 3D FLC with Symmetric triangle reference function

(b) Controlled by a reference function based 3D FLC with Gaussian reference function

(c) Controlled by a reference function based 3D FLC with Cauchy reference function

(d) Controlled by a reference function based 3D FLC with Laplace reference function

Figure 9: Continued.

(g) Controlled by an expert based 3D FLC

Figure 9: Control performance comparisons. From left to right in (a)-(g): catalyst temperature varying with time and space, manipulated input, and catalyst temperature at steady state.

Before the proof of Theorem 2, we first present some preparation work.

When $x_{z}=\bar{x}_{z} \in \Omega$, the firing level of the fired rule $\bar{R}^{l}(l=$ $1, \ldots, N)$ is

$$
\begin{equation*}
\mu_{\varphi^{l}}\left(\bar{x}_{z}\right)=\sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}\left(\bar{x}_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right) . \tag{11}
\end{equation*}
$$

The inference result of $\bar{R}^{l}$ is given by

$$
D^{l}(u)= \begin{cases}0 & \text { if } \mu_{\varphi^{l}}\left(\bar{x}_{z}\right)=0 \text { or } u \neq b^{l},  \tag{12}\\ b^{l} \mu_{\varphi^{l}}\left(\bar{x}_{z}\right) & \text { in other case, }\end{cases}
$$

where $u$ is the output variable of the 3D FLC, which is corresponding to the " $u$ " of the consequent part of the fired
rule. The composition result of all the fired rules is given as follows:

$$
\begin{equation*}
D(u)=\bigcup_{l=1}^{N} D^{l}(u) . \tag{13}
\end{equation*}
$$

Based on the above preparation, Lemma 3 is presented as follows.

Lemma 3. Under the condition of Theorem 1 there exists a reference function based $3 D \operatorname{FLC} F\left(x_{z}\right)$ such that

$$
\begin{equation*}
D(u)\left|u-g\left(x_{z}\right)\right| \leq D(u) * \varepsilon \quad \text { for each } u \in R \tag{14}
\end{equation*}
$$

Proof. Let $a_{z} \in \Omega$. As $g(\cdot)$ is continuous at $a_{z}$, for each $i=$ $1, \ldots, s$ there exists a $\delta_{a_{z}}^{i}>0$ such that

$$
\begin{equation*}
\left|x_{z}^{i}-a_{z}^{i}\right| \leq \delta_{a_{z}}^{i} \quad i=(1, \ldots, s) \Longleftrightarrow\left|g\left(x_{z}\right)-g\left(a_{z}\right)\right| \leq \varepsilon \tag{15}
\end{equation*}
$$

For each $a_{z} \in \Omega$, set

$$
\begin{equation*}
O_{a_{z}}=\left\{x_{z}| | x_{z}^{i}-a_{z}^{i} \mid \leq \delta_{a_{z}}^{i}(i=1, \ldots, s)\right\} . \tag{16}
\end{equation*}
$$

Then, $O_{a_{z}}$ is open on $R^{p \times s}$ and $\Omega \subseteq \bigcup_{a_{z} \in \Omega} O_{a_{z}}$. As $\Omega$ is compact, there exists a finite subfamily $O_{a_{z}^{1}}, O_{a_{z}^{2}}, \ldots, O_{a_{z}^{k}}$ such that

$$
\begin{equation*}
\Omega \subseteq O_{a_{z}^{1}} \bigcup O_{a_{z}^{2}} \bigcup \cdots \bigcup O_{a_{z}^{k}} \tag{17}
\end{equation*}
$$

We can construct a 3D FLC $F\left(x_{z}\right)$, defined by

$$
\begin{align*}
& \mu_{\varphi^{l}}\left(x_{z}\right)= \begin{cases}\sum_{j=1}^{p} a_{j}^{l} \prod_{i=1}^{s} \beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right) & x_{z} \in O_{a_{z}^{l}} \\
0 & x_{z} \notin O_{a_{z}^{l}}\end{cases}  \tag{18}\\
& (l=1, \ldots, N),
\end{aligned} \begin{aligned}
& b^{l}=\left\{\begin{array}{ll}
g\left(x_{z}\right) & x_{z} \in O_{a_{z}^{l}} \\
0 & x_{z} \notin O_{a_{z}^{l}}
\end{array} \quad(l=1, \ldots, N) .\right.
\end{align*}
$$

When $x_{z}=\bar{x}_{z} \in \Omega$, we have the following.
(1) If $D(u)=0$, the lemma is trivial.
(2) If $D(u)>0$, then $D(u)=\bigcup_{l=1}^{N} D^{l}(u)>0$; hence, there exists a $k \in[1, N]$ such that $D^{k}(u)>0$. Therefore, we further have that the following.
(a) From $D^{k}(u) \neq 0$, it follows that $\mu_{\varphi^{k}}\left(\bar{x}_{z}\right) \neq 0$; that is, $\bar{x}_{z} \in O_{a_{z}^{k}}$. In terms of the continuity of $g(\cdot)$, we have $\left|g\left(\bar{x}_{z}\right)-g\left(a_{z}^{k}\right)\right| \leq \varepsilon$.
(b) From $D^{k}(u) \neq 0$, it follows that $D^{k}(u)=$ $b^{k} \mu_{\varphi^{k}}\left(\bar{x}_{z}\right) \neq 0$, and then $b^{k} \neq 0$. We have $u=b^{k}=$ $g\left(a_{z}^{k}\right)$.

Hence, we have

$$
\begin{equation*}
\left|u-g\left(\bar{x}_{z}\right)\right| \leq\left|u-g\left(a_{z}^{k}\right)\right|+\left|g\left(a_{z}^{k}\right)-g\left(\bar{x}_{z}\right)\right|<0+\varepsilon=\varepsilon . \tag{20}
\end{equation*}
$$

In terms of Lemma 3, the proof of Theorem 2 can be given as follows.

Proof of Theorem 2. Consider

$$
\begin{align*}
\left|F\left(\bar{x}_{z}\right)-g\left(\bar{x}_{z}\right)\right| & \leq \frac{\sum D(u)\left|u-g\left(\bar{x}_{z}\right)\right|}{\sum D(u)}  \tag{21}\\
& \leq \frac{\varepsilon * \sum D(u)}{\sum D(u)} \leq \varepsilon .
\end{align*}
$$

## 5. Application

5.1. A Catalytic Packed-Bed Reactor. This designed 3D FLC is applied to a catalytic packed-bed reactor $[1,4,18]$ shown in Figure 5, where a reaction of the form $A \rightarrow B$ takes place on the catalyst. The reaction is endothermic and a jacket is used to heat the reactor. A dimensionless model that describes this nonlinear tubular chemical reactor is provided as follows:

$$
\begin{align*}
\frac{\partial T_{r}}{\partial t}= & -\frac{\partial T_{r}}{\partial z}+\frac{1}{P_{e T}} \frac{\partial^{2} T_{r}}{\partial z^{2}} \\
& -B_{T} B_{C} C_{A} \exp \left(\frac{\gamma_{r} T_{r}}{1+T_{r}}\right)+\beta_{T}\left(u-T_{r}\right)  \tag{22}\\
\frac{\partial C_{A}}{\partial t}= & -\frac{\partial C_{A}}{\partial z}+\frac{1}{P_{e C}} \frac{\partial^{2} C_{A}}{\partial z^{2}}-B_{C} C_{A} \exp \left(\frac{\gamma_{r} T_{r}}{1+T_{r}}\right)
\end{align*}
$$

subject to the boundary conditions

$$
\begin{gather*}
z=0, \quad P_{e T} T_{r}=\frac{\partial T_{r}}{\partial z}, \quad P_{e C}\left(C_{A}-1\right)=\frac{\partial C_{A}}{\partial z}, \\
z=1, \quad \frac{\partial T_{r}}{\partial z}=0, \quad \frac{\partial C_{A}}{\partial z}=0, \tag{23}
\end{gather*}
$$

where $T_{r}, C_{A}$, and $u$ denote the dimensionless temperature, the concentration of reactant $A$, and jacket temperature, respectively; $t$ and $z$ denote the dimensionless time and space; $P_{e T}$ and $P_{e C}$ are the heat and mass Peclet numbers, $B_{T}$ is a dimensionless heat of reaction, $B_{C}$ is a dimensionless preexponential factor, $\gamma_{r}$ is a dimensionless activation energy, and $\beta_{T}$ is a dimensionless heat transfer coefficient. The values of the process parameters are given as follows:

$$
\begin{align*}
& P_{e T}=5.0, \quad P_{e C}=5.0, \quad B_{C}=0.00001, \\
& B_{T}=1.0, \quad \beta_{T}=15.62, \quad \gamma_{r}=22.14 . \tag{24}
\end{align*}
$$

The control problem is to maintain a desired reaction rate via tuning the jacket temperature to control catalyst temperature. In this application, the reactor began to work at one steady state; because of the requirement of operation conditions, the reference value of temperature is increased by $8 \%$. Thus, the control objective is to make the temperature of reactor well track the new reference value along the space domain.

### 5.2. Design of Reference Function Based 3D FLC

(1) Data Collection. The spatiotemporal data is collected from the catalytic packed-bed reactor, which is controlled by
an expert based 3D FLC [4]. Five point sensors are located along the length of the reactor with $Z=[00.40 .60 .81]$ for collecting the spatial distribution of the temperature $T_{r}$. Two spatial inputs are error and error in change; that is, $e^{*}(Z, k)=$ $T_{s d}(Z)-T_{s}(Z, k)$ and $\Delta e^{*}(Z)=e^{*}(Z, k)-e^{*}(Z, k-1)$. The detailed design of the expert based 3D FLC, including 3D fuzzifier, 3D rule inference, and defuzzifier, can refer to [4]. The scaling factors for the spatial error, the spatial error in change, and the incremental output are set as 0.5 , 0.1 , and 0.3 , respectively. The sample period is 0.1 s , and the sampling duration is 6 s . Thus, we have 60 input-output data pairs (shown in Figure 6), each of which is represented by $\left(\left[e(z)_{k}, \Delta e(z)_{k}\right], \Delta u_{k}\right), k=1, \ldots, 60$.
(2) KF Generation. The reference functions (Symmetric triangle, Gaussian, Cauchy, Laplace, Hyperbolic Secant, and Squared Sinc) as listed in Table 1 are employed, respectively, to generate 3D MFs and then formulate SFBFs. SFBFs are used for KFs in an SVR learning.
(3) SVR Learning. With the spatiotemporal data set as above, the SVM learning algorithm is used for the support vector learning. It should be noted that the SFBFs in (3) are taken as the KFs. As a result, spatial support vectors are extracted and their associated learning parameters are obtained. For instance, when the Gaussian type reference function is used for KF generation, eight support vectors were learned from 60 spatiotemporal data pairs when $C=1000, \varepsilon=0.00005$, and $d=0.1$, as shown in Figure 7, where the spatiotemporal input data is decomposed into multiple two-dimensional graphical representations over the space domain.
(4) $3 D$ Fuzzy Rule Construction. In terms of the learned results of the SVR in the previous step, we establish 3D fuzzy rules. For instance, with the Gaussian type reference function, eight 3D fuzzy rules are constructed. The first four 3D fuzzy rules are presented as follows.
$\bar{R}_{1}:$ if $e(z)$ is close to $[-0.0580 \quad-0.0963$ $-0.0988-0.1000-0.1005]^{\prime}$ and $\Delta e(z)$ is close to $\left[\begin{array}{lllll}-0.0580 & -0.0963 & -0.0988 & -0.1000 & -0.1005\end{array}\right]^{\prime}$, then $\Delta u$ is close to -9.5172 .
$\bar{R}_{2}:$ if $e(z)$ is close to $[-0.0479 \quad-0.0817$ $-0.0839-0.0848-0.0851]^{\prime}$ and $\Delta e(z)$ is close to $\left[\begin{array}{lllll}0.0101 & 0.0146 & 0.0149 & 0.0152 & 0.0154\end{array}\right]^{\prime}$, then $\Delta u$ is close to -298.9862 .
$\bar{R}_{3}:$ if $e(z)$ is close to $[-0.0414-0.0714$ $-0.0731-0.0735-0.0737]^{\prime}$ and $\Delta e(z)$ is close to $\left[\begin{array}{lllll}0.0065 & 0.0104 & 0.0109 & 0.0113 & 0.0114\end{array}\right]^{\prime}$, then $\Delta u$ is close to 249.2288 .
$\bar{R}_{4}:$ if $e(z)$ is close to $[-0.0359-0.0626$ $-0.0638-0.0639-0.0639]^{\prime}$ and $\Delta e(z)$ is close to $\left[\begin{array}{lllll}0.0055 & 0.0087 & 0.0092 & 0.0096 & 0.0098\end{array}\right]^{\prime}$, then $\Delta u$ is close to 148.1335 .

The first four 3D fuzzy rules and their associated 3D MF distributions can be depicted in Figure 8, which show the inherent spatial nature of the 3D fuzzy control system.
(5) 3D Fuzzy Controller Integration. Based on the 3D fuzzy rules established in step (4), we obtain a complete 3D FLC by combining 3D fuzzifier, 3D rule inference, and defuzzifier. The resultant 3D FLC will be used as a controller for the catalytic packed-bed reactor.
5.3. Control Performance Validation. The designed reference function based 3D FLC using SVR learning is validated on the nonlinear catalytic packed-bed reactor. We employed six kinds of reference functions, that is, Symmetric triangle, Gaussian, Cauchy, Laplace, Hyperbolic Secant, and Squared Sinc, and finally produced six 3D FLCs. With the reference function based 3D FLC as the controller, the catalyst temperature varying with time and space, manipulated input, and the catalyst temperature at steady state are presented in Figures 9(a)-9(f). The control performance is given in Table 2, where steady-state error (SSE), integral of the absolute error (IAE), and integral of time multiplied by absolute error (ITAE) [2] are used as the performance criteria. In terms of Figures 9 (a)-9(f) and Table 2, we can find that different reference functions will yield different control performance. In this application, Gaussian, Cauchy, Hyperbolic Secant, and Squared Sinc reference functions result in good control performance, while Symmetric triangle and Laplace reference functions lead to poor control performance. The results illustrate that KF will influence the control performance; thus, in the actual application, we should choose proper KF to design a 3D FLC.

As a comparison, we do another control experiment; that is, the expert knowledge-based 3D FLC from [4] is taken as the controller. As for its detailed design including 3D MF, 3D rule base, 3D inference, fuzzification, and defuzzification, one can refer to [4]. The scaling factors for the spatial error, the spatial error in change, and the incremental output are set as $0.5,0.1$, and 0.3 , respectively. The controlled profiles and control performance are given in Figure 9(g) and Table 2, respectively.

From Figure 9 and Table 2, we can find that with a proper reference function, the reference function based 3D FLC has nearly the same control performance as the expert knowledge-based 3D FLC. It means that the proposed spatial SVR learning method can well extract the control laws hidden in a spatiotemporal input-output dataset and formulate them in the form of 3D fuzzy rules.

## 6. Conclusions

A reference function based 3D FLC design methodology using SVR learning is proposed for spatially distributed dynamic systems. Utilizing the concept of reference function, the 3D FLC can access more kinds of 3D MFs, such as Symmetric triangle, Gaussian, Cauchy, Laplace, Hyperbolic Secant, and Squared Sinc. Based on the mathematical expressions of reference function based 3D FLC, we define spatial fuzzy basis functions and then find an equivalence

Table 2: Performance comparisons.

| Reference function based 3D FLC | 3D fuzzy rules | ISS $\left(\times 10^{-3}\right)$ | IAE $\left(\times 10^{-1}\right)$ | ITAE $\left(\times 10^{-1}\right)$ |
| :--- | :---: | :---: | :---: | :---: |
| Symmetric triangle reference function | 7 | 103.4 | 16.163 | 65.457 |
| Gaussian reference function | 8 | 4.7 | 1.864 | 3.526 |
| Cauchy reference function | 8 | 4.7 | 1.881 | 3.537 |
| Laplace reference function | 7 | 108 | 16.525 | 67.408 |
| Hyperbolic Secant reference function | 28 | 4.8 | 1.918 | 3.567 |
| Squared Sinc reference function | 5 | 4.7 | 1.859 | 3.515 |
| Expert-based 3D FLC | 25 | 4.8 | 1.931 | 3.676 |

relationship between a 3D FLC and an SVR by connecting spatial fuzzy basis functions in the 3D FLC to KFs in the SVR. On the basis of the equivalence relationship, a 3D FLC can be designed using the SVR learning; that is, the learned spatial support vectors as the optimal leading data points can be directly used for 3D fuzzy control rule generation. The proposed reference function based 3D FLC design can be carried out in five steps: data collection, KF generation, SVR learning, 3D fuzzy rule construction, and 3D fuzzy controller integration. Besides, the universal approximation capability of the proposed 3D fuzzy system is discussed. Finally, effectiveness of the proposed 3D FLC design methodology is validated on a catalytic packed-bed reactor.

## Appendices

## A. Nonlinear Mapping Derivation of a Reference Function Based 3D FLC

Let $x_{z}=\left(x_{1}(z), \ldots, x_{s}(z)\right)$ be a spatial input vector. Then, the 3D MF of the $i$ th spatial input $x_{i}(z)$ is given as

$$
\begin{equation*}
\mu_{i}=\beta_{i}^{l}\left(x_{i}(z)\right)=\beta_{i}\left(x_{i}(z)-\tau_{i}^{l}\right) \tag{A.1}
\end{equation*}
$$

and the Gaussian type 2D MF of the $i$ th spatial input $x_{i}(z)$ at the sensing location $z=z_{j}$ is given as

$$
\begin{equation*}
\mu_{i j}=\beta_{i}^{l}\left(x_{i}\left(z_{j}\right)\right)=\beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right) \tag{A.2}
\end{equation*}
$$

Via a 3D fuzzifier, the spatial input vector $x_{z}$ in the universe of discourse $X$ can be transformed into a spatial fuzzy input $\bar{A}_{X}$ as below:

$$
\begin{align*}
& \bar{A}_{X} \\
& =\sum_{z \in Z} \sum_{x_{1}(z) \in X_{1}} \cdots \sum_{x_{s}(z) \in X_{s}} \mu_{\bar{A}_{X}}\left(x_{1}(z), \ldots, x_{s}(z), z\right) \\
& \quad /\left(x_{1}(z), \ldots, x_{s}(z), z\right) \\
& =\sum_{z \in Z} \sum_{x_{1}(z) \in X_{1}} \cdots \sum_{x_{s}(z) \in X_{s}} \mu_{X_{1}}\left(x_{1}(z), z\right) * \cdots * \mu_{X_{s}}\left(x_{s}(z), z\right) \\
& \quad /\left(x_{1}(z), \ldots, x_{s}(z), z\right), \tag{A.3}
\end{align*}
$$

where $*$ denotes the t -norm operation.

Assume that 3D fuzzy rules are designed with the following form:

$$
\begin{align*}
& \bar{R}^{0}: \text { if } x_{1}(z) \text { is } \bar{C}_{1}^{0} \text { and } \cdots \text { and } x_{s}(z) \text { is } \bar{C}_{s}^{0} \text { then } u \text { is } b^{0}, \\
& \bar{R}^{l}: \text { if } x_{1}(z) \text { is } \bar{C}_{1}^{l} \text { and } \cdots \text { and } x_{s}(z) \text { is } \bar{C}_{s}^{l} \text { then } u \text { is } b^{l}, \tag{A.4}
\end{align*}
$$

where $\bar{C}_{i}^{0}$ is a universal 3D fuzzy set, whose MF at sensing location $z=z_{j}$ is $\beta_{i}^{0}\left(x_{i}\left(z_{j}\right)\right) \equiv 1 ; \bar{C}_{i}^{l}$ is a 3D fuzzy set, whose MF at sensing location $z=z_{j}$ is $\beta_{i}^{l}\left(x_{i}\left(z_{j}\right)\right): R \rightarrow[0,1], i=$ $1, \ldots, s ; b^{0}$ and $b^{l}$ are constants.

Then, for each fired rule, a fuzzy relation is obtained as below:

$$
\begin{equation*}
\bar{R}^{l}: \bar{C}_{1}^{l} \times \cdots \times \bar{C}_{s}^{l} \longrightarrow b^{l}, \quad l=0,1,2, \ldots, N \tag{A.5}
\end{equation*}
$$

A 3D rule inference integrates the spatial information processing and the traditional inference and contains three main operations: spatial information fusion, dimension reduction, and traditional inference. Firstly, using the spatial information fusion operation, we have a spatially distributed set $W^{l}$ over the space domain with the grade of the MF derived as

$$
\begin{align*}
& \mu_{W^{l}}(z) \\
& =\mu_{\bar{A}_{X^{\circ}}\left(\bar{C}_{1} \times \cdots \times \bar{C}_{s}^{l}\right)}\left(x_{z}, z\right) \\
& =\sup _{x_{1}(z) \in X_{1}, \ldots, x_{s}(z) \in X_{s}}\left[\mu_{\bar{A}_{X}}\left(x_{z}, z\right) * \mu_{\bar{C}_{1}^{l} \times \cdots \times \bar{C}_{s}^{l}}\left(x_{z}, z\right)\right] \\
& =\left\{\sup _{x_{1}(z) \in X_{1}}\left[\mu_{X_{1}}\left(x_{1}(z), z\right) * \mu_{\bar{C}_{1}^{l}}\left(x_{1}(z), z\right)\right]\right\} \\
& \quad * \cdots *\left\{\sup _{x_{s}(z) \in X_{s}}\right. \\
& \left.\quad\left[\mu_{X_{s}}\left(x_{s}(z), z\right) * \mu_{\bar{C}_{s}}\left(x_{s}(z), z\right)\right]\right\} \\
& =\prod_{i=1}^{s} \beta_{i}\left(x_{i}(z)-\tau_{i}^{l}\right) \tag{A.6}
\end{align*}
$$

where "product" is used for t-norm (*) and singleton fuzzifier is used.

Then, utilizing a weighted aggregation [3] dimension reduction operation, a 2D set $\chi^{l}$ is obtained. Consider the following:

$$
\begin{align*}
\mu_{\chi^{l}} & =a_{1} \mu_{W^{l}}\left(z_{1}\right)+a_{2} \mu_{W^{l}}\left(z_{2}\right)+\cdots+a_{p} \mu_{W^{l}}\left(z_{p}\right) \\
& =\sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right) \tag{A.7}
\end{align*}
$$

Finally, traditional inference operation (Mamdani implication operation) and linear defuzzifier [14] are carried out successively. We have the nonlinear mathematical expression as follows:

$$
\begin{align*}
u\left(x_{z}\right) & =b^{0}+\sum_{l=1}^{N} b^{l} \sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \mu_{\bar{C}_{i}^{l}}\left(x_{i}\left(z_{j}\right)\right) \\
& =b^{0}+\sum_{l=1}^{N} b^{l} \sum_{j=1}^{p} a_{j} \prod_{i=1}^{s} \beta_{i}\left(x_{i}\left(z_{j}\right)-\tau_{i j}^{l}\right) . \tag{A.8}
\end{align*}
$$

## B. Mathematical Preliminaries of $\varepsilon$-Support Vector Regression

In this study, we focus on $\varepsilon$-support vector regression ( $\varepsilon$ SVR). Suppose that we have a training set $D=\left\{\left[x_{i}, y_{i}\right] \in\right.$ $\left.R^{s} \times R, i=1, \ldots, q\right\}$ consisting of $q$ pairs $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)$, $\ldots,\left(x_{q}, y_{q}\right)$, where the inputs are $s$-dimensional vectors and the labels are continuous values. In $\varepsilon$-SVR, the goal is to find a function $f(x, w)$ so that for all training patterns $x$ has a maximum deviation $\varepsilon$ from the target values $y_{i}$ and has a maximum margin. The $\varepsilon$-insensitive loss function is defined as follows:

$$
|y-f(x, w)|_{\varepsilon}= \begin{cases}0 & \text { if }|y-f(x, w)| \leq \varepsilon  \tag{B.1}\\ |y-f(x, w)|-\varepsilon & \text { otherwise }\end{cases}
$$

To make the SVR nonlinear, we may map the input vector $x \in R^{s}$ into the vector $v$ of a high-dimensional feature space, $v=\Theta(x)$, where $\Theta$ represents a mapping $R^{s} \rightarrow R^{f}$, and formulate a linear regression problem in this feature space, and then an optimization problem will be solved. The optimization problem can also be solved in a dual space. By introducing the Lagrange multipliers $\alpha_{i}$ and $\alpha_{i}^{*}$, the primal optimization problem can be formulated in its dual form as follows:

$$
\begin{gather*}
\max _{\alpha_{i} \alpha_{i}^{*}}\left\{-\frac{1}{2} \sum_{i=1}^{q} \sum_{j=1}^{q}\left(\alpha_{i}^{*}-\alpha_{i}\right)\left(\alpha_{j}^{*}-\alpha_{j}\right)\left\langle\Theta\left(x_{i}\right) \cdot \Theta\left(x_{j}\right)\right\rangle\right. \\
\left.-\varepsilon \sum_{i=1}^{q}\left(\alpha_{i}^{*}+\alpha_{i}\right)+\sum_{i=1}^{q}\left(\alpha_{i}^{*}-\alpha_{i}\right) y_{i}\right\} \tag{B.2}
\end{gather*}
$$



Layer 1
Layer 2
Layer 3
Layer 1: input $x=\left(x_{1}, \ldots, x_{s}\right)$
Layer 2: support vectors $x^{1}, \ldots, x^{N}$ and KFs $K\left(x, x^{1}\right), \ldots, K\left(x, x^{N}\right)$
Layer 3: output $u=\sum_{1=1}^{N}\left(\alpha_{\mathrm{i}}^{*}-\alpha_{\mathrm{i}}\right) K\left(x, x^{i}\right)+b$
Figure 10: Three-layer network structure of an SVR.
subject to

$$
\begin{equation*}
\sum_{j=1}^{q} \alpha_{i}^{*}=\sum_{i=1}^{q} \alpha_{i}, \quad 0 \leq \alpha_{i}^{*} \leq C, 0 \leq \alpha_{i} \leq C, i=1, \ldots, q \tag{B.3}
\end{equation*}
$$

where the constant $C$ is a design parameter chosen by the user, which determines the tradeoff between the complexity of $f(x, w)$ and the approximate error.

Solving the dual quadratic programming problem, we can find an optimal weight vector $w$ and an optimal bias $b$ of the regression hypersurface given in (B.4):

$$
\begin{gather*}
w=\sum_{i=1}^{q}\left(\alpha_{i}^{*}-\alpha_{i}\right) \Theta\left(x_{i}\right) \\
b=\frac{1}{q}\left(\sum_{i=1}^{q}\left(y_{i}-\left\langle w \cdot \Theta\left(x_{i}\right)\right\rangle\right)\right) . \tag{B.4}
\end{gather*}
$$

Then, the best regression hypersurface is given by

$$
\begin{align*}
f(x, w) & =\sum_{i=1}^{q}\left(\alpha_{i}^{*}-\alpha_{i}\right)\left\langle\Theta(x) \cdot \Theta\left(x_{i}\right)\right\rangle+b  \tag{B.5}\\
& =\sum_{i \in \mathrm{SV}}\left(\alpha_{i}^{*}-\alpha_{i}\right)\left\langle\Theta(x) \cdot \Theta\left(x_{i}\right)\right\rangle+b .
\end{align*}
$$

The training pattern $x_{i}$ with nonzero $\left(\alpha_{i}^{*}-\alpha_{i}\right)$ is called support vector (SV).

To avoid a direct mapping $\Upsilon(x)$, the kernel trick is used. AKF $K\left(x_{i}, x_{j}\right)$, which satisfies the Mercer's theorem, is introduced as below:

$$
\begin{equation*}
K\left(x_{i}, x_{j}\right)=\left\langle\Theta\left(x_{i}\right) \cdot \Theta\left(x_{j}\right)\right\rangle \tag{B.6}
\end{equation*}
$$

Using $K\left(x_{i}, x_{j}\right)$, the SVR can be constructed which operates in an infinite dimensional space. Then, the solution of the SVR has the form

$$
\begin{equation*}
f(x, w)=\sum_{i \in \mathrm{SV}}\left(\alpha_{i}^{*}-\alpha_{i}\right) K\left(x, x^{i}\right)+b \tag{B.7}
\end{equation*}
$$

Let $x^{1}, x^{2}, \ldots, x^{N}$ represent support vectors. The solution of the SVR can be described by a three-layer network structure as shown in Figure 10.

## Acknowledgment

This work was supported by the project from the National Science Foundation of China under Grant no. 61273182, no. 51375293, and no. 61171145.

## References

[1] W. H. Ray, Advanced Process Control, McGraw-Hill, New York, NY, USA, 1981.
[2] H. X. Li, X. X. Zhang, and S. Y. Li, "A three-dimensional fuzzy control methodology for a class of distributed parameter systems," IEEE Transactions on Fuzzy Systems, vol. 15, no. 3, pp. 470-481, 2007.
[3] X. X. Zhang, H. X. Li, and S. Y. Li, "Analytical study and stability design of a 3-D fuzzy logic controller for spatially distributed dynamic systems," IEEE Transactions on Fuzzy Systems, vol. 16, no. 6, pp. 1613-1625, 2008.
[4] X. X. Zhang, S. Y. Li, and H. X. Li, "Structure and BIBO stability of a three-dimensional fuzzy two-term control system," Mathematics and Computers in Simulation, vol. 80, no. 10, pp. 1985-2004, 2010.
[5] X. X. Zhang, H. X. Li, and C. K. Qi, "Spatially constrained fuzzy-clustering-based sensor placement for spatiotemporal fuzzycontrol system," IEEE Transactions on Fuzzy Systems, vol. 18, no. 5, pp. 946-957, 2010.
[6] V. Kecman, Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models, The MIT Press, London, UK, 2001.
[7] Y. Jiang, X. X. Zhang, T. Zou, and G. T. Cao, "A novel 3D fuzzy logic controller design using table look-up scheme," in Proceedings of the International Conference on Computer Application and System Modeling (ICCASM '10), pp. V4-388-V4-393, Taiyuan, China, October 2010.
[8] X. X. Zhang, J. J. Li, Y. Jiang, B. L. Su, C. K. Qi, and T. Zou, "Fuzzy clustering based spatiotemporal fuzzy logic controller design," in Proceedings of the 10th World Congress on Intelligent Control and Automation, pp. 3167-3172, Beijing, China, July 2012.
[9] X. X. Zhang, Y. Jiang, T. Zou, C. K. Qi, and G. T. Cao, "Datadriven based 3-D fuzzy logic controller design using nearest neighborhood clustering and linear support vector regression," in Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ '11), pp. 374-380, Taipei, Taiwan, June 2011.
[10] X. X. Zhang, J. D. Qi, B. L. Su, S. W. Ma, and H. B. Liu, "A clustering and SVM regression learning-based spatiotemporal fuzzy logic controller with interpretable structure for spatially distributed systems," Journal of Applied Mathematics, vol. 2012, Article ID 841609, 24 pages, 2012.
[11] X. X. Zhang, Y. Jiang, H. X. Li, and S. Y. Li, "SVR learning-based spatiotemporal fuzzy logic controller for nonlinear spatially
distributed dynamic systems," IEEE Transactions on Neural Networks and Learning Systems, vol. 24, no. 10, pp. 1635-1647, 2013.
[12] D. Dubois and H. Prade, "Operations on fuzzy numbers," International Journal of Systems Science, vol. 9, no. 6, pp. 613626, 1978.
[13] Y. Chen and J. Z. Wang, "Support vector learning for fuzzy rule-based classification systems," IEEE Transactions on Fuzzy Systems, vol. 11, no. 6, pp. 716-728, 2003.
[14] H. Ying, Fuzzy Control and Modeling: Analytical Foundations and Applications, IEEE Press, New York, NY, USA, 2000.
[15] L. X. Wang, A Course in Fuzzy Systems and Control, PrenticeHall, Upper Saddle River, NJ, USA, 1997.
[16] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Statistics and Computing, vol. 14, no. 3, pp. 199-222, 2004.
[17] R. Babuska, Fuzzy Modeling for Control, Kluwer Academic, Boston, Mass, USA, 1998.
[18] P. D. Christofides and P. Daoutidis, "Nonlinear control of diffusion-convection-reaction processes," Computers and Chemical Engineering, vol. 20, no. 2, pp. S1071-S1076, 1996.

## Research Article

# Novel Observer-Based Suboptimal Digital Tracker for a Class of Time-Delay Singular Systems 

Nien-Tsu Hu, ${ }^{1}$ Ter-Feng Wu, ${ }^{2}$ Sendren Sheng-Dun Xu, ${ }^{3}$ and Hsu-Chih Huang ${ }^{2}$<br>${ }^{1}$ Integrated Logistical Support Center, Chung-Shan Institute of Science and Technology, Taoyuan 32599, Taiwan<br>${ }^{2}$ Department of Electrical Engineering, National Ilan University, Ilan 26047, Taiwan<br>${ }^{3}$ Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei 10607, Taiwan

Correspondence should be addressed to Ter-Feng Wu; tfwu@niu.edu.tw
Received 12 July 2013; Accepted 28 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 Nien-Tsu Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper presents a novel suboptimal digital tracker for a class of time-delay singular systems. First, some existing techniques are utilized to obtain an equivalent regular time-delay system, which has a direct transmission term from input to output. The equivalent regular time-delay system is important as it enables the optimal control theory to be conveniently combined with the digital redesign approach. The linear quadratic performance index, specified in the continuous-time domain, can be discretized into an equivalent decoupled discrete-time performance index using the newly developed extended delay-free model. Additionally, although the extended delay-free model is large, its advantage is the elimination of all delay terms (which included a new extended state vector), simplifying the proposed approach. As a result, the proposed approach can be applied to a class of time-delay singular systems. An illustrative example demonstrates the effectiveness of the proposed design methodology.


## 1. Introduction

The singular systems naturally arise in describing large-scale systems, and there are several examples occurring in power and interconnected systems. In general, an interconnection of state variable subsystems is conveniently described as a singular system, even though an overall state space representation may not even exist. Over the past decades, much research into singular systems has solved many complex problems concerning, for example, the stability [1-4], impulsive modes [5], controllability, observability [6], and the sufficient and necessary conditions for the impulse controllability and observability of time-varying singular systems [7-11]. However, the main purpose of such work is either to stabilize the singular system or to prove its controllability and observability. Here, the key note of this paper is about tracking the issue.

This investigation considers a time-delay system. The overwhelming majority of practical control systems are described by continuous-time settings with input, output, and state time delays. Those delays arise from inherent physical phenomena and are commonly encountered in
various engineering systems. Several authors [12-15] have studied the linear quadratic optimal analog controllers for the analog system with input and state delays. Recently, robust control and filtering for both continuous-time and discretetime nominal/uncertain systems with time delays have been thoroughly studied by Mahmoud [16]. Despite much progress in both analog control theory and digital control theory over the last few decades, effective digital control of analog plants with input and state delays (input-state delayed hybrid control systems) is still being developed [17, 18].

The objective of this paper is to develop a novel observerbased suboptimal digital tracker for a class of time-delay singular systems. The developed digital tracker can make the outputs of the digitally controlled time-delay singular system track the desired reference signals. First, the timedelay singular system is converted into a regular time-delay system that contains a direct transmission term from input to output. Then, for effective utilization of the well-developed discrete-time optimal control theory for a regular time-delay system, it is converted into a new extended discrete delayfree model. The performance cost function is discretized using the extended discrete delay-free model. When the
states of the continuous time-delay singular system are not available for measurements, a suboptimal digital observer for the original continuous time-delay singular system is constructed by using the duality of the digital redesign technique for the controller and the digital-to-analog model conversion technique [19]. As a result, the proposed novel observer-based suboptimal digital tracker is able to make the output of the digitally controlled analog time-delay system track the desired reference signals.

The rest of the paper is organized as follows. Section 2 presents the problem description and preliminary results. Section 3 presents the novel optimal tracker and a novel observer-based suboptimal tracker for the time-delay singular system and proposes a systematic design methodology for designing a set of high-performance trackers for a class of time-delay systems. Finally, an illustrative example is given to demonstrate the effectiveness of the proposed approach.

## 2. Problem Description and Preliminaries

2.1. Problem Description. Consider the following continuous time-delay singular system:

$$
\begin{gather*}
E \dot{x}_{c}(t)=A x_{c}(t)+\sum_{i=1}^{N_{1}} \widehat{A}_{i} x_{c}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} B_{j} u_{c}\left(t-\tau_{i, j}\right),  \tag{1a}\\
y_{c}(t)=C x_{c}\left(t-\tau_{o}\right), \tag{lb}
\end{gather*}
$$

where $x_{c}(t) \in \Re^{n}$ is the state vector, $u_{c}(t) \in \Re^{m}$ is the control input vector, and $y_{c}(t) \in \mathfrak{R}^{p}$ is the output vector. $E, A, \widehat{A}_{i}$, $B_{j}$, and $C$ are known constant system matrices of appropriate dimensions and $E$ is a singular matrix. The corresponding state time delay $\tau_{s, i}, i=1,2, \ldots, N_{1}$, input time delay $\tau_{i, j}$, $j=1,2, \ldots, N_{2}$, and output time delay $\tau_{o}$ are assumed to be known.

The continuous time-delay singular system (1a) and (lb) may be in impulsive modes. Directly designing the controller or observer for (la) and (lb) is very difficult because impulsive modes are uncontrollable. To solve this problem, the regular pencil, the standard pencil, and the preliminary feedback control methods are used to eliminate impulsive modes and then obtain an equivalent regular time-delay system that can be applied to the original continuous time-delay singular system (la) and (1b). The following section systematically develops the design of the novel controller and observer using the equivalent regular time-delay system.
2.2. Preliminaries. The regular pencil and standard pencil are defined below.

Definition 1 (regular pencil [20]). Let $E$ and $A$ be two square constant matrices. If $\operatorname{det}(s E-A) \neq 0$, for all $s$, then $(s E-A)$ is called a regular pencil.

Definition 2 (standard pencil [21]). Let $\left(s E_{n}-A_{n}\right)$ be a regular pencil. If there exists scalars $\alpha$ and $\beta$ such that $\alpha E_{n}+\beta A_{n}=I_{n}$, then $\left(s E_{n}-A_{n}\right)$ is called a standard pencil.

Notably, for any regular pencil, $(s E-A)$ can be easily transformed into a standard pencil by multiplying
$(\alpha E+\beta A)^{-1}$ to $E$ and $A$, respectively, where $\alpha$ and $\beta$ are scalars such that $\operatorname{det}(\alpha E+\beta A) \neq 0$. Therefore, the matrix coefficients of a standard pencil $\left(s E_{n}-A_{n}\right)$ become

$$
\begin{align*}
& E_{n}=\left(\alpha E_{r}+\beta A\right)^{-1} E,  \tag{2a}\\
& A_{n}=(\alpha E+\beta A)^{-1} A . \tag{2b}
\end{align*}
$$

The modified system retains its state vector $x_{c}(t)$ and the matrices $E_{n}$ and $A_{n}$ have the following nice properties.

Lemma 3 (see [22]). Consider
(a) $E_{n} A_{n}=A_{n} E_{n}$, meaning that $E_{n}$ and $A_{n}$ commute each other;
(b) $E_{n}$ and $A_{n}$ have the same eigenspaces.

The above properties enable a singular system to be decomposed into a reduced-order regular subsystem and a nondynamic subsystem.

## 3. Main Results

3.1. Decomposition of Time-Delay Singular System. By (2a) and (2b) the regular pencil $(s E-A)$ can be transformed into a standard pencil $\left(s E_{n}-A_{n}\right)$. Notably since $E_{n}$ is a singular matrix, which has at least one zero eigenvalue, $\beta$ cannot be equal to zero. Hence, multiplying (la) by $(\alpha E+\beta A)^{-1}$ can yield the following equation:

$$
\begin{equation*}
E_{n} \dot{x}_{c}(t)=A_{n} x_{c}(t)+\sum_{i=1}^{N_{1}} \widehat{A}_{n, i} x_{c}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} B_{n, j} u_{c}\left(t-\tau_{i, j}\right), \tag{3}
\end{equation*}
$$

where

$$
\begin{align*}
E_{n} & =(\alpha E+\beta A)^{-1} E, \\
\widehat{A}_{n, i} & =(\alpha E+\beta A)^{-1} \widehat{A}_{i},  \tag{4}\\
B_{n, j} & =(\alpha E+\beta A)^{-1} B_{j} .
\end{align*}
$$

Since $\alpha E_{n}+\beta A_{n}=I_{n}$, the pencil $\left(s E_{n}-A_{n}\right)$ is a standard one, and has the properties that are mentioned in Lemma 3. To decompose system (3), the state $x_{c}(t)$ is converted into $\bar{x}_{c}(t)$ by

$$
\begin{equation*}
x_{c}(t)=M \bar{x}_{c}(t), \tag{5}
\end{equation*}
$$

where the constant matrix $M$ is a block modal matrix of $E_{n}$ and determined by means of the extended matrix sign function [23,24]. The $M$ matrix of state space transformation is as follows.

Step 1. Find $\operatorname{sign}\left(\widetilde{E}_{n}\right)$ using the extended matrix sign function with an adequate $\omega$, where

$$
\begin{equation*}
\widetilde{E}_{n}=\left(E_{n}-\omega I_{n}\right)\left(E_{n}+\omega I_{n}\right)^{-1} \tag{6}
\end{equation*}
$$

Step 2. Find $\operatorname{sign}^{+}\left(\widetilde{E}_{n}\right)=(1 / 2)\left[I_{n}+\operatorname{sign}\left(\widetilde{E}_{n}\right)\right]$ and $\operatorname{sign}^{-}\left(\widetilde{E}_{n}\right)=$ $(1 / 2)\left[I_{n}-\operatorname{sign}\left(\widetilde{E}_{n}\right)\right]$.

Step 3. Construct the matrix $M=\left[\operatorname{ind}\left(\operatorname{sign}^{+}\left(\widetilde{E}_{n}\right)\right)\right.$ $\left.\operatorname{ind}\left(\operatorname{sign}^{-}\left(\widetilde{E}_{n}\right)\right)\right]$, where ind $(\cdot)$ represents the collection of linearly independent column vectors of $(\cdot)$.

Substituting (5) into (3) and multiplying by $M^{-1}$ on the left yield

$$
\begin{align*}
M^{-1} E_{n} M \dot{\bar{x}}_{c}(t)= & M^{-1} A_{n} M \bar{x}_{c}(t)+\sum_{i=1}^{N_{1}} M^{-1} \widehat{A}_{n, i} M \bar{x}_{c}\left(t-\tau_{s, i}\right) \\
& +\sum_{j=1}^{N_{2}} M^{-1} B_{n, j} u_{c}\left(t-\tau_{i, j}\right) \\
= & \frac{1}{\beta}\left(I_{n}-\alpha E_{n}\right) \bar{x}_{c}(t) \\
& +\sum_{i=1}^{N_{1}} M^{-1} \widehat{A}_{n, i} M \bar{x}_{c}\left(t-\tau_{s, i}\right) \\
& +\sum_{j=1}^{N_{2}} M^{-1} B_{n, j} u_{c}\left(t-\tau_{i, j}\right) . \tag{7}
\end{align*}
$$

If $M^{-1} \widehat{A}_{n, i} M$ can be diagonalized, then (7) yields,

$$
\left.\begin{array}{l}
{\left[\begin{array}{c|c}
\bar{E}_{1} & O \\
\hline & \bar{E}_{2}
\end{array}\right] \dot{\bar{x}}_{c}(t)} \\
\quad=\left[\begin{array}{c|c}
\frac{1}{\beta}\left(I_{k}-\alpha \bar{E}_{1}\right) & O \\
\hline O & \frac{1}{\beta}\left(I_{n-\kappa}-\alpha \bar{E}_{2}\right)
\end{array}\right] \bar{x}_{c}(t)  \tag{8}\\
\quad+\sum_{i=1}^{N_{1}}\left[\frac{\bar{A}_{1, i}}{\hdashline O}\right. \\
\hline \bar{A}_{2, i}
\end{array}\right] \bar{x}_{c}\left(t-\tau_{s, i}\right) .
$$

where $\bar{x}_{c}(t)=\left[\bar{x}_{s}^{T}(t), \bar{x}_{f}^{T}(t)\right]^{T}$ and $M^{-1} E_{n} M=$ block diagonal $\left\{\bar{E}_{1}, \bar{E}_{2}\right\} . \bar{E}_{1}$ is invertible with $\operatorname{rank}\left(\bar{E}_{1}\right)=\operatorname{deg}\{\operatorname{det}$ $\left.\left(s E_{r}-A\right)\right\} \triangleq k,\left[\bar{B}_{1, j}^{T}, \bar{B}_{2, j}^{T}\right]^{T}=M^{-1} B_{n, j}$, and $\bar{E}_{2}$ is a nilpotent matrix with dimension $(n-k) \times(n-k)$. Since $\operatorname{det}\left(I_{n-k}-\alpha \bar{E}_{2}\right)=$ 1 , it is invertible. Simplifying (8) by premultiplying the block diagonal $\left\{\bar{E}_{1}^{-1}, \beta\left(I_{n-k}-\alpha \bar{E}_{2}\right)^{-1}\right\}$ on both sides, one has

$$
\left.\begin{array}{rl}
{\left[\begin{array}{c|c}
I_{k} & O \\
\hline O & \bar{E}_{f}
\end{array}\right] \dot{\bar{x}}_{c}(t)=} & {\left[\begin{array}{c|c}
\bar{A}_{s} & O \\
\hline O & I_{n-\kappa}
\end{array}\right] \bar{x}_{c}(t)} \\
& +\sum_{i=1}^{N_{1}}\left[\begin{array}{c|c}
\widetilde{A}_{1, i} & O \\
\hline O & \widetilde{A}_{2, i}
\end{array}\right] \bar{x}_{c}\left(t-\tau_{s, i}\right)  \tag{9}\\
& +\sum_{j=1}^{N_{2}}\left[\bar{B}_{s, j}\right. \\
\hline \bar{B}_{f, j}
\end{array}\right] u_{c}\left(t-\tau_{i, j}\right),
$$

where

$$
\begin{gathered}
\bar{E}_{f}=\beta\left(I_{n-\kappa}-\alpha \bar{E}_{2}\right)^{-1} \bar{E}_{2}, \\
\bar{A}_{s}=\frac{1}{\beta}\left(\bar{E}_{1}^{-1}-\alpha I_{\kappa}\right),
\end{gathered}
$$

$$
\begin{gather*}
\widetilde{A}_{1, i}=\bar{E}_{1}^{-1} \bar{A}_{1, i} \\
\widetilde{A}_{2, i}=\beta\left(I_{n-\kappa}-\alpha \bar{E}_{2}\right)^{-1} \bar{A}_{2, i}, \\
\bar{B}_{s, j}=\bar{E}_{1}^{-1} \bar{B}_{1, j} \\
\bar{B}_{f, j}=\beta\left(I_{n-\kappa}-\alpha \bar{E}_{2}\right)^{-1} \bar{B}_{2, j} . \tag{10}
\end{gather*}
$$

Remarkably, since

$$
\begin{equation*}
\operatorname{rank}(E)-\operatorname{deg}\{\operatorname{det}(s E-A)\}=\operatorname{rank}\left(\bar{E}_{f}\right) \tag{11}
\end{equation*}
$$

it is much easier to determine the number of the impulsive mode using the above equation relating to (9).

For simplicity, only those singular systems that include at least one impulsive mode are discussed. First, assume that the singular system (9) has $q$; then, $\operatorname{rank}\left(\bar{E}_{f}\right)=q$. By a previously proposed method [12], the preliminary feedback gain $K_{f, j}$ is found and $K_{f, j}$ is proven to eliminate the impulsive modes. For the time-delay singular system (9), the proposed method yields a similar result (Appendix A) to that previously developed method [12] and the linear preliminary feedback control is

$$
\begin{align*}
u_{c}\left(t-\tau_{i, j}\right) & =-K_{f, j} \widehat{x}_{c, f}(t)+v_{c}\left(t-\tau_{i, j}\right) \\
& =-\left[O_{m \times k}, K_{f, j}\right] \widehat{x}_{c}(t)+v_{c}\left(t-\tau_{i, j}\right) \tag{12}
\end{align*}
$$

The time-delay singular system (9) can be transformed into (Appendix A)

$$
\begin{align*}
E_{k} \dot{\widehat{x}}_{c}(t)= & A_{k} \widehat{x}_{c}(t)+\sum_{i=1}^{N_{1}} \widehat{A}_{k, i} \widehat{x}_{c}\left(t-\tau_{s, i}\right) \\
& +\sum_{j=1}^{N_{2}} B_{k, j} v_{c}\left(t-\tau_{i, j}\right) \tag{13}
\end{align*}
$$

where

$$
\begin{align*}
& E_{k}=\left[\begin{array}{l|l}
I_{k} & O \\
\hline O & \widehat{E}_{f}
\end{array}\right] \\
& A_{k}=\left[\begin{array}{c|c}
\widehat{A}_{s} & -\sum_{j=1}^{N_{2}} \widehat{B}_{s, j} K_{f, j} \\
\hline O & I_{n-k}-\sum_{j=1}^{N_{2}} \widehat{B}_{f, j} K_{f, j}
\end{array}\right],  \tag{14}\\
& \widehat{A}_{k, i}=\left[\begin{array}{c|c}
\widetilde{A}_{1, i} & O \\
\hline O & \widetilde{A}_{2, i}
\end{array}\right] \\
& B_{k, j}=\left[\begin{array}{l}
\widehat{B}_{s, j} \\
\hline \widehat{B}_{f, j}
\end{array}\right]
\end{align*}
$$

in which

$$
\begin{gathered}
\widehat{E}_{f}=U^{-1} \bar{E}_{f} U, \\
\widehat{A}_{s}=\bar{A}_{s}, \\
\widetilde{A}_{1, i}=\bar{E}_{1}^{-1} \bar{A}_{1, i},
\end{gathered}
$$

$$
\begin{gather*}
\widetilde{A}_{2, i}=\beta\left(I_{n-\kappa}-\alpha \bar{E}_{2}\right)^{-1} \bar{A}_{2, i} \\
\widehat{B}_{s, j}=\bar{B}_{s, j} \\
\widehat{B}_{f, j}=U^{-1} \bar{B}_{f, j}, \tag{15}
\end{gather*}
$$

and $U$ is a modal matrix of $\bar{E}_{f}$ with dimension $(n-k) \times$ ( $n-k$ ) such that $U^{-1} \bar{E}_{f} U$ is in the Jordan block form. The time-delay singular system in (13) is obtained by applying the linear preliminary feedback control law $u(t)$ from (12) to the system that is given by (9). Equation (13) has the $q$ finite modes (where $\left.q=\operatorname{rank}\left(\bar{E}_{f}\right)=\operatorname{rank}\left(\widehat{E}_{f}\right)\right)$ and the $k$ original finite modes.All of these finite modes are guaranteed to be controllable. The next task is to decompose the singular system into a reduced-order regular system with $(k+q)$ controllable finite modes and the nondynamic equation with $(n-k-q)$ infinite nondynamic ones. This task can be accomplished by using previously outlined steps. First, the regular form is transformed into a standard one by multiplying (13) by $\left(\gamma E_{k}+\eta A_{k}\right)^{-1}$, where $\gamma$ and $\eta$ are arbitrary scalars such that $\left(\gamma E_{k}+\eta A_{k}\right)$ is invertible. Therefore,

$$
\begin{align*}
\left(\gamma E_{k}\right. & \left.+\eta A_{k}\right)^{-1} E_{k} \dot{\hat{x}}_{c}(t) \\
= & \left(\gamma E_{k}+\eta A_{k}\right)^{-1} A_{k} \widehat{x}_{c}(t) \\
& +\sum_{i=1}^{N_{1}}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} \widehat{A}_{k, i} \widehat{x}_{c}\left(t-\tau_{s, i}\right)  \tag{16}\\
& +\sum_{j=1}^{N_{2}}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} B_{k, j} \nu_{c}\left(t-\tau_{i, j}\right)
\end{align*}
$$

Let

$$
\begin{equation*}
\widehat{x}_{c}(t)=\widetilde{M} \tilde{x}_{c}(t), \tag{17}
\end{equation*}
$$

where the constant matrix $\widetilde{M}$ is determined by using the extended matrix sign function. The procedure is the same as that elucidated above for finding $M$, except that it operates on $\left(\gamma E_{k}+\eta A_{k}\right)^{-1} E_{k}$. Substituting (17) into (16) and multiplying by $\bar{M}^{-1}$ yield

$$
\begin{aligned}
& \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} E_{k} \widetilde{M} \dot{\tilde{x}}_{c}(t) \\
& =\widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} A_{k} \widetilde{M} \widetilde{x}_{c}(t) \\
& \quad+\sum_{i=1}^{N_{1}} \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} \widehat{A}_{k, i} \widetilde{M} \widetilde{x}_{c}\left(t-\tau_{s, i}\right) \\
& \quad+\sum_{j=1}^{N_{2}} \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} B_{k, j} v_{c}\left(t-\tau_{i, j}\right) \\
& =\widetilde{M}^{-1} \frac{1}{\eta}\left[I_{n}-\gamma\left(\gamma E_{k}+\eta A_{k}\right)^{-1} E_{k}\right] \widetilde{M} \widetilde{x}_{c}(t) \\
& \quad+\sum_{i=1}^{N_{1}} \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} \widehat{A}_{k, i} \widetilde{M} \widetilde{x}_{c}\left(t-\tau_{s, i}\right)
\end{aligned}
$$

$$
\begin{align*}
& +\sum_{j=1}^{N_{2}} \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} B_{k, j} v_{c}\left(t-\tau_{i, j}\right) \\
= & \frac{1}{\eta}\left[I_{n}-\gamma \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} E_{k} \widetilde{M}\right] \widetilde{x}_{c}(t) \\
& +\sum_{i=1}^{N_{1}} \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} \widehat{A}_{k, i} \widetilde{M} \widetilde{x}_{c}\left(t-\tau_{s, i}\right) \\
& +\sum_{j=1}^{N_{2}} \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} B_{k, j} v_{c}\left(t-\tau_{i, j}\right) . \tag{18}
\end{align*}
$$

That is,

$$
\begin{align*}
& {\left[\begin{array}{c|c}
\bar{E}_{s k} & O \\
\hline O & \bar{E}_{f k}
\end{array}\right] \dot{\tilde{x}}_{c}(t)} \\
& \quad=\left[\begin{array}{c|c}
\frac{1}{\eta}\left(I_{k+q}-\gamma \bar{E}_{s k}\right) & O \\
\hline O & \frac{1}{\eta}\left(I_{n-k-q}-\gamma \bar{E}_{f k}\right)
\end{array}\right] \tilde{x}_{c}(t)  \tag{19}\\
& \quad+\sum_{i=1}^{N_{1}}\left[\begin{array}{c|c}
\Lambda_{1, i} & O \\
\hline O & \Lambda_{2, i}
\end{array}\right] \tilde{x}_{c}\left(t-\tau_{s, i}\right) \\
& \quad+\sum_{j=1}^{N_{2}}\left[\frac{\bar{B}_{s k, j}}{\left[\bar{B}_{f k, j}\right.}\right] v_{c}\left(t-\tau_{i, j}\right)
\end{align*}
$$

where $\tilde{x}_{c}(t)=\left[\tilde{x}_{s}^{T}(t), \widetilde{x}_{f}^{T}(t)\right]^{T}, \widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} E_{k} \widetilde{M}=$ block diagonal $\left\{\bar{E}_{s k}, \bar{E}_{f k}\right\}=$ block diagonal $\left\{\bar{E}_{s k}, O_{(n-q-k)}\right\}$. $\bar{E}_{s k}$ is invertible with $\operatorname{rank}\left(\bar{E}_{s k}\right)=\operatorname{deg}\left\{\operatorname{det}\left(s E_{k}-A_{k}\right)\right\}=(q+k)$. $\bar{E}_{f k}$ is a null matrix and $\left[\bar{B}_{s k, j}^{T}, \bar{B}_{f k, j}^{T}\right]^{T}=\widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1}$ $B_{k, j}$. In (19), $\widetilde{M}^{-1}\left(\gamma E_{k}+\eta A_{k}\right)^{-1} \widehat{A}_{k, i}$ is assumed to be able to be diagonalized as block diagonal $\left\{\Lambda_{1, i}, \Lambda_{2, i}\right\}$. Then, (19) can be rewritten as

$$
\begin{align*}
& \begin{aligned}
& \dot{\tilde{x}}_{s}(t)= \frac{1}{\eta}\left(\bar{E}_{s k}^{-1}-\gamma I_{k+q}\right) \tilde{x}_{s}(t) \\
&+\sum_{i=1}^{N_{1}} \bar{E}_{s k}^{-1} \Lambda_{1, i} \widetilde{x}_{s}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} \bar{E}_{s k}^{-1} \bar{B}_{s k, j} v_{c}\left(t-\tau_{i, j}\right), \\
& 0=\tilde{x}_{f}(t)+\sum_{i=1}^{N_{1}} \eta \Lambda_{2, i} \tilde{x}_{f}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} \eta \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right),
\end{aligned},
\end{align*}
$$

and the time-delay singular system output (lb) can be rewritten as (Appendix B)

$$
\begin{aligned}
y_{c}(t) & =C x_{c}\left(t-\tau_{o}\right) \\
& =\left[\begin{array}{ll}
C_{1} & C_{2}
\end{array}\right]\left[\begin{array}{l}
\tilde{x}_{s}\left(t-\tau_{o}\right) \\
\tilde{x}_{f}\left(t-\tau_{o}\right)
\end{array}\right]
\end{aligned}
$$

$$
\begin{align*}
& =C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)+C_{2} \tilde{x}_{f}\left(t-\tau_{o}\right) \\
& =C_{1} \widetilde{x}_{s}\left(t-\tau_{o}\right)-C_{2} \sum_{j=1}^{N_{2}} \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right), \tag{21a}
\end{align*}
$$

where $C M V \widetilde{M}=\left[\begin{array}{ll}C_{1} & C_{2}\end{array}\right]$.
Finally, the time-delay singular system (1a) and (1b) can be decomposed as the equivalent regular time-delay system as follows:

$$
\begin{gather*}
\dot{\tilde{x}}_{s}(t)=A_{s} \widetilde{x}_{s}(t)+\sum_{i=1}^{N_{1}} \widehat{A}_{d, i} \tilde{x}_{s}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} B_{d, j} v_{c}\left(t-\tau_{i, j}\right),  \tag{22a}\\
y_{c}(t)=C_{1} \widetilde{x}_{s}\left(t-\tau_{o}\right)-\sum_{j=1}^{N_{2}} D_{j} v_{c}\left(t-\tau_{i, j}\right), \tag{22b}
\end{gather*}
$$

where

$$
\begin{gather*}
A_{s}=\frac{1}{\eta}\left(\bar{E}_{s k}^{-1}-\gamma I_{k+q}\right), \\
\widehat{A}_{d, i}=\bar{E}_{s k}^{-1} \Lambda_{1, i}  \tag{23}\\
B_{d, j}=\bar{E}_{s k}^{-1} \bar{B}_{s k, j} \\
D_{j}=C_{2} \bar{B}_{f k, j}
\end{gather*}
$$

Following the transformation, the time-delay singular system (la) and (lb) can be converted into a regular system (22a) and (22b) that contains a direct transmission term from input to output and the impulsive mode can be eliminated by means of the method [12]. In the next section, (22a) and (22b) will be used to develop the new optimal tracker and observer for a time-delay singular system (la) and (lb) with a series of time-delays. The proposed approaches are more general and applicable to actual systems.

### 3.2. Based on Digital Redesign and Optimal Control to

Discretize the Continuous Time-Delay Singular System and Construct the Performance Index
3.2.1. Discretization of Continuous Time-Delay Singular System. Consider the continuous time-delay singular system (22a) and (22b). To discretize (22a) and (22b), assume that $v_{c}(t)$ is a piecewise constant input function:

$$
\begin{equation*}
v_{c}(t)=v_{d}(k T), \quad k T \leq t<(k+1) T, \tag{24}
\end{equation*}
$$

where $T$ is the sampling period. Let the state delay time be given by $\tau_{s, i}=\rho_{i} T+\Gamma_{i}$, where $0 \leq \Gamma_{i}<T$ and $\rho_{i} \geq 0$ is an integer, and let the input delay time be given by $\tau_{i, j}=\eta_{j} T+$ $\sigma_{j}$, where $0 \leq \sigma_{j}<T$ and $\eta_{j} \geq 0$ is an integer. The timedelay singular system (22a) and (22b), by both the Newton extrapolation method and the Chebyshev quadrature method [25, 26], becomes

$$
\begin{aligned}
& \widetilde{x}_{d s}((k+1) T) \\
& \quad=G \widetilde{x}_{d s}(k T)+\sum_{i=1}^{N_{1}}\left[\widehat{G}_{i}^{(1)} \tilde{x}_{d s}\left(k T-\rho_{i} T+T\right)\right.
\end{aligned}
$$

$$
\begin{gather*}
+\widehat{G}_{i}^{(2)} \widetilde{x}_{d s}\left(k T-\rho_{i} T\right) \\
\left.+\widehat{G}_{i}^{(3)} \widetilde{x}_{d s}\left(k T-\rho_{i} T-T\right)\right] \\
+\sum_{j=1}^{N_{2}}\left[H_{j}^{(0)} v_{d}\left(k T-\eta_{j} T\right)+H_{j}^{(1)} v_{d}\left(k T-\eta_{j} T-T\right)\right] \tag{25}
\end{gather*}
$$

where

$$
\begin{gather*}
G=e^{A_{s} T}, \\
\widehat{G}_{i}^{(1)}=\frac{T}{2}\left[Q_{i}^{(2)}+Q_{i}^{(3)}\right] \widehat{A}_{d, i}, \\
\widehat{G}_{i}^{(2)}=T\left[Q_{i}^{(1)}-Q_{i}^{(3)}\right] \widehat{A}_{d, i}, \\
\widehat{G}_{i}^{(3)}=\frac{T}{2}\left[Q_{i}^{(3)}-Q_{i}^{(2)}\right] \widehat{A}_{d, i},  \tag{26}\\
H_{j}^{(0)}=\left[G^{1-\gamma_{j}}-I_{n}\right] A_{s}^{-1} B_{d, j}, \\
H_{j}^{(1)}=\left[G-G^{1-\gamma_{j}}\right] A_{s}^{-1} B_{d, j},
\end{gather*}
$$

in which

$$
\begin{gather*}
\gamma_{j}=\frac{\sigma_{j}}{T}, \quad \beta_{i}=\frac{\Gamma_{i}}{T}, \\
Q_{i}^{(1)}=\left[G-I_{n}\right]\left(A_{s} T\right)^{-1},  \tag{27}\\
Q_{i}^{(2)}=\left[Q_{i}^{(1)}-\left(1-\beta_{i}\right) I_{n}-\beta_{i} G\right]\left(A_{s} T\right)^{-1}, \\
Q_{i}^{(3)}=\left[2 Q_{i}^{(2)}-\left(1-\beta_{i}\right)^{2} I_{n}-\beta_{i}^{2} G\right]\left(A_{s} T\right)^{-1} .
\end{gather*}
$$

Some terms in (25) may be combined because of the same delay, so (25) can be reduced to

$$
\begin{align*}
\tilde{x}_{d s}((k+1) T)= & \bar{G} \tilde{x}_{d s}(k T)+\sum_{i=1}^{M_{1}} \bar{G}_{i} \tilde{x}_{d s}(k T-i T) \\
& +\bar{H} v_{d}(k T)+\sum_{j=1}^{M_{2}} \bar{H}_{j} v_{d}(k T-j T) . \tag{28}
\end{align*}
$$

For the output (22b), the time-delay state $x_{c}\left(t-\tau_{o}\right)$ for $k T \leq$ $t-\tau_{o}<(k+1) T$ must be evaluated. System (22a) and (22b) can be rewritten as

$$
\begin{aligned}
\tilde{x}_{s}\left(t-\tau_{o}\right)= & e^{A_{s}\left(t-\tau_{o}-k T\right)} \tilde{x}_{d s}(k T) \\
& +\sum_{i=1}^{N_{1}} \int_{k T}^{t-\tau_{o}} e^{A_{s}\left(t-\tau_{o}-\lambda\right)} \widehat{A}_{d, i} \tilde{x}_{d s}\left(\lambda-\tau_{s, i}\right) d \lambda \\
& +\sum_{j=1}^{N_{2}} \int_{k T}^{t-\tau_{o}} e^{A_{s}\left(t-\tau_{o}-\lambda\right)} B_{d, j} v_{d}\left(\lambda-\tau_{i, j}\right) d \lambda \\
= & \delta_{1}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T) \\
& +\sum_{i=1}^{N_{1}}\left[\widehat{\delta}_{i}^{(1)}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}\left(k T-\rho_{i} T+T\right)\right.
\end{aligned}
$$

$$
\begin{align*}
&+\widehat{\delta}_{i}^{(2)}\left(t-\tau_{o}-k T\right) \widetilde{x}_{d s}\left(k T-\rho_{i} T\right) \\
&+\widehat{\delta}_{i}^{(3)}\left(t-\tau_{o}-k T\right) \\
&\left.\times \widetilde{x}_{d s}\left(k T-\rho_{i} T-T\right)\right] \\
&+\sum_{j=1}^{N_{2}}\left[\varphi_{j}^{(0)}\left(t-\tau_{o}-k T\right) v_{d}\left(k T-\eta_{j} T\right)\right. \\
&+\varphi_{j}^{(1)}\left(t-\tau_{o}-k T\right) \\
&\left.\times v_{d}\left(k T-\eta_{j} T-T\right)\right] \tag{29}
\end{align*}
$$

where

$$
\begin{gather*}
\delta_{1}\left(t-\tau_{o}-k T\right)=e^{A_{s}\left(t-\tau_{o}-k T\right)}, \\
\widehat{\delta}_{i}^{(1)}\left(t-\tau_{o}-k T\right)=\frac{T}{2}\left[q_{i}^{(2)}+q_{i}^{(3)}\right] \widehat{A}_{d, i}, \\
\widehat{\delta}_{i}^{(2)}\left(t-\tau_{o}-k T\right)=T\left[q_{i}^{(1)}-q_{i}^{(3)}\right] \widehat{A}_{d, i}, \\
\widehat{\delta}_{i}^{(3)}\left(t-\tau_{o}-k T\right)=\frac{T}{2}\left[q_{i}^{(3)}-q_{i}^{(2)}\right] \widehat{A}_{d, i}, \\
\varphi_{j}^{(0)}\left(t-\tau_{o}-k T\right) \\
= \begin{cases}O_{n \times m}, & t-\tau_{o}<\sigma_{j}, \\
{\left[e^{-A_{s} \sigma_{j}} \delta_{1}\left(t-\tau_{o}-k T\right)-I_{n}\right] A_{s}^{-1} B_{d, j},} & t-\tau_{o} \geq \sigma_{j},\end{cases} \\
\varphi_{j}^{(1)}\left(t-\tau_{o}-k T\right) \\
= \begin{cases}{\left[\delta_{1}\left(t-\tau_{o}-k T\right)-I_{n}\right] A_{s}^{-1} B_{d, j},} & t-\tau_{o}<\sigma_{j} \\
\delta_{1}\left(t-\tau_{o}-k T\right)\left[I_{n}-e^{-A_{s} \sigma_{j}}\right] A_{s}^{-1} B_{d, j}, & t-\tau_{o} \geq \sigma_{j},\end{cases} \tag{30}
\end{gather*}
$$

in which

$$
\begin{align*}
q_{i}^{(1)}= & {\left[\delta_{1}\left(t-\tau_{o}-k T\right)-I_{n}\right]\left(A_{s} T\right)^{-1}, } \\
q_{i}^{(2)}= & {\left[q_{i}^{(1)}-\left(\frac{t-\tau_{o}-k T}{T}-\beta_{i}\right) I_{n}\right.} \\
& \left.-\beta_{i} \delta_{1}\left(t-\tau_{o}-k T\right)\right]\left(A_{s} T\right)^{-1},  \tag{31}\\
q_{i}^{(3)}= & {\left[2 q_{i}^{(2)}-\left(\frac{t-\tau_{o}-k T}{T}-\beta_{i}\right)^{2} I_{n}\right.} \\
& \left.+\beta_{i}^{2} \delta_{1}\left(t-\tau_{o}-k T\right)\right]\left(A_{s} T\right)^{-1} .
\end{align*}
$$

Also, some terms in (29) may be combined as in (28), and (29) may be rewritten as

$$
\begin{aligned}
\tilde{x}_{s}\left(t-\tau_{o}\right)= & \bar{\delta}_{0}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T) \\
& +\sum_{i=1}^{M_{1}} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T-i T)
\end{aligned}
$$

$$
\begin{gather*}
+\bar{\varphi}_{0}\left(t-\tau_{o}-k T\right) v_{d}(k T) \\
+\sum_{j=1}^{M_{2}} \bar{\varphi}_{j}\left(t-\tau_{o}-k T\right) v_{d}(k T-j T) . \tag{32}
\end{gather*}
$$

Then, the output (22b) can be rewritten as

$$
\begin{align*}
y_{c}(t)= & C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-\sum_{j=1}^{N_{2}} D_{j} v_{c}\left(t-\tau_{i, j}\right) \\
= & C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T) \\
& +\sum_{i=1}^{M_{1}} C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T-i T) \\
& +C_{1} \bar{\varphi}_{0}\left(t-\tau_{o}-k T\right) v_{d}(k T)  \tag{33}\\
& +\sum_{j=1}^{M_{2}} C_{1} \bar{\varphi}_{j}\left(t-\tau_{o}-k T\right) v_{d}(k T-j T) \\
& -\sum_{j=1}^{N_{2}}\left[D_{j}^{(0)} v_{d}\left(k T-\eta_{j} T\right)\right. \\
& \left.\quad+D_{j}^{(1)} v_{d}\left(k T-\eta_{j} T-T\right)\right]
\end{align*}
$$

where

$$
\begin{gather*}
D_{j}^{(0)}=D_{j}^{*}\left(B_{d, j}^{T} B_{d, j}\right)^{-1} H_{j}^{(0)}, \\
D_{j}^{(1)}=D_{j}^{*}\left(B_{d, j}^{T} B_{d, j}\right)^{-1} H_{j}^{(1)},  \tag{34}\\
D_{j}^{*}=\left[\begin{array}{ll}
D_{j} & O
\end{array}\right]^{T} .
\end{gather*}
$$

Similarly, some terms in (33) can be combined, so (33) can be rewritten as

$$
\begin{align*}
y_{c}(t)= & C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T) \\
& +\sum_{i=1}^{M_{1}} C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T-i T) \\
& +C_{1} \bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right) v_{d}(k T)  \tag{35}\\
& +\sum_{j=1}^{M_{2}} C_{1} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right) v_{d}(k T-j T) .
\end{align*}
$$

Thus, the discretization of continuous time-delay singular system (22a) and (22b) is carried out using (28) and (35).
3.2.2. Establishing Performance Index for Discrete Time-Delay Singular System. The optimal state-feedback control law minimizes the following performance cost function:

$$
J=\int_{0}^{t_{f}}\left\{\left[C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-\sum_{j=1}^{N_{2}} D_{j} v_{c}\left(t-\tau_{i, j}\right)-r(t)\right]^{T}\right.
$$

$$
\begin{align*}
& \times Q\left[C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-\sum_{j=1}^{N_{2}} D_{j} v_{c}\left(t-\tau_{i, j}\right)-r(t)\right] \\
& \left.+v_{c}^{T}(t) R v_{c}(t)\right\} \tag{36}
\end{align*}
$$

where $Q$ is the positive semidefinite matrix, $R$ is the positive definite matrix, $r(t) \in \mathfrak{R}^{q}$ is the reference input vector, and the final time $t_{f}<\infty$. To discretize the cost function $J$, given by (36), $t_{f}=N T$ is chosen and $J$ can be rewritten as

$$
J=\sum_{k=0}^{N-1} \int_{k T}^{(k+1) T}\left\{\left[C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)\right.\right.
$$

$$
\begin{align*}
& \left.-\sum_{j=1}^{N_{2}} D_{j} v_{c}\left(t-\tau_{i, j}\right)-r(t)\right]^{T} \\
& \times Q\left[C_{1} \widetilde{x}_{s}\left(t-\tau_{o}\right)\right. \\
& \left.\quad-\sum_{j=1}^{N_{2}} D_{j} v_{c}\left(t-\tau_{i, j}\right)-r(t)\right] \\
& \left.+v_{c}^{T}(t) R v_{c}(t)\right\} \tag{37}
\end{align*}
$$

Let $r^{*}(k T) \in \Re^{q}$ be the piecewise-constant reference input vector to be determined in terms of $r(k T)$ for the tracking purpose. Then, cost function (37) can be rewritten as [27]

$$
\begin{align*}
& J=\sum_{k=0}^{N-1} \int_{k T}^{(k+1) T}\left\{\left[\bar{\delta}_{0}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T)+\sum_{i=1}^{M_{1}} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T-i T)+\bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right) v_{d}(k T)\right.\right. \\
& \left.+\sum_{j=1}^{M_{2}} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right) v_{d}(k T-j T)-r^{*}(k T)\right]^{T} \\
& \times Q\left[\bar{\delta}_{0}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T)+\sum_{i=1}^{M_{1}} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right) \tilde{x}_{d s}(k T-i T)+\bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right) v_{d}(k T)\right. \\
& \left.\left.+\sum_{j=1}^{M_{2}} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right) v_{d}(k T-j T)-r^{*}(k T)\right]+v_{d}^{T}(k T) R v_{d}(k T) d t\right\} \\
& =\cdots \\
& =\sum_{k=0}^{N-1}\left[\tilde{x}_{d s}^{T}(k T) \tilde{x}_{d s}^{T}(k T-T) \cdots \tilde{x}_{d s}^{T}\left(k T-M_{1} T\right) v_{d}^{T}(k T-T) \cdots \cdots v_{d}^{T}\left(k T-M_{2} T\right) r^{* T}(k T) v_{d}^{T}(k T)\right]  \tag{38}\\
& \times\left[\begin{array}{ccccccccc}
Q_{1} & Q_{31} & \cdots & Q_{3 M_{1}} & M_{21} & \cdots & M_{2 M_{2}} & -M_{3} & M_{1} \\
Q_{31}^{T} & Q_{211} & \cdots & Q_{21 M_{1}} & M_{511} & \cdots & M_{51 M_{2}} & -M_{61} & M_{41} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
Q_{3 M_{1}}^{T} & Q_{2 M_{1} 1} & \cdots & Q_{2 M_{1} M_{1}} & M_{5 M_{1} 1} & \cdots & M_{5 M_{1} M_{2}} & -M_{6 M_{1}} & M_{4 M_{1}} \\
M_{21}^{T} & M_{511}^{T} & \cdots & M_{5 M_{1} 1}^{T} & R_{211} & \cdots & R_{21 M_{2}} & -M_{81} & R_{41} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\
M_{2 M_{2}}^{T} & M_{51 M_{2}}^{T} & \cdots & M_{5 M_{1} M_{2}}^{T} & R_{2 M_{2} 1} & \cdots & R_{2 M_{2} M_{2}} & -M_{8 M_{2}} & R_{4 M_{2}} \\
-M_{3}^{T} & -M_{61}^{T} & \cdots & -M_{6 M_{1}}^{T} & -M_{81}^{T} & \cdots & -M_{8 M_{2}}^{T} & R_{3} & -M_{7} \\
M_{1}^{T} & M_{41}^{T} & \cdots & M_{4 M_{1}}^{T} & R_{41}^{T} & \cdots & R_{4 M_{2}}^{T} & -M_{7}^{T} & R_{1}
\end{array}\right] \\
& \times\left[\begin{array}{lllllll}
\tilde{x}_{d s}(k T) & \tilde{x}_{d s}(k T-T) & \cdots & \tilde{x}_{d s}\left(k T-M_{1} T\right) & v_{d}(k T-T) & \cdots & v_{d}\left(k T-M_{2} T\right)
\end{array} r^{*}(k T) v_{d}(k T)\right]^{T},
\end{align*}
$$

where

$$
\begin{align*}
& Q_{1}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& Q_{2 i j}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\delta}_{j}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& Q_{3 i}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& M_{1}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& M_{2 j}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& M_{3}=\int_{k T}^{(k+1) T} {\left[\left(C_{1} \bar{\delta}_{0}\left(t-\tau_{o}-k T\right)\right)^{T} Q\right] d t, } \\
& M_{4 i}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right)\right)^{T}\right.  \tag{39}\\
&\left.\times Q\left(C_{1} \bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t,
\end{align*}
$$

$$
X_{d}(k T)=\left[\begin{array}{lllllll}
\tilde{x}_{d s}(k T) & \tilde{x}_{d s}(k T-T) & \cdots & \tilde{x}_{d s}\left(k T-M_{1} T\right) & v_{d}(k T-T) & \cdots & v_{d}\left(k T-M_{2} T\right) \tag{40}
\end{array} r^{*}(k T)\right]^{T}
$$

The extended delay-free system that is equivalent to the original time-delay singular system (28) and (35) is obtained as

$$
\begin{gather*}
X_{d}((k+1) T)=\widehat{G}_{e} X_{d}(k T)+\widehat{H}_{e} v_{d}(k T),  \tag{41a}\\
y_{d}(k T)=\widehat{C}_{e} X_{d}(k T) . \tag{41b}
\end{gather*}
$$

We assume that the reference input $r(t)$ is a step function with a constant magnitude, $r^{*}((k+1) T)=r^{*}(k T)$. Designing a system based on such a reference input can lead to predictable time-response characteristics. Although our design methodology is based on a step function, it should be pointed out that the resulting control system, if properly designed, enables to give good time responses for any arbitrary reference input $r(t)$. Also, the reference input $r^{*}(k T)$ is entered in the last row of $X_{d}(k T)$ at the beginning of step $k$. As a result, the extended new system does not have any time-delay terms and it can be utilized to simplify the representation of the cost function (38). Now, (38) can be rewritten as

$$
\begin{aligned}
& M_{5 i j}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& M_{6 i}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\delta}_{i}\left(t-\tau_{o}-k T\right)\right)^{T} Q\right] d t, \\
& M_{7}=\int_{k T}^{(k+1) T}\left[Q\left(C_{1} \bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& M_{8 j}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right)\right)^{T} Q\right] d t, \\
& R_{1}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\varphi}_{0}^{*}\left(t-\tau_{o}-k T\right)\right)+R\right] d t, \\
& R_{2 i j}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\varphi}_{i}^{*}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\times Q\left(C_{1} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t, \\
& R_{3}=\int_{k T}^{(k+1) T} Q_{2} d t, \\
& R_{4 j}=\int_{k T}^{(k+1) T}\left[\left(C_{1} \bar{\varphi}_{i}^{*}\left(t-\tau_{o}-k T\right)\right)^{T}\right. \\
&\left.\quad \times Q\left(C_{1} \bar{\varphi}_{j}^{*}\left(t-\tau_{o}-k T\right)\right)\right] d t .
\end{aligned}
$$

Construct an extended virtual state vector:

Then, define a new virtual weighting matrix

$$
\begin{equation*}
\widetilde{Q}=\widehat{Q}-\widehat{M} \widehat{R}^{-1} \widehat{M}^{T} \tag{44}
\end{equation*}
$$

and a new virtual control input

$$
\begin{equation*}
S(k T)=\widehat{R}^{-1} \widehat{M}^{T} X_{d}(k T)+v_{d}(k T) . \tag{45}
\end{equation*}
$$

Substituting (44) and (45) into (42) results in a decoupled performance index:

$$
\begin{equation*}
J=\sum_{k=0}^{N-1}\left[X_{d}^{T}(k T) \widetilde{\mathrm{Q}} X_{d}(k T)+S^{T}(k T) \widehat{R} S(k T)\right] . \tag{46}
\end{equation*}
$$

Substituting (45) into the extended delay-free singular system (41a) and (41b) yields

$$
\begin{align*}
X_{d} & ((k+1) T) \\
& =\widehat{G}_{e} X_{d}(k T)+\widehat{H}_{e} v_{d}(k T) \\
& =\widehat{G}_{e} X_{d}(k T)+\widehat{H}_{e}\left[S(k T)-\widehat{R}^{-1} \widehat{M}^{T} X_{d}(k T)\right]  \tag{47}\\
& =\widetilde{G}_{e} X_{d}(k T)+\widehat{H}_{e} S(k T),
\end{align*}
$$

where $\widetilde{G}_{e}=\widehat{G}_{e}-\widehat{H}_{e} \widehat{R}^{-1} \widehat{M}^{T}$.
Notably, the quadratic optimal control of the system that is given by (41a) and (41b) with the performance index that is given by (42) is equivalent to the quadratic optimal control of the system that is given by (47) with the performance index that is given by (46). The development of the desired optimal virtual control vector $S(k T)$ that minimizes the performance index that is given by (46) can be described as follows.

### 3.3. Development of Optimal Tracker for Time-Delay Singular

 System with States Available. Let the Hamilton function depend on the cost function (46) [28]:$$
\begin{align*}
H_{f}(k T)= & {\left[X_{d}^{T}(k T) \widetilde{Q} X_{d}(k T)+S^{T}(k T) \widehat{R} S(k T)\right] } \\
& +\lambda^{T}((k+1) T)\left[\widetilde{G}_{e} X_{d}(k T)+\widehat{H}_{e} S(k T)\right] \tag{48}
\end{align*}
$$

where $\lambda(k T)$ is a costate (Lagrange multiplier). Based on the well-developed optimal control theory [29, 30], the state equation is

$$
\begin{equation*}
X_{d}((k+1) T)=\widetilde{G}_{e} X_{d}(k T)+\widehat{H}_{e} S(k T), \tag{49}
\end{equation*}
$$

and the costate equation is

$$
\begin{equation*}
\lambda(k T)=\widetilde{G}_{e}^{T} \lambda((k+1) T)+\widetilde{\mathrm{Q}} X_{d}(k T) \tag{50}
\end{equation*}
$$

with the stationary condition

$$
\begin{equation*}
0=\widehat{H}_{e}^{T} \lambda((k+1) T)+\widehat{R} S(k T) \tag{51}
\end{equation*}
$$

or

$$
\begin{equation*}
S(k T)=-\widehat{R}^{-1} \widehat{H}_{e}^{T} \lambda((k+1) T), \tag{52a}
\end{equation*}
$$

and the boundary condition is

$$
\begin{equation*}
\lambda(N T)=\widetilde{Q} X_{d}(N T) \tag{52b}
\end{equation*}
$$

Assume that $\lambda(k T)$ can be written as follows:

$$
\begin{equation*}
\lambda(k T)=P(k T) X_{d}(k T) \tag{53}
\end{equation*}
$$

where $P(k T)$ is a real symmetric matrix of appropriate dimension. So far, the original optimal tracking problem has been transformed into an optimal regulator problem.

To derive the optimal regulator, (53) is substituted into (50):

$$
\begin{align*}
& P(k T) X_{d}(k T) \\
& \quad=\widetilde{G}_{e}^{T} P((k+1) T) X_{d}((k+1) T)+\widetilde{Q} X_{d}(k T), \tag{54}
\end{align*}
$$

and (52a), (52b), and (53) are substituted into (49):

$$
\begin{align*}
X_{d}((k+1) T)= & \widetilde{G}_{e} X_{d}(k T)+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} \lambda((k+1) T) \\
= & \widetilde{G}_{e} X_{d}(k T)+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T)  \tag{55}\\
& \times X_{d}((k+1) T)
\end{align*}
$$

or

$$
\begin{equation*}
X_{d}((k+1) T)=\left[I+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T)\right]^{-1} \widetilde{G}_{e} X_{d}(k T) \tag{56}
\end{equation*}
$$

Also, substituting (56) into (54) yields

$$
\begin{align*}
P(k T) X_{d}(k T)= & \widetilde{G}_{e}^{T} P((k+1) T) \\
& \times\left[I+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T)\right]^{-1}  \tag{57}\\
& \times \widetilde{G}_{e} X_{d}(k T)+\widetilde{Q} X_{d}(k T)
\end{align*}
$$

or

$$
\begin{align*}
& \left\{P(k T)-\widetilde{G}_{e}^{T} P((k+1) T)\right. \\
& \left.\quad \times\left[I+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T)\right]^{-1} \widetilde{G}_{e}-\widetilde{Q}\right\} X_{d}(k T)=0 . \tag{58}
\end{align*}
$$

The above equation must hold for all $X_{d}(k T)$. Hence,

$$
\begin{align*}
P(k T)= & \widetilde{Q}+\widetilde{G}_{e}^{T} P((k+1) T) \\
& \times\left[I+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T)\right]^{-1} \widetilde{G}_{e} \tag{59}
\end{align*}
$$

Equation (59) is called the Riccati equation. With reference to (52a), (52b), and (53), when at $k=N$,

$$
\begin{equation*}
\lambda(N T)=\widetilde{Q} X_{d}(N T)=P(N T) X_{d}(N T) \tag{60}
\end{equation*}
$$

or

$$
\begin{equation*}
P(N T)=\widetilde{Q} \tag{61}
\end{equation*}
$$

From (59) and (61), all $P(k T)$ for $0 \leq k \leq N$ can be obtained. With reference to (53) and (56), the desired optimal virtual control input that is given by (52a) now becomes

$$
\begin{align*}
S(k T)= & -\widehat{R}^{-1} \widehat{H}_{e}^{T} \lambda((k+1) T) \\
= & -\widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T) X_{d}((k+1) T) \\
= & -\widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T) \\
& \times\left[I+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T} P((k+1) T)\right]^{-1} \widetilde{G}_{e} X_{d}(k T)  \tag{62}\\
= & -\widehat{R}^{-1} \widehat{H}_{e}^{T}\left[P^{-1}((k+1) T)+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T}\right]^{-1} \\
& \times \widetilde{G}_{e} X_{d}(k T) \\
= & -K(k T) X_{d}(k T),
\end{align*}
$$

where $K(k T)=\widehat{R}^{-1} \widehat{H}_{e}^{T}\left[P^{-1}((k+1) T)+\widehat{H}_{e} \widehat{R}^{-1} \widehat{H}_{e}^{T}\right]^{-1} \widetilde{G}_{e}$. From (45), the original optimal controller $v_{d}(k T)$ is obtained as follows:

$$
\begin{equation*}
v_{d}(k T)=S(k T)-\widehat{R}^{-1} \widehat{M}^{T} X_{d}(k T)=-\widehat{K}(k T) X_{d}(k T), \tag{63}
\end{equation*}
$$

where $\widehat{K}(k T)=K(k T)+\widehat{R}^{-1} \widehat{M}^{T}$. Notice that if there are no state and input time delays, the above development can be reduced to that in [30]. Equation (63) can be represented in the following form:

$$
\begin{align*}
v_{d}( & k T) \\
= & -\widehat{K}(k T) X_{d}(k T) \\
= & -K_{d}(k T) \\
& \times\left[\begin{array}{llll}
\widetilde{x}_{d s}^{T}(k T) & \widetilde{x}_{d s}^{T}(k T-T) & \cdots & \widetilde{x}_{d s}^{T}\left(k T-M_{1} T\right)
\end{array}\right]^{T} \\
& -F_{d}(k T)\left[\begin{array}{llll}
v_{d}^{T}(k T-T) & \cdots & v_{d}^{T}\left(k T-M_{2} T\right)
\end{array}\right]^{T} \\
& -E_{d}(k T) r^{*}(k T), \tag{64}
\end{align*}
$$

where $\widehat{K}(k T)=\left[\begin{array}{lll}K_{d}(k T) & F_{d}(k T) & E_{d}(k T)\end{array}\right]$, in which

$$
\begin{gather*}
K_{d}(k T)=\left[\begin{array}{lllll}
K_{d}^{(0)}(k T) & K_{d}^{(1)}(k T) & \cdots & K_{d}^{\left(M_{1}\right)}(k T)
\end{array}\right], \\
F_{d}(k T)=\left[\begin{array}{llll}
F_{d}^{(1)}(k T) & \cdots & F_{d}^{\left(M_{2}\right)}(k T)
\end{array}\right] \tag{65}
\end{gather*}
$$

Here, the discrete optimal controller (64) for the continuous time-delay system (22a) and (22b) has been completely derived. Figure 1 presents the realization of the time-varying piecewise-constant optimal controller (64) for the digitally controlled continuous time-delay singular system.

### 3.4. Development of Observer-Based Suboptimal Tracker for

 Time-Delay Singular System with States Unavailable. When the states of a continuous time-delay system (22a) and (22b) are not available for measurement, the continuous-time statescan be constructed using the recently developed continuous time-delay observers [28, 31, 32]. However, the developed digital tracker (64) requires the extended discrete-time state $\mathrm{X}_{d}(k T)$ in (41a) and (41b). Using the prediction-based digital redesign [27], a new observer-based suboptimal tracker for the time-delay singular system can be designed as follows.

According to the digitally redesigned observer [27] and controller [27], the extended digitally redesigned observer and controller can be represented as

$$
\begin{gather*}
\widehat{X}_{d}((k+1) T)=\widehat{G}_{o} \widehat{X}_{d}(k T)+\widehat{H}_{o} v_{d}(k T) \\
\quad+L_{d}\left[y_{d}(k T)-\widehat{C}_{e} \widehat{X}_{d}(k T)\right],  \tag{66a}\\
v_{d}(k T)=-\widehat{K}(k T) \widehat{X}_{d}(k T), \tag{66b}
\end{gather*}
$$

where $\widehat{\mathrm{X}}_{d}(k T) \in \mathfrak{R}^{p}$ is the estimate of the extended state $\mathrm{X}_{d}(k T) \in \Re^{p}$ in (41a) and (41b),

$$
\begin{align*}
& \widehat{G}_{o}=\widehat{G}_{e}-L_{d} \widehat{C}_{e} \widehat{G}_{e},  \tag{66c}\\
& \widehat{H}_{o}=\widehat{H}_{e}-L_{d} \widehat{C}_{e} \widehat{H}_{e} . \tag{66d}
\end{align*}
$$

Additionally, $\widehat{G}_{e}=e^{\widehat{A} T}$ and $\widehat{H}_{e}=\left[\widehat{G}_{e}-I_{p}\right] \widehat{A}^{-1} \widehat{B}$, where $\widehat{A}=(1 / T) \ln \left(\widehat{G}_{e}\right)$ and $\widehat{B}=\widehat{A}\left(\widehat{G}_{e}-I\right)^{-1} \widehat{H}_{e}$. To determine the extended observer gain $L_{d}$ in (66a), the equivalent extended continuous-time observer (41a) and (41b) and controller (64) can be represented [19], on the basis of the analog observer and controller, as

$$
\begin{gather*}
\dot{\widehat{X}}_{c}(t)=\widehat{A} \widehat{X}_{c}(t)+\widehat{B} v_{c}(t)+L_{c}\left[y_{c}(t)-\widehat{C}_{e} \widehat{X}_{c}(t)\right],  \tag{67a}\\
v_{c}(t)=-\widehat{K}_{c} \widehat{X}_{c}(t) \tag{67b}
\end{gather*}
$$

The algorithm for computing the analog system matrix $\widehat{A}$ in (67a) from the digital system matrix $\widehat{G}_{e}$ in (41a) via the geometric-series method [15] is as follows:

$$
\begin{align*}
\widehat{A} & =\frac{1}{T} \ln \left(\widehat{G}_{e}\right) \\
& =\frac{2}{T}\left\{\widehat{R}+\frac{1}{3} \widehat{R}^{3}+\cdots+\frac{1}{n} \widehat{R}^{n}\left[I_{p}-\frac{1}{(1+2 / n)} \widehat{R}^{2}\right]^{-1}\right\} \\
& \text { for }\left|\sigma\left(\widehat{R}^{2}\right)\right|<\left(1+\frac{2}{n}\right) \\
& \cong \frac{2}{T} \widehat{R}\left[I_{p}-\frac{1}{3} \widehat{R}^{2}\right]^{-1} \quad \text { for } n=1 \\
& \cong \frac{2}{T} \widehat{R}\left[I_{p}-\frac{4}{15} \widehat{R}^{2}\right]\left[I_{p}-\frac{3}{5} \widehat{R}^{2}\right]^{-1} \text { for } n=3 \\
& \cong \cdots, \tag{67c}
\end{align*}
$$

where $\widehat{R}=\left[\widehat{G}_{e}-I_{p}\right]\left[\widehat{G}_{e}+I_{p}\right]^{-1}$ and $|\sigma(\circ)|$ denotes the absolute eigenspectrum of ( $\circ$ ). Based on the dual concept of the digital redesign, the analog observer gain $L_{c}$ in (67a)


Figure 1: Digital redesign for time-delay singular system.
and the digitally redesigned observer gain $L_{d}$ in (66a) can be represented, respectively, as

$$
\begin{gather*}
L_{c}=P_{o b} \widehat{C}_{e}^{T} R^{-1}  \tag{68}\\
L_{d}=\left(\widehat{G}_{e}-I_{p}\right) \widehat{A}^{-1} L_{c}\left[I+\widehat{C}_{e}\left(\widehat{G}_{e}-I_{p}\right) \widehat{A}^{-1} L_{c}\right]^{-1} \tag{69}
\end{gather*}
$$

where $P_{o b}$ is the positive-definite and symmetric solution of the following Riccati equation:

$$
\begin{equation*}
\widehat{A} P_{o b}+P_{o b} \widehat{A}^{T}-P_{o b} \widehat{C}_{e}^{T} R^{-1} \widehat{C}_{e} P_{o b}+\widehat{C}_{e}^{T} Q \widehat{C}_{e}=0 \tag{70}
\end{equation*}
$$

in which $Q \geq 0$ and $R>0$ with appropriate dimensions.
Owing to the extended virtual state vector in (40), the matrix $\widehat{G}_{e}$ in (41a) and (41b) and (67a), (67b), and (67c) is singular. The matrices $\widehat{A}$ and $\widehat{B}$ in (67a), (67b), and (67c) cannot be directly determined. To solve this problem, an alternative is derived via the matrix sign function method $[23,24]$ as follows.

Following the procedures shown in Section 3 [23, 24], the transformed matrix is

$$
\begin{equation*}
\widehat{G}_{b}=\left(\widehat{G}_{e}-\zeta_{2} I_{p}\right)\left(\widehat{G}_{e}+\zeta_{2} I_{p}\right)^{-1} \tag{71}
\end{equation*}
$$

where $\widehat{G}_{e} \in \Re^{p \times p}$ and $\zeta_{2}$ is a radius of a circle from the origin of the coordinates. Additionally, the associated matrix sign functions are

$$
\begin{gather*}
\operatorname{Sign}\left(\widehat{G}_{b}\right)=\widehat{G}_{b}\left(\sqrt[2]{\widehat{G}_{b}^{2}}\right)^{-1},  \tag{72}\\
\operatorname{Sign}^{-}\left(\widehat{G}_{b}\right)=\frac{1}{2}\left(I_{p}-\operatorname{Sign}\left(\widehat{G}_{b}\right)\right),  \tag{73a}\\
\operatorname{Sign}^{+}\left(\widehat{G}_{b}\right)=\frac{1}{2}\left(I_{p}+\operatorname{Sign}\left(\widehat{G}_{b}\right)\right), \tag{73b}
\end{gather*}
$$

respectively. A fast and stable algorithm for computing the matrix sign function [23,24] is given as follows.

For the order of the desired convergence rate $r=2$, one has

$$
\begin{gather*}
Q(l+1)=\frac{1}{2}\left[Q(l)+Q^{-1}(l)\right] \\
Q(0)=\widehat{G}_{b}  \tag{74}\\
\lim _{l \rightarrow \infty} Q(l)=\operatorname{Sign}\left(\widehat{G}_{b}\right), \quad \text { for } l=0,1,2, \ldots
\end{gather*}
$$

By [19, 20], a transformation matrix $T_{m}$ can be found such that

$$
\widehat{G}_{m}=T_{m}^{-1} \widehat{G}_{e} T_{m}=\left[\begin{array}{c|c}
\widehat{G}_{m 1} & 0_{(p-g) \times g}  \tag{75}\\
\hline 0_{g \times(p-g)} & \widehat{G}_{m 2}
\end{array}\right]
$$

where $\widehat{G}_{m 1}$ is a nonsingular matrix and $\widehat{G}_{m 2}$ is a singular matrix whose eigenvalues are all null. Finally, the matrix $\widehat{A}$ is obtained by the following equation:

$$
\widehat{A}=T_{m} \widehat{G}_{\ln } T_{m}^{-1}=T_{m}\left[\begin{array}{c|c}
\widehat{G}_{\ln 1} & 0  \tag{76a}\\
\hline 0 & \widehat{G}_{\ln 2}
\end{array}\right] T_{m}^{-1}
$$

where $\widehat{G}_{\ln 1}=(1 / T) \ln \left(\sqrt[3]{\widehat{G}_{m 1}}\right)^{3}=(3 / T) \ln \left(\sqrt[3]{\widehat{G}_{m 1}}\right)$ and $\widehat{G}_{\ln 2}=\nu I_{g}$, in which $v$ is a large negative real constant. The algorithm for finding $\ln \left(\sqrt[3]{\widehat{G}_{m 1}}\right)$ in (76a) can be found in (67c) [19]. If the matrix $\widehat{G}_{m 1}$ has any negative real eigenvalue, then the principal third root of $\widehat{G}_{m 1}$ is not defined for $\arg \left(\lambda_{i}\right) \neq \pi$ $[23,24]$. The first part $\widehat{G}_{\ln 1}$ in (76a) can be rewritten as

$$
\begin{equation*}
\widehat{G}_{\ln 1}=\frac{3}{T} \ln \left(\sqrt[3]{\widehat{G}_{m 1}}\right)=\frac{3}{T} \ln \left[\left(\sqrt[3]{\widehat{G}_{m 1} e^{i \theta}}\right) e^{-(i \theta / 3)}\right] \tag{76b}
\end{equation*}
$$

where the matrix $\widehat{G}_{m 1}$ is rotated by a small positive real angle $\theta$. The second part $\widehat{G}_{\ln 2}$ in (76a) is utilized to recover the property $\ln (0)=-\infty$. Matrix $\widehat{B}$ can be evaluated as

$$
\begin{align*}
\widehat{B} & =\widehat{A}\left(\widehat{G}_{e}-I_{p}\right)^{-1} \widehat{H}_{e} \\
& =\left[\left(\widehat{G}_{e}-I_{p}\right) \widehat{A}^{-1}\right]^{-1} \widehat{H}_{e}  \tag{77}\\
& =\left(T I_{p}+\frac{\widehat{A} T^{2}}{2!}+\frac{\widehat{A}^{2} T^{3}}{3!}+\cdots\right)^{-1} \widehat{H}_{e} .
\end{align*}
$$

Substituting (76a) and (76b) into the Riccati equation (70) and solving it yield the observer gain matrices in (68) and (69). Figure 2 presents the implementation of the observerbased suboptimal tracker for the time-delay singular system.

## 4. An Illustrative Example

Consider a continuous time-delay singular system, described in (la) and (lb), with

$$
\begin{gather*}
E=\left[\begin{array}{llllll}
1 & 2 & 1 & 1 & -3 & -2 \\
0 & 2 & 2 & 1 & -3 & -3 \\
1 & 2 & 1 & 1 & -3 & -2 \\
1 & 2 & 1 & 3 & -5 & -4 \\
0 & 2 & 1 & 1 & -2 & -2 \\
1 & 0 & 0 & 0 & -1 & 0
\end{array}\right], \quad A=I_{6}, \\
\widehat{A}_{1}=\left[\begin{array}{ccccccc}
0.447 \\
0 & 0 & 0 & 0 & 0.447 & 0 \\
0 & 0 & 0.2236 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.447 & -0.8944 & 0 \\
0 & 0 & 0 & 0 & 0.447 & 0 \\
0 & 0 & 0 & 0 & -0.8944 & 0.447
\end{array}\right], \\
B_{1}=\left[\begin{array}{ccccc}
1 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0
\end{array}\right]^{T}, \quad C=\left[\begin{array}{ccccc}
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 \\
0
\end{array}\right], \\
N_{1}=N_{2}=1, \tag{78}
\end{gather*}
$$

Let the sampling period $T=0.01$ (s) and apply the reference input $r(t)=[0.5 \sin (t) 0.5 \cos (t)]^{T}$ to the system. The initial condition is $x_{c}(0)=(M V \widetilde{M})\left[\widetilde{x}_{s}^{T}(0) \tilde{x}_{f}^{T}(0)\right]^{T}=$ $\left[\begin{array}{llllll}0 & 0 & 0 & 0 & 0 & 0\end{array}\right]^{T}, \tilde{x}_{s}(0)=\left[\begin{array}{llll}0 & 0 & 0 & 0\end{array}\right]^{T}$, and $\tilde{x}_{f}=\left[\begin{array}{ll}0 & 0\end{array}\right]^{T}$.

Since $0 \times E+A=I_{6}$, and according to the definition of the standard form, $\{E, A\}$ is in standard form. If $\alpha=0$ and $\beta=1$ are set, then $E_{n}=E, A_{n}=A, \widehat{A}_{n, 1}=\widehat{A}_{1}$, and $B_{n, 1}=B_{1}$. Since $E_{n}$ is singular, $E_{n}$ includes some zero eigenvalues, and the bilinear transform must be performed to find the similarity transformation matrix $M$ of $E_{n}$. Assume $\omega=0.5$; the algorithm that was described in Section 3 yields,

$$
\widetilde{E}_{n}=\left[\begin{array}{cccccc}
0.3333 & 1.6 & -2.4 & 0.16 & 0.9067 & 2.24 \\
0 & 0.6 & 1.6 & 0.16 & -1.76 & -1.76 \\
1.3333 & 1.6 & -3.4 & 0.16 & 0.9067 & 2.24 \\
1.3333 & 1.6 & -2.4 & 0.76 & -0.6933 & 0.64 \\
0 & 1.6 & -2.4 & 0.16 & 1.24 & 2.24 \\
1.3333 & 0 & 0 & 0 & -1.3333 & -1
\end{array}\right]
$$

$$
\begin{align*}
& \operatorname{sign}\left(\widetilde{E}_{n}\right)=\left[\begin{array}{cccccc}
1 & 2 & 2 & 0 & -4 & -2 \\
0 & 1 & 2 & 0 & -2 & -2 \\
2 & 2 & 1 & 0 & -4 & -2 \\
2 & 2 & 2 & 1 & -6 & -4 \\
0 & 2 & 2 & 0 & -3 & -2 \\
2 & 0 & 0 & 0 & -2 & -1
\end{array}\right], \\
& M=\left[\begin{array}{cccccc}
1 & 1 & 0 & 0 & -1 & -1 \\
0 & 1 & 0 & 0 & 0 & -1 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 1 & 1 & -1 & -1 & -1 \\
0 & 1 & 0 & 0 & -1 & -1 \\
1 & 0 & 0 & -1 & 0 & 0
\end{array}\right] . \tag{79}
\end{align*}
$$

Based on Section 3.1 and Appendices, the time-delay singular system can be decomposed as follows:

$$
\begin{gather*}
\dot{\tilde{x}}_{s}(t)=A_{s} \tilde{x}_{s}(t)+\widehat{A}_{d, 1} \tilde{x}_{s}\left(t-\tau_{s, 1}\right)+B_{d, 1} v_{c}\left(t-\tau_{i, 1}\right),  \tag{80a}\\
y_{c}(t)=C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-D_{1} v_{c}\left(t-\tau_{i, 1}\right), \tag{80b}
\end{gather*}
$$

where

$$
\begin{gather*}
A_{s}=\left[\begin{array}{cccc}
1 & 0 & 0 & -0.5 \\
0 & 0.5 & -0.25 & -0.5 \\
0 & 0 & 0.5 & 0 \\
0 & 0 & 0 & 0.5
\end{array}\right], \\
\widehat{A}_{d, 1}=\left[\begin{array}{cccc}
0.4472 & 0 & 0 & -0.2236 \\
0 & 0.2236 & -0.1118 & -0.2236 \\
0 & 0 & 0.2236 & 0 \\
0 & 0 & 0 & 0.2236
\end{array}\right], \\
B_{d, 1}=\left[\begin{array}{cc}
0.5 & 0.5 \\
-0.25 & -0.25 \\
0.5 & 0.5 \\
0.5 & -0.5
\end{array}\right],  \tag{81}\\
C_{1}=\left[\begin{array}{ccc}
1 & 0 & 1 \\
1 & 0 & 0 \\
0
\end{array}\right] \\
D_{1}=\left[\begin{array}{cc}
0 & 2 \\
0.5 & 1.5
\end{array}\right]
\end{gather*}
$$

and the other parameters are listed below:

$$
\begin{gathered}
\widetilde{M}=\left[\begin{array}{llllll}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0
\end{array}\right], \\
\gamma=2, \\
\eta=-1, \\
\Lambda_{1,1}=\left[\begin{array}{ccccc}
0.4472 & 0 & 0 & -0.2982 \\
0 & 0.1491 & -0.0994 & -0.1988 \\
0 & 0 & 0.1491 & 0 \\
0 & 0 & 0 & 0.1491
\end{array}\right],
\end{gathered}
$$



Figure 2: Observer-based suboptimal tracker for the digitally redesigned time-delay singular system.


Figure 3: Output responses of time-delay singular system with states available by the new digital redesign approach.

$$
\begin{gather*}
\Lambda_{2,1}=\left[\begin{array}{cc}
-0.4472 & 0 \\
0 & -0.4472
\end{array}\right], \\
B_{s k, 1}=\left[\begin{array}{cc}
0.3333 & 0.6667 \\
-0.3333 & -0.1111 \\
0.3333 & 0.3333 \\
0.3333 & -0.3333
\end{array}\right], \\
V=I_{6}, \quad K_{f, 1}=\left[\begin{array}{ccc}
0 & 1 & 0 \\
0 & -1 & 0
\end{array}\right] . \tag{82}
\end{gather*}
$$

Following the proposed method in Section 3, the schemes of Figures 1 and 2 are implemented. For simplification, the numerical analysis is not presented and Figures 3 and 4 show the results of the simulation.

Comparing with the offline observer/Kalman filter identification (OKID) method, the advantages of the proposed approach can be shown in [33, 34]. Following [33, 34], let the unknown system (80a) and (80b) be excited by the whitenoise control force $u(t)=\left[u_{1}(t) u_{2}(t)\right]^{T}$ with a zero mean and covariance $\operatorname{diag}\left[\operatorname{cov}\left(u_{1}(t)\right), \operatorname{cov}\left(u_{2}(t)\right)\right]=\operatorname{diag}\left[\begin{array}{ll}0.2 & 0.2\end{array}\right]$. The input-output sampled data is given in Figure 5.


Figure 4: Output responses of time-delay singular system with states unavailable by new observer-based suboptimal approach.


Figure 5: System I/O data for identification.

The identified system $(\widehat{G}, \widehat{H}, \widehat{C})$ and observer gain $(F)$ matrices for the unknown system (80a) and (80b) are given as
$\widehat{G}=\left[\begin{array}{ccccc}8.7538 & -4.3746 & 8.9478 & -0.5128 & 0 \\ 10.2956 & 5.5361 & 1.4682 & -3.6730 & 0 \\ -19.5769 & 9.2613 & -19.7421 & 9.5377 & 0 \\ -0.0821 & 0.6656 & -1.0711 & -359.3919 & 0 \\ 0 & 0 & 0 & 0 & -112.9831\end{array}\right]$,

$$
\begin{gathered}
\widehat{H}=\left[\begin{array}{cc}
-0.2470 & -0.2116 \\
-0.8808 & -0.8879 \\
-0.5713 & -0.5007 \\
22.2478 & 21.6048 \\
0 & 0
\end{array}\right], \\
\widehat{C}=\left[\begin{array}{ccccc}
-1.0867 & -0.5871 & 0.7911 & -0.0579 & 0 \\
0.5676 & -0.5710 & -0.4333 & -0.0488 & 0
\end{array}\right],
\end{gathered}
$$




$$
\text { Error }=y_{s 1}(k T)-y_{\text {okid } 1}(k T)
$$

(a)


--- $y_{\text {okid2 }}(k T)$ : OKID method


[^0](b)

Figure 6: (a) The comparison between the system output $y_{s 1}(k T)$ and its observer-based output $y_{\text {okid } 1}(k T)$ by OKID. (b) The comparison between the system output $y_{s 2}(k T)$ and its observer-based output $y_{\text {okid2 }}(k T)$ by OKID.


- $y_{s 1}(k T):$ system
..- $y_{\mathrm{id} 1}(k T)$ : the proposed method


$$
- \text { Error }=y_{s 1}(k T)-y_{\mathrm{id} 1}(k T)
$$

(a)


- $y_{s 2}(k T)$ : system
..-- $y_{\mathrm{id} 2}(k T)$ : the proposed method

(b)

FIgURe 7: (a) The comparison between the system output $y_{s 1}(k T)$ and its observer-based output $y_{\mathrm{id1}}(k T)$ by the proposed method. (b) The comparison between the system output $y_{s 2}(k T)$ and its observer-based output $y_{\mathrm{id} 2}(k T)$ by the proposed method.

$$
F=\left[\begin{array}{cc}
1.4216 & -2.0615  \tag{83}\\
0.7740 & 1.0579 \\
0.5564 & -0.8225 \\
-0.0043 & -0.0041 \\
0 & 0
\end{array}\right]
$$

Then, the observer-based outputs by OKID compared with the actual system outputs for the unknown system (80a) and (80b) are shown in Figure 6.

To overcome the effect of modeling error, an improved observer (69) with the high-gain property has been proposed in this paper, where the observer gain matrices are given as

$$
\begin{align*}
L_{d}=\left[\begin{array}{cccccccc}
-0.3231 & -0.5708 & 0.3103 & -0.1051 & 0 & -0.0628 \\
0.4008 & -1.0446 & -0.4013 & -0.1646 & 0 & 0.1289 & \\
& -0.0833 & -0.1119 & 6.4407 & 0 & 1.1006 & 0.1373 & \\
& -0.1585 & -0.4047 & 10.2422 & 0 & 1.6436 & 0.3760 & \\
& -16.6049 & 228.3244 & 0.0003 & 0.0124 & 0.0119 & 0 & 0 \\
& -25.8086 & 363.6177 & 0.0001 & 0.0191 & 0.0184 & 0 & 0
\end{array}\right]^{T} .
\end{align*}
$$

Then, the comparisons between the actual outputs and the proposed method outputs for the unknown system (80a) and (80b) are shown in Figure 7.

Obviously, the proposed method is better than OKID method on the tracking error performance from Figures 6 and 7.

## 5. Conclusion

This paper presents a systematic methodology for developing novel observer-based suboptimal digital trackers for a class of time-delay singular systems. The time-delay property and singular system have been attracting more attention in recent years. The proposed controller and observer depend on the concepts of optimal control and the digital redesign with high-gain property to ensure effective tracking and favorable state matching performance. In future works, we will pay more attention to the online application and the realtime implementation of fault tolerant control system with performance optimization by using the proposed methods.

## Appendices

## A. Transformation of the Time-Delay Singular Systems

The following steps yield the preliminary feedback gain $K_{f, j}$ and prove that $K_{f, j}$ can eliminate impulsive modes.

Let

$$
\begin{equation*}
\bar{x}_{c}(t)=V \widehat{x}_{c}(t), \tag{A.1}
\end{equation*}
$$

where $\widehat{x}_{c}(t)=\left[\widehat{x}_{c, s}^{T}(t), \widehat{x}_{c, f}^{T}(t)\right]^{T}=\left[\bar{x}_{c, s}^{T}(t),\left(U^{-1} \bar{x}_{c, f}(t)\right)^{T}\right]^{T}$ and $V=\left[\begin{array}{l|l}I_{k} \mid O \\ \hline O \mid U\end{array}\right] \cdot U$ is a modal matrix of $\bar{E}_{f}$ with dimension $(n-k) \times(n-k)$ such that $U^{-1} \bar{E}_{f} U$ is in the Jordan block
form. Substituting (A.1) into (9) and multiplying by $V^{-1}$ yield the following equation:

$$
\begin{align*}
& {\left[\begin{array}{c|c}
I_{k} & O \\
\hline O & \widehat{E}_{f}
\end{array}\right] \dot{\hat{x}}_{c}(t)} \\
& =\left[\begin{array}{c|c}
\widehat{A}_{s} & O \\
\hline O & I_{n-k}
\end{array}\right] \widehat{x}_{c}(t) \\
& +\sum_{i=1}^{N_{1}}\left[\begin{array}{c|c}
\widehat{A}_{1, i} & O \\
\hline O & \widehat{A}_{2, i}
\end{array}\right] \widehat{x}_{c}\left(t-\tau_{s, i}\right)  \tag{A.2}\\
& +\sum_{j=1}^{N_{2}}\left[\frac{\widehat{B}_{s, j}}{\widehat{B}_{f, j}}\right] u_{c}\left(t-\tau_{i, j}\right),
\end{align*}
$$

where $\widehat{E}_{f}=U^{-1} \bar{E}_{f} U, \widehat{A}_{s}=\bar{A}_{s}, \widehat{A}_{1, i}=\widetilde{A}_{1, i}, \widehat{A}_{2, i}=U^{-1} \widetilde{A}_{2, i} U$, $\widehat{B}_{s, j}=\bar{B}_{s, j}$, and $\widehat{B}_{f, j}=U^{-1} \bar{B}_{f, j}$. Notably, $\widehat{E}_{f}$ is in the Jordan block form with $d$ blocks of sizes $\mu_{1}, \mu_{2}, \ldots, \mu_{d}$, where $\sum_{i=1}^{d} \mu_{i}=$ column (row) number of $\widehat{E}_{f}$. In (A.2), the statedelay $\widehat{x}_{c}\left(t-\tau_{s, i}\right)$ can be equal to $W_{s, i} \widehat{x}_{c}(t)$, where $W_{s, i}$ is a block diagonal $\left\{\psi_{1, i}, \psi_{2, i}\right\}$. Therefore, (A.2) can be rewritten as

$$
\begin{align*}
& {\left[\begin{array}{c|c}
I_{k} & O \\
\hline \mathrm{O} & \widehat{E}_{f}
\end{array}\right] \dot{\hat{x}}_{c}(t)} \\
& =\left[\begin{array}{c|c}
\widehat{A}_{s}+\sum_{i=1}^{N_{1}} \widehat{A}_{1, i} \psi_{1, i} & O \\
\hline O & I+\sum_{i=1}^{N_{1}} \widehat{A}_{2, i} \psi_{2, i}
\end{array}\right] \widehat{x}_{c}(t)  \tag{A.3}\\
& \quad+\sum_{j=1}^{N_{2}}\left[\frac{\widehat{B}_{s, j}}{\widehat{B}_{f, j}}\right] u_{c}\left(t-\tau_{i, j}\right) \\
& \triangleq \\
& =\left[\begin{array}{c|c}
A_{s 1} & O \\
\hline O & A_{s 2}
\end{array}\right] \widehat{x}_{c}(t)+\sum_{j=1}^{N_{2}}\left[\frac{\widehat{B}_{s, j}}{\widehat{B}_{f, j}}\right] u_{c}\left(t-\tau_{i, j}\right) .
\end{align*}
$$

From (A.3), the fast subsystem is

$$
\begin{equation*}
\widehat{E}_{f} \dot{\widehat{x}}_{c, f}(t)=A_{s 2} \widehat{x}_{c, f}(t)+\sum_{j=1}^{N_{2}} \widehat{B}_{f, j} u_{c}\left(t-\tau_{i, j}\right), \tag{A.4}
\end{equation*}
$$

so

$$
\begin{equation*}
\widehat{E}_{f}^{*} \dot{\hat{x}}_{c, f}(t)=\widehat{x}_{c, f}(t)+\sum_{j=1}^{N_{2}} \widehat{B}_{f, j}^{*} U_{c, j}(t), \tag{A.5}
\end{equation*}
$$

where $\widehat{E}_{f}^{*}=A_{s 2}^{-1} \widehat{E}_{f}, \widehat{B}_{f, j}^{*}=A_{s 2}^{-1} \widehat{B}_{f, j}$, and $U_{c, j}(t)=u_{c}\left(t-\tau_{i, j}\right)$.
Taking the Laplace transformation of the fast subsystem (A.5), one obtains

$$
\begin{align*}
\widehat{X}_{c, f}(s) & =\left(s \widehat{E}_{f}^{*}-I_{n-k}\right)^{-1}\left(\widehat{E}_{f}^{*} \widehat{x}_{c, f}(0)+\sum_{j=1}^{N_{2}} \widehat{B}_{f, j}^{*} U_{c, j}(s)\right) \\
& =-\sum_{i=0}^{l-1} s^{i}\left(\widehat{E}_{f}^{*}\right)^{i}\left(\widehat{E}_{f}^{*} \widehat{x}_{c, f}(0)+\sum_{j=1}^{N_{2}} \widehat{B}_{f, j}^{*} U_{c, j}(s)\right) . \tag{A.6}
\end{align*}
$$

Taking the inverse Laplace transformation of the above equation, one has

$$
\begin{align*}
\widehat{x}_{c, f}(t)= & -\sum_{i=1}^{l-1}\left(\widehat{E}_{f}^{*}\right)^{i} \widehat{x}_{c, f}(0) \delta^{(i-1)}(t) \\
& -\sum_{i=0}^{l-1} \sum_{j=1}^{N_{2}}\left(\widehat{E}_{f}^{*}\right)^{i} \widehat{B}_{f, j}^{*} U_{c, j}^{(i)}(t) \tag{A.7}
\end{align*}
$$

where $\delta(t)$ and $\delta^{(i)}(t)$ denote the delta function and the $i$ th derivative of the delta function, respectively. From the above equation, the impulsive modes of the fast state are induced from inconsistent initial conditions of the fast state or discontinuous control input (or its derivatives). By [12], determination of the preliminary feedback gain $K_{f, j}=$ $\left[k_{1, j}, k_{2, j}, \ldots, k_{n-k, j}\right]_{m \times(n-k)}$, where $k_{\xi, j}$ is of dimension $m \times 1$ for $\xi=1,2, \ldots,(n-k)$, is summarized as follows.
(1) If $\mu_{i} \geq 1$, where $1 \leq i \leq d$, and its corresponding Jordan block is a null matrix, then

$$
\begin{gather*}
k_{\mu_{1}+\mu_{2}+\cdots+\mu_{i-1}+1, j}=O_{m \times 1}, \\
k_{\mu_{1}+\mu_{2}+\cdots+\mu_{i-1}+2, j}=O_{m \times 1}, \\
\vdots  \tag{A.8}\\
k_{\mu_{1}+\mu_{2}+\cdots+\mu_{i-1}+\mu_{i}, j}=O_{m \times 1} .
\end{gather*}
$$

(2) If $\mu_{i}>1$, where $1 \leq i \leq d$, and its corresponding Jordan block is not a null matrix, then

$$
\begin{gathered}
k_{\mu_{1}+\mu_{2}+\cdots+\mu_{i-1}+1, j}=\left[\begin{array}{c}
\delta\left(\widehat{b}_{\left(\mu_{1}+\mu_{2}+\cdots+\mu_{i}\right) 1, j}\right) \\
\delta\left(\widehat{b}_{\left(\mu_{1}+\mu_{2}+\cdots+\mu_{i}\right) 2, j}\right) \\
\vdots \\
\delta\left(\widehat{b}_{\left(\mu_{1}+\mu_{2}+\cdots+\mu_{i}\right) m, j}\right)
\end{array}\right], \\
k_{\mu_{1}+\mu_{2}+\cdots+\mu_{i-1}+2, j}=O_{m \times 1} \\
\vdots \\
k_{\mu_{1}+\mu_{2}+\cdots+\mu_{i-1}+\mu_{i}, j}=O_{m \times 1}
\end{gathered}
$$

where

$$
\begin{gathered}
\widehat{B}_{f, j}^{*} \triangleq\left[\begin{array}{c}
\widehat{b}_{(k+1), j} \\
\widehat{b}_{(k+2), j} \\
\vdots \\
\widehat{b}_{n, j}
\end{array}\right]_{(n-k) \times m}, \\
\widehat{b}_{i, j} \triangleq\left[\widehat{b}_{i 1, j} \widehat{b}_{i 2, j,}, \ldots, \widehat{b}_{i m, j}\right]_{1 \times m} \\
\delta\left(\widehat{b}_{i \xi, j}\right) \triangleq \begin{cases}0, & \text { if } \widehat{b}_{i \xi, j}=0, \\
1, & \text { if } \widehat{b}_{i \xi, j}>0, \quad \xi=1,2, \ldots, m . \\
-1, & \text { if } \hat{b}_{i \xi, j}<0,\end{cases}
\end{gathered}
$$

Let

$$
\begin{align*}
U_{c, j}(t)=u_{c}\left(t-\tau_{i, j}\right) & =-K_{f, j} \widehat{x}_{c, f}(t)+V_{c, j}(t) \\
& =-K_{f, j} \widehat{x}_{c, f}(t)+v\left(t-\tau_{i, j}\right) \\
& =-\left[O_{m \times k}, K_{f, j}\right] \widehat{x}_{c}(t)+v_{c}\left(t-\tau_{i, j}\right) . \tag{A.11}
\end{align*}
$$

Substituting (A.11) into (A.2) yields (13).

## B. Output Transformation of the Time-Delay Singular Systems

Equation (20b) can be decomposed as follows:

$$
\begin{align*}
& 0=\tilde{x}_{f, s}(t)+\sum_{i=1}^{N_{1}} \Delta_{1, i} \tilde{x}_{f, s}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} \bar{B}_{s, j}^{*} v_{c}\left(t-\tau_{i, j}\right),  \tag{B.1a}\\
& 0=\tilde{x}_{f, f}(t)+\sum_{i=1}^{N_{1}} \Delta_{2, i} \tilde{x}_{f, f}\left(t-\tau_{s, i}\right)+\sum_{j=1}^{N_{2}} \bar{B}_{f, j}^{*} v_{c}\left(t-\tau_{i, j}\right), \tag{B.1b}
\end{align*}
$$

where $\tilde{x}_{f}(t)=\left[\frac{\tilde{x}_{f, s}(t)}{\bar{x}_{f, f}(t)}\right], \bar{B}_{f k, j}=\left[\begin{array}{c}\bar{B}_{s, j}^{*} \\ \bar{B}_{f, j}^{\prime}\end{array}\right]$, and $\eta \Lambda_{2, i}=$ block diagonal $\left\{\Delta_{1, i}, \Delta_{2, i}\right\}$ is assumed. Based on (B.1a) and (B.lb), the following equations hold:

$$
\begin{align*}
& \sum_{i=1}^{N_{1}+1} \Delta_{1, i} \tilde{x}_{f, s}\left(t-\tau_{s, i}\right)=-\sum_{j=1}^{N_{2}} \bar{B}_{s, j}^{*} v_{c}\left(t-\tau_{i, j}\right),  \tag{B.2a}\\
& \sum_{i=1}^{N_{1}+1} \Delta_{2, i} \tilde{x}_{f, f}\left(t-\tau_{s, i}\right)=-\sum_{j=1}^{N_{2}} \bar{B}_{f, j}^{*} v_{c}\left(t-\tau_{i, j}\right), \tag{B.2b}
\end{align*}
$$

where $i=k$ and $\Delta_{1, k}=\Delta_{1, k}=I, \tau_{s, k}=0$. Similarly, from (21a),

$$
\begin{align*}
y_{c}(t) & =C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)+C_{2} \tilde{x}_{f}\left(t-\tau_{o}\right) \\
& =C_{1} \widetilde{x}_{s}\left(t-\tau_{o}\right)+\left[\begin{array}{ll}
\widetilde{C}_{1} & \widetilde{C}_{2}
\end{array}\right]\left[\begin{array}{c}
\widetilde{x}_{f, s}\left(t-\tau_{o}\right) \\
\widetilde{x}_{f, f}\left(t-\tau_{o}\right)
\end{array}\right] \\
& =C_{1} \widetilde{x}_{s}\left(t-\tau_{o}\right)+\widetilde{C}_{1} \widetilde{x}_{f, s}\left(t-\tau_{o}\right)+\widetilde{C}_{2} \tilde{x}_{f, f}\left(t-\tau_{o}\right), \tag{B.3}
\end{align*}
$$

where $C_{2}=\left[\begin{array}{ll}\widetilde{C}_{1} & \widetilde{C}_{2}\end{array}\right]$. From (B.3) denotes the following equation is satisfied:

$$
\begin{aligned}
\sum_{i=1}^{N_{1}+1} y_{c}\left(t-\tau_{i}^{*}\right)= & \sum_{i=1}^{N_{1}+1} C_{1} \widetilde{x}_{s}\left(t-\tau_{o}-\tau_{i}^{*}\right) \\
& +\sum_{i=1}^{N_{1}+1} \widetilde{C}_{1} \widetilde{x}_{f, s}\left(t-\tau_{o}-\tau_{i}^{*}\right) \\
& +\sum_{i=1}^{N_{1}+1} \widetilde{C}_{2} \widetilde{x}_{f, f}\left(t-\tau_{o}-\tau_{i}^{*}\right)
\end{aligned}
$$

$$
\left.\begin{array}{rl}
= & \sum_{i=1}^{N_{1}+1} C_{1} \widetilde{x}_{s}\left(t-\tau_{s, i}\right) \\
& -\widetilde{C}_{1} \sum_{i=1}^{N_{1}+1} \sum_{j=1}^{N_{2}} \Delta_{1, i}^{-1} \bar{B}_{s, j}^{*} v_{c}\left(t-\tau_{i, j}\right) \\
& -\widetilde{C}_{2} \sum_{i=1}^{N_{1}+1} \sum_{j=1}^{N_{2}} \Delta_{2, i}^{-1} \bar{B}_{f, j}^{*} v_{c}\left(t-\tau_{i, j}\right) \\
= & \sum_{i=1}^{N_{1}+1} C_{1} \widetilde{x}_{s}\left(t-\tau_{s, i}\right)-\left[\widetilde{C}_{1} \widetilde{C}_{2}\right] \\
& \times \sum_{i=1}^{N_{1}+1} \sum_{j=1}^{N_{2}}\left[\begin{array}{cc}
\Delta_{1, i}^{-1} & O \\
O & \Delta_{2, i}^{-1}
\end{array}\right]\left[\bar{B}_{s, j}^{*}\right. \\
\bar{B}_{f, j}^{*}
\end{array}\right] v_{c}\left(t-\tau_{i, j}\right) .
$$

where $\tau_{i}^{*}=\tau_{s, i}-\tau_{o}$. One of the terms $\tau_{i}^{*}$ in (B.4) is set to $\tau_{i}^{*}=0$ and (B.4) can be represented as

$$
\begin{align*}
y_{c}(t) & +\sum_{i=1}^{N_{1}} y_{c}\left(t-\tau_{i}^{*}\right) \\
= & C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-C_{2}\left(\eta \Lambda_{2, k}\right)^{-1} \\
& \times \sum_{j=1}^{N_{2}} \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right)+\sum_{i=1}^{N_{1}} C_{1} \widetilde{x}_{s}\left(t-\tau_{s, i}\right) \\
& \quad-C_{2} \sum_{i=1}^{N_{1}} \sum_{j=1}^{N_{2}}\left(\eta \Lambda_{2, i}\right)^{-1} \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right) \\
= & C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-C_{2}(I)^{-1} \sum_{j=1}^{N_{2}} \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right) \\
& +\sum_{i=1}^{N_{1}} C_{1} \widetilde{x}_{s}\left(t-\tau_{s, i}\right) \\
& -C_{2} \sum_{i=1}^{N_{1}} \sum_{j=1}^{N_{2}}\left(\eta \Lambda_{2, i}\right)^{-1} \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right) . \tag{B.5}
\end{align*}
$$

From (B.5),

$$
\begin{equation*}
y_{c}(t)=C_{1} \tilde{x}_{s}\left(t-\tau_{o}\right)-C_{2} \sum_{j=1}^{N_{2}} \bar{B}_{f k, j} v_{c}\left(t-\tau_{i, j}\right) \tag{B.6}
\end{equation*}
$$

## Acknowledgment

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract no. NSC 101-2511-S-197-002.

## References

[1] J. Lin and Z. Gao, "Exponential admissibility and $H_{\infty}$ control of switched singular time-delay systems: an average dwell time approach," Journal of Applied Mathematics, vol. 2012, Article ID 482792, 28 pages, 2012.
[2] P. Wang, C. Han, and B. Ding, "Stability of discrete-time networked control systems and its extension for robust $H_{\infty}$ control," International Journal of Systems Science, vol. 44, no. 2, pp. 275-288, 2013.
[3] L. Xie, T. Liu, G. Lu, J. Liu, and S. T. C. Wong, "Stochastic robust stability analysis for Markovian jump discrete-time delayed neural networks with multiplicative nonlinear perturbations," in Advances in Neural Networks, vol. 3971 of Lecture Notes in Computer Science, pp. 172-178, Springer, 2006.
[4] X. Zhang, S. Li, and H. Li, "Structure and BIBO stability of a three-dimensional fuzzy two-term control system," Mathematics and Computers in Simulation, vol. 80, no. 10, pp. 1985-2004, 2010.
[5] J. S.-H. Tsai, C.-T. Wang, and L. S. Shieh, "Model conversion and digital redesign of singular systems," Journal of the Franklin Institute, vol. 330, no. 6, pp. 1063-1086, 1993.
[6] B. G. Mertzios, M. A. Christodoulou, B. L. Syrmos, and F. L. Lewis, "Direct controllability and observability time domain conditions of singular systems," IEEE Transactions on Automatic Control, vol. 33, no. 8, pp. 788-791, 1988.
[7] C.-J. Wang and H.-E. Liao, "Impulse observability and impulse controllability of linear time-varying singular systems," Automatica, vol. 37, no. 11, pp. 1867-1872, 2001.
[8] J. Yu, G. Sun, and H. R. Karimi, "Fault-reconstruction-based cascaded sliding mode observers for descriptor linear systems," Mathematical Problems in Engineering, vol. 2012, Article ID 623426, 20 pages, 2012.
[9] J. Wu, H. R. Karimi, and P. Shi, "Network-based $H_{\infty}$ output feedback control for uncertain stochastic systems," Information Sciences, vol. 232, pp. 397-410, 2013.
[10] H. R. Karimi, "Observer-based mixed $\mathrm{H}_{2} / H_{\infty}$ control design of linear systems with time-varying delays: an LMI approach," International Journal of Control, Automation and Systems, vol. 6, no. 1, pp. 1-14, 2008.
[11] M. Chadi and H. R. Karimi, "Robust observer design for unknown inputs Takagi-Sugeno models," IEEE Transactions on Fuzzy Systems, vol. 21, no. 1, pp. 158-164, 2013.
[12] S. Zhou and W. Zhang, "Discrete-time indefinite stochastic LQ control via SDP and LMI methods," Journal of Applied Mathematics, vol. 2012, Article ID 638762, 14 pages, 2012.
[13] T. Zou, "Offset-free strategy by double-layered linear model predictive control," Journal of Applied Mathematics, vol. 2012, Article ID 808327, 14 pages, 2012.
[14] H. Zhang, J. Cao, and W. Jiang, "Controllability criteria for linear fractional differential systems with state delay and impulses," Journal of Applied Mathematics, vol. 2013, Article ID 146010, 9 pages, 2013.
[15] C.-J. Wang and J.-S. Chiou, "A stability condition with delaydependence for a class of switched large-scale time-delay systems," Journal of Applied Mathematics, vol. 2013, Article ID 360170, 7 pages, 2013.
[16] M. S. Mahmoud, Robust Control and Filtering for Time-Delay Systems, vol. 5 of Control Engineering (New York), Marcel Dekker, New York, NY, USA, 2000.
[17] T. Chen and B. A. Francis, Optimal Sample-Data Control Systems, Springer, New York, NY, USA, 1995.
[18] D. Li and Y. Xi, "Constrained feedback robust model predictive control for polytopic uncertain systems with time delays," International Journal of Systems Science, vol. 42, no. 10, pp. 16511660, 2011.
[19] L. S. Shieh, J. S. H. Tsai, and S. R. Lian, "Determining continu-ous-time state equations from discrete-time state equations via the principal qth root method," IEEE Transactions on Automatic Control, vol. 31, no. 5, pp. 454-457, 1986.
[20] F. R. Ganmacher, The Theory of Matrices II, Chelsea, New York, NY, USA, 1974.
[21] R. Nikoukhah, A. S. Willsky, and B. C. Levy, "Boundary-value descriptor systems: well-posedness, reachability and observability," International Journal of Control, vol. 46, no. 5, pp. 17151737, 1987.
[22] S. L. Campbell, Singular Systems of Differential Equations II, Pitman, New York, NY, USA, 1982.
[23] L. S. Shieh, Y. T. Tsay, and C. T. Wang, "Matrix sector functions and their applications to systems theory," IEE Proceedings D: Control Theory and Applications, vol. 131, no. 5, pp. 171-181, 1984.
[24] J. S. H. Tsai, L. S. Shieh, and R. E. Yates, "Fast and stable algorithms for computing the principal nth root of a complex matrix and the matrix sector function," Computers and Mathematics with Applications, vol. 15, no. 11, pp. 903-913, 1988.
[25] W.-M. Wang, S.-M. Guo, and L.-S. Shieh, "Discretization of cascaded continuous-time controllers for state and input delayed systems," International Journal of Systems Science, vol. 31, no. 3, pp. 287-296, 2000.
[26] S.-M. Guo, W. Wang, and L.-S. Shieh, "Discretisation of two degree-of-freedom controller and system with state, input and output delays," IEE Proceedings-Control Theory and Applications, vol. 147, no. 1, pp. 87-96, 2000.
[27] S.-M. Guo, L.-S. Shieh, G. Chen, and C.-F. Lin, "Effective chaotic orbit tracker: a prediction-based digital redesign approach," IEEE Transactions on Circuits and Systems I, vol. 47, no. 11, pp. 1557-1570, 2000.
[28] M. Hou, P. Zítek, and R. J. Patton, "An observer design for linear time-delay systems," IEEE Transactions on Automatic Control, vol. 47, no. 1, pp. 121-125, 2002.
[29] F. L. Lewis, Optimal Control, John Wiley \& Sons, New York, NY, USA, 1986.
[30] K. Ogata, Discrete-Time Control Systems, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.
[31] K. Watanabe and T. Ouchi, "An observer of systems with delays in state variables," International Journal of Control, vol. 41, no. 1, pp. 217-229, 1985.
[32] J. Leyva-Ramos and A. E. Pearson, "Asymptotic modal observer for linear autonomous time lag systems," IEEE Transactions on Automatic Control, vol. 40, no. 7, pp. 1291-1294, 1995.
[33] J. S.-H. Tsai, N.-T. Hu, P.-C. Yang, S.-M. Guo, and L.-S. Shieh, "Modeling of decentralized linear observer and tracker for a class of unknown interconnected large-scale sampled-data
nonlinear systems with closed-loop decoupling property," Computers and Mathematics with Applications, vol. 60, no. 3, pp. 541562, 2010.
[34] N. T. Hu, J. S. H. Tsai, S. M. Guo, L. S. Shieh, and Y. Chen, "Loworder multi-rate linear time-invariant decentralized trackers using the new observer-based sub-optimal method for unknown sampled-data nonlinear time-delay system with closed-loop decoupling," Optimal Control Applications and Methods, vol. 32, no. 4, pp. 433-475, 2011.

## Research Article

# Practical Aspects of Broken Rotor Bars Detection in PWM Voltage-Source-Inverter-Fed Squirrel-Cage Induction Motors 

Hong-yu Zhu, ${ }^{1,2,3,4}$ Jing-tao Hu, ${ }^{1,2,3}$ Lei Gao, ${ }^{1,2,3}$ and Hao Huang ${ }^{1,3}$<br>${ }^{1}$ Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China<br>${ }^{2}$ University of Chinese Academy of Sciences, Beijing 100049, China<br>${ }^{3}$ Key laboratory of Industrial Information Technology, Chinese Academy of Sciences, Shenyang 110016, China<br>${ }^{4}$ School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114044, China<br>Correspondence should be addressed to Hong-yu Zhu; zhuhongyu@sia.cn

Received 5 June 2013; Revised 18 September 2013; Accepted 23 October 2013
Academic Editor: Xianxia Zhang
Copyright © 2013 Hong-yu Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Broken rotor bars fault detection in inverter-fed squirrel cage induction motors is still as difficult as the dynamics introduced by the control system or the dynamically changing excitation (stator) frequency. This paper introduces a novel fault diagnosis techniques using motor current signature analysis (MCSA) to solve the problems. Switching function concept and frequency modulation theory are firstly used to model fault current signal. The competency of the amplitude of the sideband components at frequencies $(1 \pm 2 s) f_{s}$ as indices for broken bars recognition is subsequently studied in the controlled motor via openloop constant voltage/frequency control method. The proposed techniques are composed of five modules of anti-aliasing signal acquisition, optimal-slip-estimation based on torque-speed characteristic curve of squirrel cage motor with different load types, fault characteristic frequency determination, nonparametric spectrum estimation, and fault identification for achieving MCSA efficiently. Experimental and simulation results obtained on 3 kW three-phase squirrel-cage induction motors show that the model and the proposed techniques are effective and accurate.


## 1. Introduction

Squirrel-cage induction motors have dominated the field of electromechanical energy conversion. They consume more than $60 \%$ of the electrical energy produced and are present in the main industrial applications [1, 2]. Although still considered proverbially robust, all components of induction machines are subject to increased stress, particularly when operated in a controlled mode and with repeated load cycles [3]. According to studies, broken rotor bars and cracked end-ring faults in the rotor cage are responsible for about $5-10 \%$ of all breakdowns and incipient detection of these events remains a key issue $[4,5]$. The main reason why early detection is important is that although broken rotor bars may not cause immediate failure, there can be serious secondary effects associated with their presence [6, 7]. If faults are undetected, they may lead to potentially catastrophic failures. Thus, health-monitoring techniques to prevent the induction motor failures are of great concern in the industry and are gaining an increasing attention.

Motor current signature analysis (MCSA) has been successfully applied to detect broken-rotor bar faults by investigating the sideband components around the supplied current fundamental frequency (i.e., the line frequency) $f$

$$
\begin{equation*}
f_{b}=(1 \pm 2 s) f \tag{1}
\end{equation*}
$$

where $f_{b}$ are the sideband frequencies associated with the broken rotor bars and $s$ is the per unit motor slip [4, 8, 9]. Some quantitative conclusion based on the left sideband values (LSB) of the amplitude-frequency spectrum plot of a motor phase current even recommends $-50 \mathrm{~dB} \sim-35 \mathrm{~dB}$ of the sideband harmonics as the threshold level for generating a warning or alarm [10]. In addition, MCSA has been tested in many industrial cases since the 1980 s with good results [1116].

Even though numerous successful main-fed motor fault detection methods are reported in the literature, bibliography relative to inverter-fed motors on this topic seems to be poor. With the increased emphasis on energy conservation and
lower cost, induction machines are supplied and controlled by inverters, and the use of inverter drives increases day by day. As a result, dynamic performance is excellent due to all kinds of control methods and the next steps in drive development are going to be driven by attempts to increase reliability and reduce maintenance costs. By using the current sensor feedback and microprocessing unit, the new tread for low-cost protection applications seems to be drive-integrated fault diagnosis systems without using any external hardware. Thus, it becomes more demanding to detect faults by using MCSA in these drives.

The introduction of voltage-source-inverter-fed (VSIfed) motors has produced significant changes in the field of diagnosis and control, needing further research in order to overcome various challenges. Contrary to the motor line current taken directly from the main, the inverter-fed motor line current includes remarkable noise (inherent floor noise which reduces the possibility of true fault pattern recognition using line current spectrum) due to the high-frequency switching devices, EMI and EMC effects, and so forth. The current signal for rotor fault diagnosis needs precise and high resolution information, which means the diagnosis system demands additional hardware such as a low-pass filter, highprecision AD converter, and additional software. Moreover, closed-loop control in IM drives introduces new difficulties in broken rotor bar detection, as traditional fault indicators tend to be masked by a control algorithm. Therefore, the methods developed for an open-loop operation are not able to produce adequate and reliable information on the extent of the fault and have to be adapted $[4,17,18]$. All these influences complicate the utilization of frequency analysis methods. VSI-fed motor faults have been analyzed, and the initial results are given in the literature but further investigation is still required.

The classic MCSA method works well under constant load torque and with high-power motors, and it has mainly been designed for fixed frequency supply, such as for machines connected to the electrical grid. To obtain satisfactory test results, in [7], recommendations are given from the author's experience. As stated in this paper, literally, "the load on the motor should be sufficiently high to produce at least $35 \%$ of rated full load rotor slip for a reliable single-test broken bar analysis." and "if the motor to be tested is fed from a variable-speed converter drive, the frequency of the drive should be locked at the $50 / 60 \mathrm{~Hz}$ power supply frequency for the test." Yet, difficulties emerge when inverted-fed motors are applied to drive fans, pumps, or other mechanisms involving speed control for energy-saving purpose. In these cases, the excitation frequency will truly depend on the speed reference and the load applied to the system. Therefore, unlike the utility-driven case, the stator excitation frequency will dynamically change and the position of the current harmonics appearing on the stator-current spectrum due to electrical faults is highly dependent on the mechanical motor load and excitation frequency, which affects the slip frequency. As a consequence, the conventional MCSA must be amended to accommodate the new scenarios. Unfortunately, to the best of our knowledge, in the published literatures there is no research work on this subject.


Figure 1: Schematic diagram of the PWM VSI-fed adjustable speed drive.

In this context, following explicitly derivation from a simplified fault signal model, a new online fault diagnosis technique based on MCSA for inverter-fed squirrel-cage induction motors is present. Compared with traditional MCSA, the novelty of the proposed method is that broken rotor bars fault in the controlled motor via open-loop constant voltage/frequency control method could be identified even if the motor operates at different excitation frequencies. To do this, oversampling signal acquisition technique is used to suppress significantly noise contained in the invertedfed motor line current, and fault-indicative frequencies with variable excitation frequency are determined by torque-speed characteristic curve of squirrel-cage motor with different load types. To obtain satisfactory results, nonparametric spectrum estimation algorithm and fault identification are subsequently presented. Including this introductory section, this paper is organized into six sections. Section 2 presents a theoretical analysis model of stator current of inverter-fed squirrel-cage motor, which is based on switching function concept and modulation theory. A detailed description of the harmonic components contained in current is given. Section 3 elaborates the new broken rotor bars fault diagnosis techniques. In Section 5, the proposed techniques are validated by laboratory tests; the method is applied to different stator currents obtained from healthy and faulty machines. Experimental and simulation results as well as the corresponding analysis and discussion are presented in Section 4. Finally, conclusions and recommendations are presented in the last section.

## 2. Analytic Model of Stator Current Signature for Squirrel-Cage Induction Motor with Constant Volt-per-Hertz Control Technique

2.1. Analytic Model of Stator Current Signature for No-Fault Squirrel-Cage Induction Motor. Switching function concept is a powerful tool in understanding and optimizing the performance of the converter [19, 20]. In [20], an analytical approach for characterizing the current harmonics and interharmonics of the VSI-fed ASD injected into the supply system in steady is presented using the switching function concept, applicable to PWM VSI that is studied in this paper. Figure 1 shows the schematic diagram of a typical PWM VSI-fed adjustable speed drive, where $S_{i a}, S_{i b}$, and $S_{i c}$ represent the rectifier converter current switching functions that correlate
the three-phase AC source currents $i_{a}^{\prime}, i_{b}^{\prime}$, and $i_{c}^{\prime}$ and the inverter input DC current $i_{d}, S_{u a}, S_{u b}$, and $S_{u c}$ represent the inverter voltage switching functions that correlate inverter dc input voltage and output phase voltages $u_{a}, u_{b}$ and $u_{c}, Z$ is the impedance operator seen from the inverter output terminals corresponding to the neutral point of induction motor. In order to calculate and analyse the harmonic currents generated by the VSI-fed ASD, an analytical model based on the modulation theory and switching function concept is proposed and expressed by (2)~(6). After obtaining the phase currents of PWM VSI-fed motor, the harmonic components of the current might be extracted by the use of a certain spectrum analysis method as follows:

$$
\begin{gather*}
\left(i_{a}^{\prime}, i_{b}^{\prime}, i_{c}^{\prime}\right)=i_{d}\left(S_{i a}, S_{i b}, S_{i c}\right)  \tag{2}\\
i_{d}=\left(i_{a}, i_{b}, i_{c}\right)\left(S_{u a}, S_{u b}, S_{u c}\right)^{\prime}  \tag{3}\\
u_{d}=\left(u_{a}^{\prime}, u_{b}^{\prime}, u_{c}^{\prime}\right)\left(S_{i a}, S_{i b}, S_{i c}\right)^{\prime},  \tag{4}\\
\left(u_{a}, u_{b}, u_{c}\right)=u_{d}\left(S_{u a}, S_{u b}, S_{u c}\right),  \tag{5}\\
\quad\left(i_{a}, i_{b}, i_{c}\right)=\frac{\left(u_{a}, u_{b}, u_{c}\right)}{Z} \tag{6}
\end{gather*}
$$

As shown in Figure 1, for the two-level natural sampled PWM with a triangular wave of carrier signal, the three-phase voltage switching functions $S_{u a}, S_{u b}$ and $S_{u c}$ can be expressed as the following double Fourier series, as follows:

$$
\begin{aligned}
S_{u a}(t)= & \frac{M}{2} \sin \left(\omega_{s} t\right)+\frac{2}{\pi} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \\
& \cdot \sin \left(\frac{m \pi}{2}\right) \cdot \sin \left(m \omega_{c} t\right) \\
& +\frac{2}{\pi} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{\infty} \frac{J_{n}(m M \pi / 2)}{m} \sin \left[\frac{(m+n) \pi}{2}\right] \\
& \times \sin \left(m \omega_{c} t+n \omega_{s} t\right), \\
S_{u b}(t)= & \frac{M}{2} \sin \left(\omega_{s} t-\frac{2 \pi}{3}\right)+\frac{2}{\pi} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \\
& \cdot \sin \left(\frac{m \pi}{2}\right) \cdot \sin \left(m \omega_{c} t\right) \\
& +\frac{2}{\pi} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{\infty} \frac{J_{n}(m M \pi / 2)}{m} \sin \left[\frac{(m+n) \pi}{2}\right] \\
& \times \sin \left[m \omega_{c} t+n\left(\omega_{s} t-\frac{2 \pi}{3}\right)\right], \\
S_{u c}(t)= & \frac{M}{2} \sin \left(\omega_{s} t+\frac{2 \pi}{3}\right)+\frac{2}{\pi} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \\
& \cdot \sin \left(\frac{m \pi}{2}\right) \cdot \sin \left(m \omega_{c} t\right)
\end{aligned}
$$

$$
\begin{align*}
& +\frac{2}{\pi} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{\infty} \frac{J_{n}(m M \pi / 2)}{m} \sin \left[\frac{(m+n) \pi}{2}\right] \\
& \times \sin \left[m \omega_{c} t+n\left(\omega_{s} t+\frac{2 \pi}{3}\right)\right] \tag{7}
\end{align*}
$$

For a hypothetical ideal motor supplied from a balanced three-phase source of sinusoidal voltages and driving a constant load, the following waveform of selected phase-a stator current may be assumed by substituting (7) into (5) and (6), where the commutation effect is ignored and proper origin of coordination selected for convenience of analysis is as follows:

$$
\begin{align*}
i_{a}= & \frac{M E_{d}}{2\left|Z\left(\omega_{s}\right)\right|} \cos \left[\omega_{s} t-\varphi\left(Z\left(\omega_{s}\right)\right)\right] \\
& +\frac{2 E_{d}}{\pi\left|Z\left(m N \omega_{s}\right)\right|} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \sin \left(\frac{m \pi}{2}\right) \\
& \times \cos \left[m N \omega_{s} t-\varphi\left(Z\left(m N \omega_{s}\right)\right)\right] \\
& +\frac{2 E_{d}}{\pi\left|Z\left((m N+n) \omega_{s}\right)\right|} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{ \pm \infty} \frac{J_{n}(m M \pi / 2)}{m} \\
& . \sin \left[\frac{(m+n) \pi}{2}\right] \cdot \cos \left[(m N+n) \omega_{s} t\right.  \tag{8}\\
= & I_{1} \cos \left(\omega_{s} t-\varphi_{1}\right)+I_{2} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \sin \left(\frac{m \pi}{2}\right) \\
& \times \cos \left(m N \omega_{s} t-\varphi_{2}\right) \\
& +I_{3} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{ \pm \infty} \frac{J_{n}(m M \pi / 2)}{m} \cdot \sin \left[\frac{(m+n) \pi}{2}\right] \\
& \times \cos \left[(m N+n) \omega_{s} t-\varphi_{3}\right],
\end{align*}
$$

where $M$ is amplitude modulation index, $M=U_{s} / U_{c} \leq 1 ; \omega_{s}$ is frequency of the modulation waveform (reference) (rad/s), $\omega_{s}=2 \pi f_{s} ; \omega_{c}$ is frequency of the carrier signal ( $\mathrm{rad} / \mathrm{s}$ ); $N$ ratio of the carrier frequency to the modulation frequency, $N=\omega_{c} / \omega_{s} \geq 1$; and $J_{0}, J_{n}$ are Bessel functions of the first kind with order 0 and order $n$, respectively.
2.2. Analytic Model of Stator Current Signature for Faulty Squirrel-Cage Induction Motor. When broken rotor bar or cracked end-ring faults develop in the rotor cage, the current, torque, and speed of the motor are affected, typically, in a periodic manner. In the case of periodic disturbances, all three line currents $i_{a}$, $i_{b}$, and $i_{c}$ are simultaneously modulated with the fundamental frequency $f_{o}$ of the fault-induced oscillation of motor variables. If only amplitude modulation and fundamental frequency $f_{o}$ are considered, current in phase-a of the supply line may now be expressed as

$$
\begin{equation*}
i_{f}=i_{a}\left[1+a \cos \left(\omega_{o} t\right)\right] \tag{9}
\end{equation*}
$$

where $a$ denotes the modulation depth (modulation index) and $\omega_{o}=2 \pi f_{o}=2 s \omega_{s}=4 \pi s f_{s}$. The value of the modulation index depends on the severity of the abnormality and motor loads.

Substituting (8) in (9) yields

$$
\begin{align*}
i_{f}= & I_{1} \cos \left(2 \pi f_{s} t-\varphi_{1}\right)+I_{2} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \sin \left(\frac{m \pi}{2}\right) \\
& \times \cos \left(2 \pi m N f_{s} t-\varphi_{2}\right) \\
& +I_{3} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{ \pm \infty} \frac{J_{n}(m M \pi / 2)}{m} \cdot \sin \left[\frac{(m+n) \pi}{2}\right] \\
& \times \cos \left[2 \pi(m N+n) f_{s} t-\varphi_{3}\right] \\
& +a I_{1}\left\{\cos \left[2 \pi(1+2 s) f_{s} t-\varphi_{1}\right]\right. \\
& \left.+\cos \left[2 \pi(1-2 s) f_{s} t-\varphi_{1}\right]\right\}  \tag{10}\\
& +\left\{I_{2} \sum_{m=1}^{\infty} J_{0}\left(\frac{m M \pi}{2}\right) \cdot \sin \left(\frac{m \pi}{2}\right)\right. \\
& +\cos \left[2 \pi(m N+2 s) f_{s} t-\varphi_{2}\right] \\
& +a I_{3} \sum_{m=1}^{\infty} \sum_{n= \pm 1}^{ \pm \infty} \frac{J_{n}(m M \pi / 2)}{m} \cdot \sin \left[\frac{(m+n) \pi}{2}\right] \\
& \cdot\left\{\cos \left[2 \pi(m N+n+2 s) f_{s} t-\varphi_{3}\right]\right. \\
& \left.+\cos \left[2 \pi(m N+n-2 s) f_{s} t-\varphi_{3}\right]\right\},
\end{align*}
$$

indicating that the spectrum of stator current for inverter-fed healthy squirrel-cage motor contains, apart from the fundamental $f_{s}$ equal to the inverter excitation (stator) frequency, $m N f_{s}$ harmonics at the carrier frequency and multiples of the carrier frequency and $(M m+n) f_{s}$ harmonics in the sidebands around each multiple of the carrier frequency. When a bar breaks, a rotor asymmetry occurs. The result is that it induces in the stator current additional frequency at $f_{b}=(1 \pm 2 s) f_{s}$, $f_{b 1}=(m N \pm 2 s) f_{s}$, and $f_{b 2}=(m N+n \pm 2 s) f_{s}$ around harmonics frequency depicted above. The amplitude of faultindicative frequencies $f_{b}=(1 \pm 2 s) f_{s}, f_{b 1}=(m N \pm 2 s) f_{s}$, and $f_{b 2}=(m N+n \pm 2 s) f_{s}$ depends on faulty severity, loads and excitation (stator) frequency, and the distance between $f_{b}, f_{b 1}$, and $f_{b 2}$ and corresponding harmonics frequency $f_{s}$, $m N f s,(M m+n) f_{s}$ is equal to $2 s f_{s}$. The amplitude of $f_{b 1}$ and $f_{b 2}$ can be considered negligible compared to that of $f_{b}$; as a result, $f_{b}=(1 \pm 2 s) f_{s}$ are adopted as fault-indicative frequencies.

In practice, the current is modulated with respect to not only the amplitude but also the phase, and the modulation process is more complex than that described by (9), but the derived equation (10) provides an adequate basis for qualitative assessment of diagnostic media.

## 3. Diagnosis Techniques of Broken Rotor Bars for Squirrel-Cage Induction Motor with Constant Voltage/Frequency Control Method

3.1. Principles of Diagnosis. A motor diagnosis technique, which contains the five processing modules illustrated in Figure 2, is presented, based on model of squirrel-cage induction motor stator current signature depicted in Section 2. The current flowing in single phase of the induction motor is sensed by anti-aliasing signal acquisition module and sent to spectrum estimation module, where the obtained time-domain signal can be decomposed into components of different frequency using Welch's periodogram method. In optimal-slip-estimation module, based on the real speed, the inverter excitation frequency, and torque-speed characteristic curve of squirrel-cage motor with different load types, the motor slip is calculated and the consequent optimal slip estimation value is transmitted to fault characteristic frequency determination module, where characteristic frequencies of broken rotor bars are calculated. Depending on whether the characteristic frequencies of $f_{b}=(1 \pm 2 s) f_{s}$ could be found in power spectrum obtained in spectrum estimation module, the conclusion of failure or not could be drawn, this work is done in fault identification module.
3.2. Method of Antialiasing Signal Acquisition. When carrying out diagnostic analysis one of the key elements to obtain good results is to properly choose acquisition parameters: the sampling frequency and the number of samples. There are two different constraints: analysis signal bandwidth and frequency resolution for the spectrum analysis.

It must be considered, when capturing the stator current signal, that the sampling frequency $f_{\text {sample }}$ plays an important role. Generally, statistical bands of fault-indicative frequencies can be ascertained from theory analysis and many MCSA experiments [2, 8, 21-23]. Table 1 summarizes the range of sideband frequencies in terms of faulty types. We see in Table 1 that the spectral analysis of stator currents might be done in low frequency range (typically between 0 and 2 kHz ) to focus on the significant phenomena [24], and in the case of broken rotor bars fault, the fault-indicative frequencies are under 400 Hz . Thus, taking into account the Nyquist criterion, a very high sampling frequency value is not mandatory in case that the motors are supplied by the ac grid. Sampling frequencies of 2 k or 5 k samples $/ \mathrm{s}$ (standard in data acquisition devices) enable good resolution analysis.

In contrast, for the inverter-fed motors, sampling procedure is more demanding. As it is known, stator current contains high-order frequency harmonics, in this case, due to switching frequencies in modern inverter, typically above 10 kHz . Thus, aliasing may occur due to wrong sampling. Antialiasing techniques have to be used in order to have a correct current spectrum and prevent a false failure alert.

For decreasing alias to an acceptable level, it is common to design a sharp-cutoff low-pass anti-aliasing filter and sample the signal at a frequency lower than the dominant frequency components such as the fundamental and the switching


Figure 2: Schematic diagram of fault diagnosis.

Table 1: Range of sideband frequencies in terms of faulty types.

| Fault types | Theoretical formula | Range | Frequencies |
| :--- | :---: | :---: | :---: |
| Air-gap eccentricity | $f_{\text {ecc }}=f_{s}\left[1 \pm m\left(\frac{1-s}{p}\right)\right]$ | Low, high | $0 \sim 900 \mathrm{~Hz}$ |
| Broken rotor bars | $f_{b}=f_{s}\left[k\left(\frac{1-s}{p}\right) \pm s\right]$ | Low | $0 \sim 400 \mathrm{~Hz}$ |
| Bearing failure | $f_{i, o}=\frac{n_{b}}{2} f_{r}\left[1 \pm \frac{b_{d}}{p_{d}} \cos \beta\right]$ | High | $0 \sim 2000 \mathrm{~Hz}$ |
| Interturn short circuit | $f_{\text {stl }}=f_{s}\left[\frac{m}{p}(1-s) \pm k\right], \quad f_{\text {sth }}=f_{s}\left[1 \pm m Z_{2}\left(\frac{1-s}{p}\right)\right]$ | Low, high | $0 \sim 1000 \mathrm{~Hz}$ |

frequencies [25]. However, such sharp-cutoff analog filters are difficult and expensive to implement, and if the system is to operate with a variable sampling rate, adjustable filters would be required. Furthermore, sharp-cutoff analog filters generally have a highly nonlinear phase response, particularly at the pass-band edge. In our proposed method, oversampling as an alternative technique is used. The principle of oversampling is briefly reviewed as follows (see [26] for details). A very simple anti-aliasing filter that has a gradual cutoff with significant attenuation is firstly applied to prefilter the motor stator phase-a current. Next, implement the A/D conversion at a higher sampling rate. After that, downsampling the obtained signal with a lower sampling frequency is implemented using a digital low-pass filter.

As for the inverter supply, several harmonics could be mixed up in case that low resolution of the band side was chosen. In general, one can take the following steps to select data acquisition parameters in order to achieve the correct resolution needed.
(1) Definition of sampling frequency. For anti-aliasing purpose, $f_{\text {sample }}$ has limitation as $f_{\text {sample }} \geqslant 2 f_{c}$, where $f_{c}$ is the highest fault-indicative frequency.
(2) Selection of required frequency resolution $\Delta f$.
(3) Specification of the number of samples $N=$ $f_{\text {sample }} / \Delta f$.
(4) Determination of sampling time $T=N / f_{\text {sample }}=$ $N T_{s}$.
3.3. Calculation of Slip and Fault-Indicative Frequencies with Variable Excitation Frequency. From Section 2, broken rotor bars can be detected by monitoring the stator current spectral components $f_{b}=(1 \pm 2 s) f_{s}$, where harmonic frequencies $f_{b}$ are a function of slip $s$ and excitation frequency $f_{s}$. In case of broken rotor bars fault, excitation frequency $f_{s}$ and the corresponding slip $s$ must be firstly determined in order to find harmonic frequencies $f_{b}$. The result of a motor diagnosis using MCSA is incorrect if the detected slip has an error [8, 27]. One of the most popular ways to obtain the information of the slip frequency is to use speed sensor. In our proposed method, slip $s$ is calculated in optimal-slip-estimation module, based on the real measured speed and torque-speed characteristic curve of squirrel-cage motor.

The torque-speed characteristic curve of different load types is shown in Figure 3. Curves (1), (2) and (3) are inherent characteristic curves of squirrel-cage motor corresponding to different excitation frequency $f_{s}$. For ease of analysis, it is assumed that torque characteristic curves (1), (2) and (3) intersect mechanical characteristic curve $f_{1}$ at nominal operating point $A$ and only excitation frequency $f_{s}<f_{1}$ is considered (if $f_{s}>f_{1}$, motor operates at point $A_{3}$ of curve (3)). In Figure 3, the value of slip $s$ corresponding to point $A$ is $s_{e}=\left(n_{01}-n_{e}\right) / n_{01}$. The no-load speedes corresponding to different excitation frequency can be expressed as $n_{01}=$ $60 f_{1} / p$ and $n_{02}=60 f_{s} / p$, where $p$ denotes the number of pole-pairs.


Figure 3: Torque-speed characteristic curve of squirrel-cage motor with different load types.

If motor works with constant torque load, the operate point shifts from $A$ to $A_{1}$, the slip $s$ corresponding to $A_{1}$ is equal to $s_{\text {constant }}=\left(n_{02}-n_{1}\right) / n_{02}$. Considering the congruent relationship between $\Delta n_{01} n_{e} A$ and $\Delta n_{02} n_{1} A_{1}$, we can deduce the specific slip formula as follows: $s_{\text {constant }}=s_{e} /\left(f_{1} / f_{s}\right)$.

If motor works with fans/pump load, the operating point shifts from $A$ to $A_{2}$ and the slip $s$ corresponding to $A_{2}$ is equal to $s_{\text {fans } / \text { pump }}=\left(n_{02}-n_{2}\right) / n_{02}$. Considering the similarity of $\Delta n_{01} n_{e} A$ to $\Delta n_{02} n_{1} A_{1}$ and relationship of $T / T_{e}=\left(n_{2} / n_{e}\right)^{2}$ between speed and torque of fans/pump, we can deduce the specific slip formula as follows: $s_{\text {fans/pump }} /\left(1-s_{\text {fans/pump }}\right)^{2}=$ $\left[s_{e} /\left(1-s_{e}\right)^{2}\right] /\left(f_{s} / f_{1}\right)$.
3.4. Nonparametric Spectrum Estimation Algorithms. MCSA techniques include parametric and nonparametric spectrum analysis of the motor current in general [28]. Among the nonparametric algorithms, we use Welch's periodogram algorithms based on DFT to compute the power spectrum of the phase-a motor current data.

Let $i_{N}[n]=\{i[0], i[1], \ldots, i[N-1]\}$ be a discrete time signal, which is obtained by sampling the phase-a motor current signal $i(t)$ for a duration of sampling time $T$. To reduce the variance of power spectrum estimate, the N point data sequence, $i_{N}[n]=\{i[0], i[1], \ldots, i[N-1]\}$, is first partitioned into $K$ overlapping segments. The length of each segment consists of Lsamples and these segments can be overlapping on each other with ( $L-S$ ) overlapping samples, where $S$ is the number of points to shift between segments. Thereafter, the periodogram of each segment is calculated and the obtained periodograms are then averaged to give the power spectrum estimate.

The length of segment $L$ is dependent on the required resolution. In order to increase the quality of power spectrum estimates, the signal segments can be windowed before calculating FFT. The proposed methods permit reduce the variance of the estimate at the expense of a decreased frequency resolution. However, it is difficult to trade off between the frequency resolution and the estimate variance. It has been noted that the use of $50 \%$ overlapping percentage among the partitioned segments leads to efficient implementation of the
fast Fourier transform (FFT) algorithm and in this case the relationship between $K$ and $L$ of segment as follows $K=$ ( $N-L / 2$ )/( $L / 2$ ).

## Algorithm 1.

Step 1. Subdividing $N$-point sampled data sequence, $i_{N}[n]=$ $\{i[0], i[1], \ldots i[N-1]\}$, into $K$ overlapping segments; the $k$ th segment data $x[n+k S], 0 \leqslant n \leqslant L-1$ and $0 \leqslant k \leqslant K-1$, is as follows:

Segment 1: $x[0], x[1], \ldots, x[L-1]$;
Segment 1: $x[S], x[S+1], \ldots, x[L+S-1]$;

Segment $K$ : $x[N-L], x[N-L+1], \ldots, x[N-1]$.
Step 2. Weighting $k$ th segment, $d[n]$ denote rectangular window function:

$$
\begin{equation*}
x^{k}[n]=d[n] x[n+k S], \quad 0 \leqslant n \leqslant L-1,0 \leqslant k \leqslant K-1 . \tag{11}
\end{equation*}
$$

Step 3. Calculating power spectrum of the $k$ th segment data:

$$
\begin{equation*}
P^{k}(f)=\frac{1}{\mathrm{ULT}} X(f)[X(f)]^{*}=\frac{1}{\mathrm{ULT}}|X(f)|^{2} \tag{12}
\end{equation*}
$$

where $X(f)=T \sum_{n=0}^{L-1} x^{(k)}[n] e^{-j 2 \pi f n T}$ denote DFT of the $k$ th segment data and $U=T \sum_{n=0}^{L-1} d^{2}[n]$ denote normalization factor.

Step 4. Averaging all segments power spectrum:

$$
\begin{equation*}
\widetilde{P}(f)=\frac{1}{K} \sum_{k=0}^{K-1} P^{k}(f) \tag{13}
\end{equation*}
$$

## 4. Experiment Setup and Signal Acquisition Methods

4.1. Experiment Setup. Schematic diagram of the experiment setup is shown in Figure 4. This system can be used to sample line current $i_{a}$ and line voltage $u_{a b}$ (if necessary, it can be arranged to sample the other line currents $i_{b}$ and $i_{c}$ and line voltages $u_{b c}$ and $u_{c a}$, too), and speed signals. The main components of the experiment setup are as follows.
(1) Three-phase 3 kW SCIM (see Table 2 for details).
(2) Digital stroboscope coupled with the shaft of the SCIM as angular-speed sensor to measure and record the time variation of the speed.
(3) DC generator coupled with the SCIM to provide its adjustable load.
(4) Mechanical coupling between SCIM and dc generator.
(5) Variable resistor bank as a variable load of the generator: the load of the generator and, consequently, the induction motor can be adjusted by varying this


Figure 4: Schematic diagram of the experiment.
resistance and/or regulating the excitation current of the generator by relevant variable resistor. The resistance of this variable-resistor bank can be selected step by step by a selector on the bank. In the operating motor, a suitable position of the selector is selected and consequently the induction motor is loaded. By regulating the output voltage of the generator inserted in the excitation current path, the load level is regulated precisely.
(6) Induction-motor drive system type Simens 440 with rating values in accordance with that of the SCIM: this drive has been mainly designed and built for openloop scalar controller in constant voltage/frequency (CV/f) method.
(7) Three-phase change-over switch to exchange the motor connections from the mains for the drive output.
(8) Signal conditioning circuits: since the used DAQ card accepts only voltage signals with maximum amplitudes of $\pm 10 \mathrm{~V}$, the type and amplitude of the signals are prepared before connection to the card. At the first stage, TBC300LTP Hall-effect current transformer is used to prepare the current signal and isolate it from the power circuit. Secondary side current is then converted to proportional voltage signals by current shunts. Then, all signals are transmitted to the DAQ card using a special shielded cable.
(9) PC equipped with a self-made data acquisition card for sampling the electrical data at a certain adjustable frequency and storing them in the memory.

Experiment setup to collect motor data and broken rotor used in the tests are illustrated in Figure 5.
4.2. Signal Acquisition Requirements. The experiments involved collecting the phase-a stator current and speed data of the induction motor for different load conditions and different excitation frequencies of $20 \mathrm{~Hz}, 32 \mathrm{~Hz}, 40 \mathrm{~Hz}$ and 50 Hz , with three broken-rotor-bar fault and without any fault. The load conditions of the motor are $25 \%$, $50 \%, 75 \%$, and $94 \%$ full load, respectively. These load

Table 2: Induction motor technical data.

| Parameter | Value | Unit |
| :--- | :---: | :---: |
| Nominal power | 3.0 | kW |
| Nominal voltage | 380 | V |
| Nominal current | 6.8 | A |
| Nominal frequency | 50 | Hz |
| Connection | $\Delta$ |  |
| Number of poles | 4 |  |
| Rotor slots number | 28 |  |

condition percentages are determined according to the motor nameplate information given in Table 2.

Signal over-sampling method has been chosen in order to avoid aliasing. The stator current of motor was sampled with a frequency of 2 kHz for main-fed and 40 kHz for inverter-fed case. In the inverter-fed case, software filters have been implemented in order to avoid aliasing. More specifically, an 8-order anti-aliasing digital butter-worth filter was implemented and resampling of the signal has been done at 2 kHz . In our case, thirty seconds long data is acquired from all two motors for each load condition at each frequency mentioned above. Thus, the analyzed frequencies vary from 0 to 1 kHz with a resolution of 0.03 Hz . For the feature extraction and discriminant analysis, starting with the first sample, the acquired data is processed with a sliding window size of 30,000 samples at a slide amount of 10,000 , resulting in 60 different data sets all together.

Figure 6 illustrates the intercept parts of 50 Hz stator current signal from the (a) main-fed and (b) inverterfed healthy squirrel-cage motors. An expert inspection of these waveforms reveals that the inverter-fed motor current waveform is heavily contaminated by the noise-like additive waveform due to PWM switching of the voltage source inverter. If we were to use this motor current waveform data to extract necessary features for fault detection and classification, we need to preprocess the data.

## 5. Experimental Results

5.1. Experiment 1. A nominally healthy squirrel-cage motor was firstly tested and the power spectrum of the stator current centered on the fundamental component supplied by (a) main and (b) inverter is shown in Figure 7. The results confirm that the motor rotor is healthy since the sidebands $f_{b}$ are not present. Comparisons of Figures 7(a) and 7(b) shows that a large amount of harmonic components is included in Figure 7(b) and the inverter supply does affect the spectrumidentifiability. Figure 8 indicates power spectrum of the stator current around the fundamental component for (a) mainfed and (b) inverter-fed fault squirrel-cage motors with three broken bars and full load. The annotations appearing in the figure denote the main sideband components around the supply frequency and corresponding amplitudes. Comparison of Figures 7 and 8 indicates that sideband components appear, which demonstrates the broken bars occurred. Although more frequency information appears, one could still identify


Figure 5: Experiment setup to collect motor data and broken rotor used in the tests.


Figure 6: 50 Hz stator current signal from the (a) main-fed and (b) inverter-fed healthy squirrel-cage motor.


FIgURe 7: Power spectrum of the stator current around the fundamental component for (a) main-fed and (b) inverter-fed healthy squirrel-cage motor.


Figure 8: Power spectrum of the stator current around the fundamental component for (a) main-fed and (b) inverter-fed fault squirrel-cage motor.


FIgURE 9: Power spectrum of the stator current around different excitation frequency of (a) 45 Hz , (b) 40 Hz , (c) 32 Hz , and (d) 20 Hz .


Figure 10: Power spectrum of the stator current around (left) $25 \%$ and (right) $50 \%$ full loads. Top row: 32 Hz and bottom row: 50 Hz reference frequency.
in Figure 8(b) broken rotor bars harmonics at 45 Hz and 55 Hz using our proposed method. Comparison of Figures 8(a) and 8(b) also indicates that, for the same level of damage at the same load, the spectrum sidebands have the same amplitude for different supply.
5.2. Experiment 2. The second experiment involved fault squirrel-cage motor operating at different excitation frequency of inverter with three broken basr and full load.

Figure 9 gives the power spectrum of the stator current around the excitation frequencies of $45 \mathrm{~Hz}, 40 \mathrm{~Hz}, 32 \mathrm{~Hz}$, and 20 Hz . As expected, the sideband components $f_{b}=(1 \pm 2 s) f_{s}$, depending on excitation frequency and load are present. Comparison of Figures 9(a), 9(b), 9(c), and 9(d) indicates that the left sideband harmonic component $(1-2 s) f_{s}$ varies from $(41 \mathrm{~Hz},-36 \mathrm{~dB})$ to $(18 \mathrm{~Hz},-42 \mathrm{~dB})$ and right harmonic $(1+$ 2s) $f_{s}$ from $(49 \mathrm{~Hz},-39 \mathrm{~dB})$ to $(22 \mathrm{~Hz},-42 \mathrm{~dB})$, when excitation frequency varies from 45 Hz to 20 Hz . An expert inspection of
these spectrums reveals that $f_{b}=(1 \pm 2 s) f_{s}$ are always located around the excitation frequency and the distance from the excitation frequency is $2 s f_{s}$. It also can be seen from Figure 9 that the amplitude of left sideband harmonic component (12s) $f_{s}$ decreases with the excitation frequency, which is the result that when excitation frequency decreases the output voltage becomes lower and much smaller current is drawn by motor.
5.3. Experiment 3. In Experiment 3, additional experiments were performed with the faulty motor at different excitation frequencies ( $32 \mathrm{~Hz}, 50 \mathrm{~Hz}$ ) under two different loads ( $25 \%$, $50 \%$ ) to assess the performance of the proposed method over the full range of motor loads. Using the proposed diagnosis method, the collected phase current data of the fault motor with three broken bars were analyzed. Figures 10(a) and 10(b) show the results with 32 Hz excitation frequency at different loads and it can be seen that the fault characteristic frequency of broken rotor bars is exposed clearly. With the increase of load, the components of characteristic frequency become more and more significant, such as in Figure 10(a) of $25 \%$ load and Figure 10 (b) of $50 \%$ load. The similar results could be received for fault motor with 50 Hz excitation frequency at different loads; see Figure 10(c) of 25\% load and Figure 10(d) of $50 \%$ load.

## 6. Conclusion

Open-loop voltage-source-inverter-fed squirrel-cage motors with constant voltage/frequency control method are widely used to drive fans, pumps, or other mechanisms involving speed control for energy-saving purpose. In these cases, the motor operates steadily with different excitation frequencies. Unlike the utility-driven case, the position of the current harmonics appearing on the stator-current spectrum due to broken rotor bar faults is highly dependent on the mechanical motor load and excitation frequency, which affects the slip frequency. As a consequence, the reliable identification and isolation of faults remains an open issue.

In this paper, a simplified fault current signal model is firstly established using switching function concept and frequency modulation theory. It is demonstrated that the inverter-fed motor current is heavily contaminated due to PWM switching of the voltage source inverter. However, the broken rotor bars fault characteristic frequency $f_{b}=$ $(1 \pm 2 s) f_{s}$, depending on faulty severity, loads, and excitation frequency, is always located around the excitation frequency and the distance from the excitation frequency is $2 s f_{s}$. Novel broken rotor bar fault diagnosis techniques using motor current signature analysis (MCSA) for open-loop voltage-source-inverter-fed squirrel-cage induction motors with constant voltage/frequency control method are subsequently proposed. Experimental results obtained on self-made 3 kW three-phase squirrel-cage induction motors are discussed. It is shown that experimental and simulation results are consistent with those of the model revealed and the proposed techniques are effective and accurate.

The method described works well under constant load torque, but some difficulties appear with regard to closedloop control-operated machines, when $f_{s}$ and $s$ vary almost simultaneously and it is impossible to employ the proposed method to diagnose broken rotor bar fault. At the moment, further research is carried out for the features, advantages, limitations, and improvements of the proposed techniques.

## Acknowledgment

This work was supported by the Chinese academy of Sciences Key Deployment Project under Grant no. KGZD-EW-302.

## References

[1] M. E. H. Benbouzid and G. B. Kliman, "What stator current processing-based technique to use for induction motor rotor faults diagnosis?" IEEE Transactions on Energy Conversion, vol. 18, no. 2, pp. 238-244, 2003.
[2] J. Cusido, L. Romeral, J. A. Ortega, J. A. Rosero, and A. G. Espinosa, "Fault detection in induction machines using power spectral density in wavelet decomposition," IEEE Transactions on Industrial Electronics, vol. 55, no. 2, pp. 633-643, 2008.
[3] T. M. Wolbank, P. Nussbaumer, H. Chen, and P. E. Macheiner, "Monitoring of rotor-bar defects in inverter-fed induction machines at zero load and speed," IEEE Transactions on Industrial Electronics, vol. 58, no. 5, pp. 1468-1478, 2011.
[4] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, "Advances in diagnostic techniques for induction machines," IEEE Transactions on Industrial Electronics, vol. 55, no. 12, pp. 4109-4126, 2008.
[5] A. H. Bonnett and C. Yung, "Increased efficiency versus increased reliability", IEEE Industry Applications Magazine, vol. 14, no. 1, pp. 29-36, 2008.
[6] J. Cusido, L. Romeral, J. A. Ortega, A. Garcia, and J. R. Riba, "Wavelet and PDD as fault detection techniques," Electric Power Systems Research, vol. 80, no. 8, pp. 915-924, 2010.
[7] I. Culbert and W. Rhodes, "Using current signature analysis technology to reliably detect cage winding defects in squirrel cage induction motors," in Proceedings of the 52nd Annual Petroleum and Chemical Industry Conference, pp. 95101, September 2005.
[8] J. H. Jung, J. J. Lee, and B. H. Kwon, "Online diagnosis of induction motors using MCSA," IEEE Transactions on Industrial Electronics, vol. 53, no. 6, pp. 1842-1852, 2006.
[9] S. H. Kia, H. Henao, and G. A. Capolino, "A high-resolution frequency estimation method for three-phase induction machine fault detection," IEEE Transactions on Industrial Electronics, vol. 54, no. 4, pp. 2305-2314, 2007.
[10] D. Matic, F. Kulić, M. Pineda-Sánchez, and I. Kamenko, "Support vector machine classifier for diagnosis in electrical machines: application to broken bar," Expert Systems with Applications, vol. 39, no. 10, pp. 8681-8689, 2012.
[11] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," IEEE Transactions on Industrial Electronics, vol. 47, no. 5, pp. 984-993, 2000.
[12] M. F. Cabanas, F. Pedrayes, M. G. Melero et al., "Unambiguous detection of broken bars in asynchronous motors by means of a flux measurement-based procedure," IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 3, pp. 891-899, 2011.
[13] R. A. Gupta, A. K. Wadhwani, and S. R. Kapoor, "Early estimation of faults in induction motors using symbolic dynamicbased analysis of stator current samples," IEEE Transactions on Energy Conversion, vol. 26, no. 1, pp. 102-114, 2011.
[14] W. T. Thomson and M. Fenger, "Case histories of current signature analysis to detect faults in induction motor drives," in Proceedings of the IEEE International Electric Machines and Drives Conference (IEMDC '03), 2003.
[15] S. Choi, B. Akin, M. M. Rahimian, and H. A. Toliyat, "Implementation of a fault-diagnosis algorithm for induction machines based on advanced digital-signal-processing techniques," IEEE Transactions on Industrial Electronics, vol. 58, no. 3, pp. 937-948, 2011.
[16] B. Akin, S. B. Ozturk, H. A. Toliyat, and M. Rayner, "DSPbased sensorless electric motor fault diagnosis tools for electric and hybrid electric vehicle powertrain applications," IEEE Transactions on Vehicular Technology, vol. 58, no. 5, pp. 21502159, 2009.
[17] A. Bellini, "Closed-loop control impact on the diagnosis of induction motors faults," IEEE Transactions on Industry Applications, vol. 36, no. 5, pp. 1318-1329, 2000.
[18] S. M. A. Cruz and A. J. M. Cardoso, "Diagnosis of rotor faults in closed-loop induction motor drives," in Proceedings of the IEEE 41st IAS Annual Meeting on Industry Applications, pp. 23462353, October 2006.
[19] B. K. Lee and M. Ehsani, "A simplified functional simulation model for three-phase voltage-source inverter using switching function concept," IEEE Transactions on Industrial Electronics, vol. 48, no. 2, pp. 309-321, 2001.
[20] G. W. Chang and S. K. Chen, "An analytical approach for characterizing harmonic and interharmonic currents generated by VSI-fed adjustable speed drives," IEEE Transactions on Power Delivery, vol. 20, no. 4, pp. 2585-2593, 2005.
[21] S. Nandi, H. A. Toliyat, and X. D. Li, "Condition monitoring and fault diagnosis of electrical motors-a review," IEEE Transactions on Energy Conversion, vol. 20, no. 4, pp. 719-729, 2005.
[22] M. Eltabach and A. Charara, "Comparative investigation of electric signal analyses methods for mechanical fault detection in induction motors," Electric Power Components and Systems, vol. 35, no. 10, pp. 1161-1180, 2007.
[23] M. Eltabach, J. Antoni, and M. Najjar, "Quantitative analysis of noninvasive diagnostic procedures for induction motor drives," Mechanical Systems and Signal Processing, vol. 21, no. 7, pp. 2838-2856, 2007.
[24] A. Yazidi, H. Henao, G. A. Capolino, M. Artioli, and F. Filippetti, "Improvement of frequency resolution for threephase induction machine fault diagnosis," in Proceedings of the IEEE 40th IAS Annual Meeting on Industry Applications, vol. 1, pp. 20-25, October 2005.
[25] I. P. Georgakopoulos, E. D. Mitronikas, and A. N. Safacas, "Detection of induction motor faults in inverter drives using inverter input current analysis," IEEE Transactions on Industrial Electronics, vol. 58, no. 9, pp. 4365-4373, 2011.
[26] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, Discrete-Time Signal Processing, Tsinghua University Press, Beijing, China, 2nd edition, 2005.
[27] J. Faiz and B. M. Ebrahimi, "A new pattern for detecting broken rotor bars in induction motors during start-up," IEEE Transactions on Magnetics, vol. 44, no. 12, pp. 4673-4683, 2008.
[28] B. Ayhan, M. Y. Chow, and M. H. Song, "Multiple discriminant analysis and neural-network-based monolith and partition fault-detection schemes for broken rotor bar in induction motors," IEEE Transactions on Industrial Electronics, vol. 53, no. 4, pp. 1298-1308, 2006.

## Research Article

# Multiple-Model Cardinality Balanced Multitarget Multi-Bernoulli Filter for Tracking Maneuvering Targets 

Xianghui Yuan, Feng Lian, and Chongzhao Han<br>Ministry of Education Key Laboratory for Intelligent Networks and Network Security (MOE KLINNS), School of Electronics and Information Engineering, Xian Jiaotong University, Xian 710049, China<br>Correspondence should be addressed to Feng Lian; lianfeng1981@mail.xjtu.edu.cn

Received 10 July 2013; Accepted 15 September 2013
Academic Editor: Xianxia Zhang
Copyright © 2013 Xianghui Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

By integrating the cardinality balanced multitarget multi-Bernoulli (CBMeMBer) filter with the interacting multiple models (IMM) algorithm, an MM-CBMeMBer filter is proposed in this paper for tracking multiple maneuvering targets in clutter. The sequential Monte Carlo (SMC) method is used to implement the filter for generic multi-target models and the Gaussian mixture (GM) method is used to implement the filter for linear-Gaussian multi-target models. Then, the extended Kalman (EK) and unscented Kalman filtering approximations for the GM-MM-CBMeMBer filter to accommodate mildly nonlinear models are described briefly. Simulation results are presented to show the effectiveness of the proposed filter.

## 1. Introduction

Recently, the random-finite-set-(RFS-) based multitarget tracking approaches [1] have attracted extensive attention. Although theoretically solid, the RFS-based approaches usually are involved with intractable computations. By introducing the finite-set statistics (FISST) [2], Mahler developed the probability hypothesis density (PHD) [3] and cardinalized PHD (CPHD) [4] filters, which have been shown to be a computationally tractable alternative to full multitarget Bayes filters in the RFS framework. The sequential Monte Carlo (SMC) implementations for the PHD and CPHD filters were devised by Zajic and Mahler [5], Sidenbladh [6], and Vo et al. [7]. Vo et al. and Zhang et al. [8-10] devised the Gaussian mixture (GM) implementations for the PHD and CPHD filters under the linear-Gaussian assumption on target dynamics, birth process, and sensor model. The PHD-based approaches have been successfully used for many real-world problems [11-13]. However, the SMC-PHD and SMC-CPHD approaches require clustering to extract state estimates from the particle population, which is expensive and unreliable [14, 15].

In 2007, Mahler proposed the multitarget multi-Bernoulli (MeMBer) [2] recursion, which is an approximation to the full multitarget Bayes recursion using multi-Bernoulli RFSs
under low clutter density scenarios. In 2009, Vo et al. showed that the MeMBer filter overestimates the number of targets and proposed a cardinality-balanced MeMBer (CBMeMBer) filter [16] to reduce the cardinality bias. Then, the SMC and GM implementations for the MeMBer and CBMeMBer filters were, respectively, proposed for generic and linearGaussian dynamic and measurement models. The MeMBer and CBMeMBer recursions propagate not the moments and cardinality distributions which are propagated by the PHD and CPHD filters but rather the approximate multitarget multi-Bernoulli posterior density. Therefore, the key advantage of the SMC-CBMeMBer filter over the SMC-PHD and SMC-CPHD filters is that the multi-Bernoulli representation of the posterior density allows reliable and inexpensive extraction of state estimates. The CBMeMBer filter has been applied for tracking multiple targets according to their audio and visual information [17].

The original CBMeMBer filter does not consider the target maneuvers. Maneuvering targets might switch between different models of operation, so tracking using a singlemodel CBMeMBer filter might fail since the filter does not match the actual system dynamics. It is well known that the interacting multiple models (IMM) approaches [18] have been proven to be very effective and have better performance
than the single-model filters in tracking a single maneuvering target without clutter. In the IMM approaches, a bank of filters, each matched with a different target motion model, operate in parallel. In general, there are three key steps in the IMM estimators: (1) mixing the model-conditioned estimates; (2) model-conditioned base-state estimation; (3) deriving the overall state estimate by combining the estimates from each model-conditioned base-state filters.

By integrating the CBMeMBer filter with the IMM algorithm, an MM-CBMeMBer filter is proposed to address the problem of tracking multiple maneuvering targets in clutter, which is much more difficult than the problem of tracking a single maneuvering target without clutter since the association between the measurements and the targets is unknown. The SMC method is used to implement the filter for generic multitarget models while the GM method is used to implement the filter for linear-Gaussian multitarget models. Then, the extended Kalman (EK) [19] and unscented Kalman (UK) [20] filtering approximations for the GM-MM-CBMeMBer filter to accommodate mildly nonlinear models are described briefly. Nonlinear and linear-Gaussian examples of multiple maneuvering targets tracking are, respectively, presented for comparing the performance of the MM-CBMeMBer filter with that of the single-model CBMeMBer filters, MM-PHD filter [21-24], and MM-CPHD filter [25]. The simulation results show that (1) the proposed filter can estimate the number and states of multiple maneuvering targets effectively, whereas the performance of the single-model CBMeMBer filters is rather poor; (2) under relatively low clutter density, the SMC-MM-CBMeMBer filter outperforms the SMC-MMPHD and SMC-MM-CPHD filters; (3) the performance of the GM-MM-CBMeMBer filter is similar to that of the GM-MMPHD filter and hence is inferior to that of GM-MM-CPHD filter.

The rest of the paper is organized as follows. Section 2 describes the problem of multiple maneuvering targets tracking. In Section 3, the MM-CBMeMBer recursion is given. The generic SMC implementation of the MM-CBMeMBer filter is described in Section 4. The analytic GM implementation of the MM-CBMeMBer filter for linear-Gaussian multitarget models and its EK and UK extensions for nonlinear multitarget models are, respectively, given in Section 5. Numerical studies are shown in Section 6. The conclusions and the future work are given in Section 7.

## 2. Problem Statement for Multiple Maneuvering Targets Tracking

The multiple maneuvering targets appear and disappear randomly against time over an observation region. At time $k$, let $\mathbf{x}_{k} \in \mathbb{R}^{n}$ denote the kinematical state of a target and $n_{k} \in \mathbb{N}$ the label of the model in effect, where $\mathbb{N}$ is the discrete set of all model labels. The models follow a discrete Markov chain with transition probability $h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)$. Let $\mathbf{y}_{k}=$ $\left(\mathbf{x}_{k}, n_{k}\right) \in \mathbb{R}^{n} \times \mathbb{N}$ denote the augmented state vector, whose transition is governed by the density

$$
\begin{equation*}
f_{k \mid k-1}\left(\mathbf{y}_{k} \mid \mathbf{y}_{k-1}\right)=f_{k \mid k-1}\left(\mathbf{x}_{k} \mid \mathbf{x}_{k-1}, n_{k}\right) h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right) \tag{1}
\end{equation*}
$$

where $f_{k \mid k-1}\left(\mathbf{x}_{k} \mid \mathbf{x}_{k-1}, n_{k}\right)$ is the kinematical state transition density conditioned on model $n_{k}$.

The measurement originates either from target or from random clutter (false alarm). Moreover, the target-generated measurements are indistinguishable from the clutter. At time $k$, let $\mathbf{z}_{k} \in \mathbb{R}^{m}$ denote the measurement vector received by a sensor. The single-measurement single-target likelihood is described by the density conditioned on model $n_{k}$

$$
\begin{equation*}
g_{k}\left(\mathbf{z}_{k} \mid \mathbf{y}_{k}\right)=f_{k \mid k}\left(\mathbf{z}_{k} \mid \mathbf{x}_{k}, n_{k}\right) \tag{2}
\end{equation*}
$$

At time $k$, let $T_{k}$ denote the number of the existing targets and $S_{k}$ the number of the measurements. Then, multiple augmented states and unlabelled sensor measurements can be represented as finite sets $Y_{k}=\left\{\left(\mathbf{x}_{k}^{(i)}, n_{k}^{(i)}\right)\right\}_{i=1}^{T_{k}}$ and $Z_{k}=$ $\left\{\mathbf{z}_{k}^{(s)}\right\}_{s=1}^{S_{k}}$, respectively. In addition, let $Z_{1: k} \triangleq Z_{1}, \ldots, Z_{k}$ denote a sequence of the measurement sets available up to and including time $k$.

## 3. MM-CBMeMBer Filter

A Bernoulli RFS $Y_{k}$ has probability $1-r_{k}$ of being empty and probability $r_{k}\left(0 \leq r_{k} \leq 1\right)$ of being a singleton whose only element is distributed according to a probability density $p_{k}$. The probability density of $Y_{k}$ is

$$
\pi\left(Y_{k}\right)= \begin{cases}1-r_{k}, & Y_{k}=\emptyset  \tag{3}\\ r_{k} p_{k}\left(\mathbf{x}_{k}, n_{k}\right), & Y_{k}=\left\{\left(\mathbf{x}_{k}, n_{k}\right)\right\}\end{cases}
$$

A multi-Bernoulli RFS $Y_{k}$ is a union of a fixed number of independent Bernoulli RFSs $Y_{k}^{(i)}, i=1, \ldots, M_{k}$, that is, $Y_{k}=\bigcup_{i=1}^{M_{k}} Y_{k}^{(i)} . Y_{k}$ is thus completely described by the multiBernoulli parameter set $\left\{\left(r_{k}^{(i)}, p_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k}}$ with the mean cardinality $\sum_{i=1}^{M_{k}} r_{k}^{(i)}$ and the probability density [2]

$$
\begin{equation*}
\pi\left(Y_{k}\right)=\prod_{j=1}^{M_{k}}\left(1-r_{k}^{(j)}\right) \sum_{1 \leq i_{1} \neq \cdots \neq i_{\left.\right|_{k} \mid} \leq M_{k}} \prod_{j=1}^{\left|Y_{k}\right|} \frac{r_{k}^{\left(i_{j}\right)} p_{k}^{\left(i_{j}\right)}\left(\mathbf{x}_{k}, n_{k}\right)}{1-r_{k}^{\left(i_{j}\right)}}, \tag{4}
\end{equation*}
$$

where $|\cdot|$ denotes the cardinality of a set.
Throughout this paper, we abbreviate a probability density of the form (4) by $\pi\left(Y_{k}\right)=\left\{\left(r_{k}^{(i)}, p_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k}}$.

Let $p_{S, k}\left(\mathbf{y}_{k-1}\right)$ denote the probability that the maneuvering target with augmented state $\mathbf{y}_{k-1}$ survives at time $k$; let $p_{D, k}\left(\mathbf{y}_{k}\right)$ denote the probability that the maneuvering target with augmented state $\mathbf{y}_{k}$ generates an observation at time $k$. RFS modeling the multiple maneuvering targets state $Y_{k}$ and the sensor measurement $Z_{k}$ are, respectively, given by the union

$$
\begin{gather*}
Y_{k}=\left[\bigcup_{\mathbf{y}_{k-1} \in Y_{k-1}} \Omega_{k \mid k-1}\left(\mathbf{y}_{k-1}\right)\right] \cup \Gamma_{k},  \tag{5}\\
Z_{k}=\left[\bigcup_{\mathbf{y}_{k} \in Y_{k}} \Theta_{k}\left(\mathbf{y}_{k}\right)\right] \cup K_{k},
\end{gather*}
$$

where $\Gamma_{k}$ denotes the multi-Bernoulli RFS of spontaneous births; the Bernoulli RFS $\Omega_{k \mid k-1}\left(\mathbf{y}_{k-1}\right)$ with $r_{k}=p_{S, k}\left(\mathbf{y}_{k-1}\right)$ and $p_{k}\left(\mathbf{y}_{k}\right)=f_{k \mid k-1}\left(\mathbf{y}_{k} \mid \mathbf{y}_{k-1}\right)$ is used to model the dynamic behavior of $\mathbf{y}_{k-1} \in Y_{k-1}$; the Bernoulli RFS $\Theta_{k}\left(\mathbf{y}_{k}\right)$ with $r_{k}=p_{D, k}\left(\mathbf{y}_{k}\right)$ and $p_{k}\left(\mathbf{z}_{k}\right)=g_{k}\left(\mathbf{z}_{k} \mid \mathbf{y}_{k}\right)$ is used to model the observation behavior of $\mathbf{y}_{k} \in Y_{k}$; the clutter is modeled as a Poisson RFS $K_{k}$ with the intensity $\kappa_{k}\left(\mathbf{z}_{k}\right)=\lambda_{c, k} f_{c, k}\left(\mathbf{z}_{k}\right)$, where $\lambda_{c, k}$ and $f_{c, k}(\cdot)$ are, respectively, the average clutter number and the probability density of clutter spatial distribution at time $k$.

Based on the above RFS models of the multiple maneuvering targets and the method of Mahler's FISST, the MMCBMeMBer filter, which implicitly requires a finite number of single-model CBMeMBer filters operate in parallel, is derived by introducing the mixing and combination strategies in the IMM approaches [18]. As the multiple-model approaches, the MM-CBMeMBer filter does not need a maneuver detection decision and undergoes a soft switching between the models. One cycle of the recursive MM-CBMeMBer algorithm can be described as follows.
(1) The Mixing and Prediction Stage. If at time $k-1$, the posterior density is a multi-Bernoulli of the form $\pi_{k-1}\left(Y_{k-1} \mid\right.$ $\left.Z_{1: k-1}\right)=\left\{\left(r_{k-1}^{(i)}, p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right)\right)\right\}_{i=1}^{M_{k-1}}$, then the mixed multiBernoulli density is

$$
\begin{equation*}
\hat{\pi}_{k-1}\left(\hat{Y}_{k-1} \mid Z_{1: k-1}\right)=\left\{\left(r_{k-1}^{(i)}, p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k}\right)\right)\right\}_{i=1}^{M_{k-1}} \tag{6}
\end{equation*}
$$

where

$$
\begin{align*}
p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k}\right) & =\sum_{n_{k-1}} p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k}, n_{k-1}\right) \\
& =\sum_{n_{k-1}} p_{k-1}^{(i)}\left(n_{k} \mid \mathbf{x}_{k-1}, n_{k-1}\right) p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right) . \tag{7}
\end{align*}
$$

Since the models switching is only decided by the model transition probability and is independent of the target kinematical state:

$$
\begin{equation*}
=\sum_{n_{k-1}} h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right) p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right) \tag{8}
\end{equation*}
$$

is a combination of the previous model-dependent densities. Finally, the mixed and predicted density is also a multiBernoulli and is given by

$$
\begin{aligned}
\pi_{k \mid k-1} & \left(Y_{k} \mid Z_{1: k-1}\right) \\
= & \left\{\left(r_{P, k \mid k-1}^{(i)}, p_{P, k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k-1}} \\
& \cup\left\{\left(r_{\Gamma, k}^{(i)}, p_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{\mathrm{r}, k}},
\end{aligned}
$$

where $\left\{\left(r_{\Gamma, k}^{(i)}, p_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{\Gamma, k}}$ are the parameters of the multiBernoulli RFS of births at time $k$ :

$$
\begin{align*}
& r_{P, k \mid k-1}^{(i)} \\
& \qquad \begin{array}{l}
=r_{k-1}^{(i)} \sum_{n_{k}} \sum_{n_{k-1}} h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right) \\
\\
\quad \times\left\langle p_{k-1}^{(i)}\left(\cdot, n_{k-1}\right), p_{S, k}\left(\cdot, n_{k-1}\right)\right\rangle, \\
\begin{aligned}
p_{P, k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)
\end{aligned} \\
=\left(\sum_{n_{k-1}} h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)\right. \\
\left.\quad \times\left\langle f_{k \mid k-1}\left(\mathbf{x}_{k} \mid \cdot, n_{k}\right), p_{k-1}^{(i)}\left(\cdot, n_{k-1}\right) p_{S, k}\left(\cdot, n_{k-1}\right)\right\rangle\right) \\
\times\left(\sum_{n_{k}} \sum_{k-1} h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)\right. \\
\quad
\end{array} \quad \begin{array}{l}
\left.\quad\left\langle p_{k-1}^{(i)}\left(\cdot, n_{k-1}\right), p_{S, k}\left(\cdot, n_{k-1}\right)\right\rangle\right)^{-1},
\end{array}
\end{align*}
$$

where $\langle\cdot, \cdot\rangle$ defines the integral inner product, that is,

$$
\begin{align*}
& \left\langle p_{k-1}^{(i)}\left(\cdot, n_{k-1}\right), p_{S, k}\left(\cdot, n_{k-1}\right)\right\rangle \\
& \quad=\int p_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right) p_{S, k}\left(\mathbf{x}_{k-1}, n_{k-1}\right) d \mathbf{x}_{k-1} . \tag{11}
\end{align*}
$$

(2) The Update Stage. If at time $k$, the mixed and predicted density is a multi-Bernoulli of the form $\pi_{k \mid k-1}\left(Y_{k} \mid Z_{1: k-1}\right)=$ $\left\{\left(r_{k \mid k-1}^{(i)}, p_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k \mid k-1}}$, then the posterior density can be approximated by a multi-Bernoulli as follows:

$$
\begin{align*}
\pi_{k}\left(Y_{k} \mid Z_{k}\right) \approx\{ & \left.\left\{r_{L, k}^{(i)}, p_{L, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k \mid k-1}}  \tag{12}\\
& \cup\left\{\left(r_{U, k}\left(\mathbf{z}_{k}\right), p_{U, k}\left(\mathbf{x}_{k}, n_{k} ; \mathbf{z}_{k}\right)\right)\right\}_{\mathbf{z}_{k} \in Z_{k}},
\end{align*}
$$

where

$$
r_{L, k}^{(i)}=r_{k \mid k-1}^{(i)} \frac{1-\sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), p_{D, k}\left(\cdot, n_{k}\right)\right\rangle}{1-r_{k \mid k-1}^{(i)} \sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), p_{D, k}\left(\cdot, n_{k}\right)\right\rangle},
$$

$$
\begin{gather*}
p_{L, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\frac{\left(1-p_{D, k}\left(\mathbf{x}_{k}, n_{k}\right)\right) p_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)}{1-\sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), p_{D, k}\left(\cdot, n_{k}\right)\right\rangle}, \\
r_{U, k}\left(\mathbf{z}_{k}\right)=\left(\sum_{i=1}^{M_{k \mid k-1}} \frac{\left(1-r_{k \mid k-1}^{(i)}\right) r_{k \mid k-1}^{(i)} \sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), g_{k}\left(\mathbf{z}_{k} \mid \cdot, n_{k}\right) p_{D, k}\left(\cdot, n_{k}\right)\right\rangle}{\left(1-r_{k \mid k-1}^{(i)} \sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), p_{D, k}\left(\cdot, n_{k}\right)\right\rangle\right)^{2}}\right) \\
\times\left(\kappa_{k}\left(\mathbf{z}_{k}\right)+\sum_{i=1}^{M_{k \mid k-1}} \frac{r_{k \mid k-1}^{(i)} \sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), g_{k}\left(\mathbf{z}_{k} \mid \cdot, n_{k}\right) p_{D, k}\left(\cdot, n_{k}\right)\right\rangle}{1-r_{k \mid k-1}^{(i)} \sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), p_{D, k}\left(\cdot, n_{k}\right)\right\rangle}\right)^{-1}, \\
p_{U, k}\left(\mathbf{x}_{k}, n_{k} ; \mathbf{z}_{k}\right)=  \tag{13}\\
\sum_{i=1}^{M_{k \mid k-1}}\left(r_{k \mid k-1}^{(i)} /\left(1-r_{k \mid k-1}^{(i)}\right)\right) p_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right) g_{k}\left(\mathbf{z}_{k} \mid \mathbf{x}_{k}, n_{k}\right) p_{D, k}\left(\mathbf{x}_{k}, n_{k}\right) \\
M_{i=1}^{(i)}\left(r_{k \mid k-1}^{(i)} /\left(1-r_{k \mid k-1}^{(i)}\right)\right) \sum_{n_{k}}\left\langle p_{k \mid k-1}^{(i)}\left(\cdot, n_{k}\right), g_{k}\left(\mathbf{z}_{k} \mid \cdot, n_{k}\right) p_{D, k}\left(\cdot, n_{k}\right)\right\rangle
\end{gather*} .
$$

(3) The Multitarget State Estimation. For the multi-Bernoulli representation $\pi_{k}\left(Y_{k} \mid Z_{k}\right)=\left\{\left(r_{k}^{(i)}, p_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k}}$, the extraction of multitarget number and state estimates are straightforward since the probability $r_{k}^{(i)}$ indicates how likely the $i$ th hypothesized track is a true track, and the posterior density $p_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)$ describes the distribution of the estimated augmented state of the track. The state estimation procedure for the MM-CBMeMBer filter [8] is summarized in Algorithm 1.

## 4. SMC-MM-CBMeMBer Filter

In this section, a generic SMC implementation of the proposed MM-CBMeMBer filter is presented for accommodating nonlinear dynamic and measurement models. In this implementation, the samples or particles, which are used to represent the multi-Bernoulli density of multiple maneuvering targets, consists of the kinematical state and model information with associated weights. One cycle of the recursive SMC-MM-CBMeMBer algorithm can be described as follows.
(1) The SMC Mixing and Prediction Stage. Suppose that at time $k-1$ the multi-Bernoulli posterior density $\widetilde{\pi}_{k-1}\left(Y_{k-1} \mid\right.$ $\left.Z_{1: k-1}\right)=\left\{\left(\widetilde{r}_{k-1}^{(i)}, \widetilde{p}_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right)\right)\right\}_{i=1}^{M_{k-1}}$ is given and each $\tilde{p}_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right), i=1, \ldots, M_{k-1}$, is composed of a set of weighted samples $\left\{\omega_{k-1}^{(i, l)}, \mathbf{x}_{k-1}^{(i, l)}, n_{k-1}^{(i, l)}\right\}_{l=1}^{L_{k-1}^{(i)}}$,

$$
\begin{equation*}
\tilde{p}_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right)=\sum_{l=1}^{L_{k-1}^{(i)}} \omega_{k-1}^{(i, l)} \delta\left(\mathbf{x}_{k-1}-\mathbf{x}_{k-1}^{(i, l)}, n_{k-1}-n_{k-1}^{(i, l)}\right) \tag{14}
\end{equation*}
$$

where $\delta\left(\mathbf{x}-\mathbf{x}^{(c)}\right)$ is Dirac delta function centered at $\mathbf{x}^{(c)}$. Then, the mixed and predicted multi-Bernoulli density $\tilde{\pi}_{k \mid k-1}\left(Y_{k} \quad \mid \quad Z_{1: k-1}\right)=\left\{\left(\widetilde{r}_{P, k \mid k-1}^{(i)}, \widetilde{p}_{P, k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k-1}} \cup$ $\left\{\left(\widetilde{r}_{\Gamma, k}^{(i)}, \widetilde{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{\mathrm{r}, k}}$ can be computed as follows:

$$
\begin{gather*}
\widetilde{r}_{P, k \mid k-1}^{(i)}=\widetilde{r}_{k-1}^{(i)} \sum_{l=1}^{L_{k-1}^{(i)}} \omega_{k-1}^{(i, l)} p_{S, k}\left(\mathbf{x}_{k-1}^{(i, l)}, n_{k-1}^{(i, l)}\right) \frac{h_{k \mid k-1}\left(n_{P, k \mid k-1}^{(i, l)} \mid n_{k-1}^{(i, l)}\right)}{\alpha_{k}^{(i)}\left(n_{P, k \mid k-1}^{(i, l} \mid n_{k-1}^{(i, l)}\right)}, \\
\widetilde{r}_{\Gamma, k}^{(i)}=\text { parameter given by birth model, } \\
\widetilde{p}_{P, k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\sum_{l=1}^{L_{k-1}^{(i)}} \omega_{P, k \mid k-1}^{(i, l)} \delta\left(\mathbf{x}_{k}-\mathbf{x}_{P, k \mid k-1}^{(i, l)}, n_{k}-n_{P, k \mid k-1}^{(i, l)}\right), \\
\widetilde{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\sum_{l=1}^{L_{\Gamma, k}^{(i)}} \omega_{\Gamma, k}^{(i, l)} \delta\left(\mathbf{x}_{k}-\mathbf{x}_{\Gamma, k}^{(i, l)}, n_{k}-n_{\Gamma, k}^{(i, l)}\right), \tag{15}
\end{gather*}
$$

where the particles $\mathbf{x}_{P, k \mid k-1}^{(i, l)}, n_{P, k \mid k-1}^{(i, l)}$ corresponding to the surviving maneuvering targets can be derived by sampling from the proposal densities $q_{k}^{(i)}\left(\cdot \mid \mathbf{x}_{k-1}, n_{k}, Z_{k}\right)$ and $\alpha_{k}^{(i)}\left(\cdot \mid n_{k-1}\right)$

$$
\begin{align*}
n_{P, k \mid k-1}^{(i, l)} & \sim \alpha_{k}^{(i)}\left(\cdot \mid n_{k-1}^{(i, l)}\right) \\
\mathbf{x}_{P, k \mid k-1}^{(i, l)} & \sim q_{k}^{(i)}\left(\cdot \mid \mathbf{x}_{k-1}^{(i, l)}, n_{P, k \mid k-1}^{(i, l)}, Z_{k}\right) \tag{16}
\end{align*} \quad l=1, \ldots, L_{k-1}^{(i)}
$$

with the associated weights

$$
\begin{gather*}
\omega_{P, k \mid k-1}^{(i, l)}=\frac{\breve{\omega}_{P, k \mid k-1}^{(i, l)}}{\sum_{l=1}^{L_{k-1}^{(i)}} \breve{\omega}_{P, k \mid k-1}^{(i, l)}}, \\
\breve{\omega}_{P, k \mid k-1}^{(i, l)}=\frac{f_{k \mid k-1}\left(\mathbf{x}_{P,(\mid k-1}^{(i, l)} \mid \mathbf{x}_{k-1}^{(i, l)}, n_{P P, k \mid k-1}^{(i, l)}\right) p_{S, k}\left(\mathbf{x}_{k-1}^{(i, l)}, n_{k-1}^{(i, l)}\right)}{q_{k}^{(i)}\left(\mathbf{x}_{P, k \mid k-1}^{(i, l)} \mid \mathbf{x}_{k-1}^{(i, l)}, n_{P, k \mid k-1}^{(i, l)}, Z_{k}\right)} \\
\cdot  \tag{17}\\
\cdot \frac{h_{k \mid k-1}^{(i)}\left(n_{P, k \mid k-1}^{(i, l)} \mid n_{k-1}^{(i, l)}\right)}{\alpha_{k}^{(i)}\left(n_{P, k \mid k-1}^{(i, l)} \mid n_{k-1}^{(i, l)}\right)} \cdot \omega_{k-1}^{(i, l)}
\end{gather*}
$$

and the particles $\mathbf{x}_{\Gamma, k}^{(i, l)}, n_{\Gamma, k}^{(i, l)}$ corresponding to the new born maneuvering targets can be derived by sampling from the proposal densities $b_{k}^{(i)}\left(\cdot \mid n_{k}, Z_{k}\right)$ and $\beta_{k}^{(i)}(\cdot)$

$$
\begin{array}{cc}
n_{\Gamma, k}^{(i, l)} \sim \beta_{k}^{(i)}(\cdot) & l=1, \ldots, L_{\Gamma, k}^{(i)} \\
\mathbf{x}_{\Gamma, k}^{(i, l)} \sim b_{k}^{(i)}\left(\cdot \mid n_{\Gamma, k}^{(i, l)}, Z_{k}\right) & \tag{18}
\end{array}
$$

$$
\begin{aligned}
& \text { given } \pi_{k}\left(Y_{k} \mid Z_{k}\right)=\left\{\left(r_{k}^{(i)}, p_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k}} ; \text { set } \widehat{T}_{k}=0, \widehat{X}_{k}=\varnothing \\
& \text { for } i=1, \ldots, M_{k} \text {, } \\
& \text { if } r_{k}^{(i)}>\text { a given threshold (i.e. 0.5); } \\
& \widehat{T}_{k}=\widehat{T}_{k}+1 \text {, } \\
& \widehat{\mathbf{x}}_{k}^{\left(\widehat{T}_{k}\right)}=\sum_{n k} \overbrace{p_{k}^{(i)}\left(n_{k}\right)}^{\text {probability of model } n_{k}} \cdot \overbrace{\int \mathbf{x}_{k} p_{k}^{(i)}\left(\mathbf{x}_{k} \mid n_{k}\right) d \mathbf{x}_{k}}^{\text {state e esimation conditioned on model } n_{k}}=\sum_{n k} \int \mathbf{x}_{k} p_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right) d \mathbf{x}_{k} \text {, } \\
& \widehat{X}_{k}:=\left[\widehat{X}_{k}, \widehat{\mathbf{x}}_{k}^{\left(\hat{T}_{k}\right)}\right] ; \\
& \text { end; } \\
& \text { end; } \\
& \text { output: } \widehat{X}_{k}=\left\{\widehat{\mathbf{x}}_{k}^{(i)}\right\}_{i=1}^{\widehat{T}_{k}}
\end{aligned}
$$

Algorithm 1: Multitarget state estimation procedure for the MM-CBMeMBer filter.
with the associated weights

$$
\begin{gather*}
\omega_{\Gamma, k}^{(i, l)}=\frac{\breve{\omega}_{\Gamma, k}^{(i, l)}}{\sum_{l=1}^{L_{\Gamma, k}^{(i)}} \breve{\omega}_{\Gamma, k}^{(i, l)}},  \tag{19}\\
\breve{\omega}_{\Gamma, k}^{(i, l)}=\frac{p_{\Gamma, k}^{(i)}\left(\mathbf{x}_{\Gamma, k}^{(i, l)}, n_{\Gamma, k}^{(i, l)}\right)}{b_{k}^{(i)}\left(\mathbf{x}_{\Gamma, k}^{(i, l)} \mid n_{\Gamma, k}^{(i, l)}, Z_{k}\right) \beta_{k}^{(i)}\left(n_{\Gamma, k}^{(i, l)}\right)} .
\end{gather*}
$$

(2) The SMC Update Stage. Suppose that at time $k$ the mixed and predicted multi-Bernoulli density $\tilde{\pi}_{k \mid k-1}\left(Y_{k} \mid Z_{1: k-1}\right)=$
$\left\{\left(\widetilde{r}_{k \mid k-1}^{(i)}, \widetilde{p}_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k \mid k-1}}$ is given and each $\widetilde{p}_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)$, $i=1, \ldots, M_{k \mid k-1}$, is composed of a set of weighted samples

$$
\begin{align*}
& \left\{\omega_{k \mid k-1}^{(i, l)}, \mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right\}_{l=1}^{L_{k \mid k-1}^{(i)}}, \\
& \quad \widetilde{p}_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} \delta\left(\mathbf{x}_{k}-\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k}-n_{k \mid k-1}^{(i, l)}\right) . \tag{20}
\end{align*}
$$

Then, the multi-Bernoulli approximation of the updated density $\widetilde{\pi}_{k}\left(Y_{k} \mid Z_{1: k}\right) \approx\left\{\left(\widetilde{r}_{L, k}^{(i)}, \widetilde{p}_{L, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k \mid k-1}} \cup\left\{\left(\widetilde{r}_{U, k}\left(\mathbf{z}_{k}\right)\right.\right.$, $\left.\left.\widetilde{p}_{U, k}\left(\mathbf{x}_{k}, n_{k} ; \mathbf{z}_{k}\right)\right)\right\}_{\mathbf{z}_{k} \in Z_{k}}$ can be computed as follows:

$$
\begin{align*}
& \widetilde{r}_{L, k}^{(i)}=\widetilde{r}_{k \mid k-1}^{(i)} \frac{1-\sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)}{1-\widetilde{r}_{k \mid k-1}^{(i)} \sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)}, \\
& \tilde{p}_{L, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{L, k}^{(i, l)} \delta\left(\mathbf{x}_{k}-\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k}-n_{k \mid k-1}^{(i, l)}\right), \\
& \tilde{r}_{U, k}\left(\mathbf{z}_{k}\right)=\left(\sum_{i=1}^{M_{k \mid k-1}} \frac{\left(1-\widetilde{r}_{k \mid k-1}^{(i)}\right) \widetilde{r}_{k \mid k-1}^{(i)} \sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} g_{k}\left(\mathbf{z}_{k} \mid \mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right) p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)}{\left(1-\widetilde{r}_{k \mid k-1}^{(i)} \sum_{l=1}^{L_{k l \mid c-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)\right)^{2}}\right)  \tag{21}\\
& \times\left(\kappa_{k}\left(\mathbf{z}_{k}\right)+\sum_{i=1}^{M_{k \mid k-1}} \frac{\widetilde{r}_{k \mid k-1}^{(i)} \sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} g_{k}\left(\mathbf{z}_{k} \mid \mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right) p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)}{1-\widetilde{r}_{k \mid k-1}^{(i)} \sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)}\right)^{-1}, \\
& \widetilde{P}_{U, k}\left(\mathbf{x}_{k}, n_{k} ; \mathbf{z}_{k}\right)=\sum_{i=1}^{M_{k \mid k-1}} \sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{U, k}^{(i, l)}\left(\mathbf{z}_{k}\right) \delta\left(\mathbf{x}_{k}-\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k}-n_{k \mid k-1}^{(i, l)}\right),
\end{align*}
$$

where

$$
\begin{gather*}
\omega_{L, k}^{(i, l)}=\frac{\breve{\omega}_{L, k}^{(i, l)}}{\sum_{l=1}^{L_{k=1}^{(i)}} \breve{\omega}_{L, k}^{(i, l)}}, \\
\breve{\omega}_{L, k}^{(i, l)}=\omega_{k \mid k-1}^{(i, l)} \frac{1-p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)}{1-\sum_{l=1}^{L_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)} \\
\omega_{U, k}^{(i, l)}\left(\mathbf{z}_{k}\right)=\frac{\breve{\omega}_{U, k}^{(i, l)}\left(\mathbf{z}_{k}\right)}{\sum_{i=1}^{M_{k \mid k-1}} \sum_{l=1}^{L_{k \mid k-1}^{(i)}} \breve{\omega}_{U, k}^{(i, l)}\left(\mathbf{z}_{k}\right)}  \tag{22}\\
\breve{\omega}_{U, k}^{(i, l)}\left(\mathbf{z}_{k}\right)=\frac{\widetilde{r}_{k \mid k-1}^{(i)}}{1-\widetilde{r}_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, l)} g_{k} \\
\times\left(\mathbf{z}_{k} \mid \mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right) p_{D, k}\left(\mathbf{x}_{k \mid k-1}^{(i, l)}, n_{k \mid k-1}^{(i, l)}\right)
\end{gather*}
$$

(3) The Resampling and Pruning Stage. It is the same as the resampling and pruning stage of the SMC-CBMeMBer filter [16].
(4) The SMC Multitarget State Estimation. Given the SMC multi-Bernoulli posterior density

$$
\begin{array}{r}
\widetilde{\pi}_{k}\left(Y_{k} \mid Z_{1: k}\right)=\left\{\left(\widetilde{r}_{k}^{(i)}, \widetilde{p}_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k}} \\
\text { with } \widetilde{p}_{k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\sum_{l=1}^{L_{k}^{(i)}} \omega_{k}^{(i, l)} \delta\left(\mathbf{x}_{k}-\mathbf{x}_{k}^{(i, l)}, n_{k}-n_{k}^{(i, l)}\right) \tag{23}
\end{array}
$$

from the method described in Algorithm 1, the SMC multitarget state estimation can be easily obtained as

$$
\begin{equation*}
\widehat{X}_{k}=\left\{\widehat{\mathbf{x}}_{k}^{(i)}\right\}_{i=1}^{\widehat{T}_{k}} \quad \text { with } \widehat{\mathbf{x}}_{k}^{(i)}=\sum_{l=1}^{L_{k}^{(i)}} \mathbf{x}_{k}^{(i, l)} \omega_{k}^{(i, l)}, i=1, \ldots, \widehat{T}_{k} \tag{24}
\end{equation*}
$$

Note that the MCMC move step [26] can be introduced for increasing the particle variety after the resample step without affecting the validity of the SMC approximation.

## 5. GM-MM-CBMeMBer Filter and Its EK and UK Extensions

An analytic solution to the MM-CBMeMBer recursion for linear-Gaussian multiple maneuvering targets models is presented in this section. The resulting filter propagates the GM multi-Bernoulli density against time. Some certain assumptions about the linear-Gaussian multiple maneuvering targets models are firstly summarized below.
(A) The dynamic and measurement models for the augmented state of each maneuvering target have the form

$$
\begin{align*}
& f_{k \mid k-1}\left(\mathbf{x}_{k}, n_{k} \mid \mathbf{x}_{k-1}, n_{k-1}\right) \\
& = \\
& \quad \mathcal{N}\left(\mathbf{x}_{k} ; F_{k}\left(n_{k}\right) \mathbf{x}_{k-1}, \mathrm{~A}_{k}\left(n_{k}\right) Q_{k}\left(n_{k}\right)\left(\mathrm{A}_{k}\left(n_{k}\right)\right)^{T}\right)  \tag{25}\\
& \quad \times h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right), \\
& g_{k}\left(\mathbf{z}_{k} \mid \mathbf{x}_{k}, n_{k}\right) \\
& = \\
& \\
& \mathcal{N}\left(\mathbf{z}_{k} ; H_{k}\left(n_{k}\right) \mathbf{x}_{k}, \mathrm{~B}_{k}\left(n_{k}\right) R_{k}\left(n_{k}\right)\left(\mathrm{B}_{k}\left(n_{k}\right)\right)^{T}\right),
\end{align*}
$$

where $\mathcal{N}(\cdot ; \mathbf{m}, P)$ denotes the density of Gaussian distribution with the mean $\mathbf{m}$ and covariance $P ; F_{k}\left(n_{k}\right), Q_{k}\left(n_{k}\right)$, and $\mathrm{A}_{k}\left(n_{k}\right)$ are, respectively, the kinematical state transition, process noise covariance, and process noise coefficient matrixes conditioned on model $n_{k} ; H_{k}\left(n_{k}\right), R_{k}\left(n_{k}\right)$, and $\mathrm{B}_{k}\left(n_{k}\right)$ are, respectively, the observation, observation noise covariance, and observation noise coefficient matrixes conditioned on model $n_{k}$.
(B) The probabilities of maneuvering target survival and maneuvering target detection are independent of the kinematical state:

$$
\begin{gather*}
p_{S, k}\left(\mathbf{x}_{k-1}, n_{k-1}\right)=p_{S, k}\left(n_{k-1}\right) \\
p_{D, k}\left(\mathbf{x}_{k}, n_{k}\right)=p_{D, k}\left(n_{k}\right) \tag{26}
\end{gather*}
$$

(C) The birth model for the maneuvering targets is a multi-Bernoulli with parameter set $\left\{\left(\bar{r}_{\Gamma, k}^{(i)}, \bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{\Gamma, k}}$, where $\bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right), i=1, \ldots, M_{\Gamma, k}$, are GM of the form

$$
\begin{align*}
\bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)= & \bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k} \mid n_{k}\right) h_{\Gamma, k}^{(i)}\left(n_{k}\right) \\
= & h_{\Gamma, k}^{(i)}\left(n_{k}\right) \sum_{j=1}^{J_{\Gamma, k}^{(i)}\left(n_{k}\right)} \omega_{\Gamma, k}^{(i, j)}\left(n_{k}\right) \\
& \times \mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{\Gamma, k}^{(i, j)}\left(n_{k}\right), P_{\Gamma, k}^{(i, j)}\left(n_{k}\right)\right), \tag{27}
\end{align*}
$$

where $h_{\Gamma, k}^{(i)}\left(n_{k}\right)$ is the distribution of model births and $\bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k} \mid\right.$ $n_{k}$ ) is the distribution of the birth kinematical state given model $n_{k} \cdot \bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k} \mid n_{k}\right)$ is GM of the form with the parameter $\operatorname{set}\left\{\varpi_{\Gamma, k}^{(i, j)}\left(n_{k}\right), \mathbf{m}_{\Gamma, k}^{(i, j)}\left(n_{k}\right), P_{\Gamma, k}^{(i, j)}\left(n_{k}\right)\right\}_{j=1}^{J_{\Gamma, k}^{(i)}\left(n_{k}\right)}$.

According to the above Assumptions A, B, and C, a closed form solution to the MM-CBMeMBer recursion, namely, the GM-MM-CBMeMBer filter, can be derived by applying the following two standard results for Gaussian functions:

$$
\begin{gather*}
\int \mathscr{N}\left(\mathbf{x} ; F \mathbf{x}^{\prime}, Q\right) \mathcal{N}\left(\mathbf{x}^{\prime} ; \mathbf{m}, P\right) d \mathbf{x}^{\prime}=\mathcal{N}\left(\mathbf{x} ; F \mathbf{m}, Q+F P F^{T}\right) \\
\mathscr{N}(\mathbf{z} ; H \mathbf{x}, R) \mathscr{N}(\mathbf{x} ; \mathbf{m}, P) \\
=\mathcal{N}\left(\mathbf{z} ; H \mathbf{m}, R+H P H^{T}\right) \mathcal{N}(\mathbf{x} ; \widehat{\mathbf{m}}, \overparen{P}) \tag{28}
\end{gather*}
$$

where

$$
\begin{gather*}
K=P H^{T}\left(H P H^{T}+R\right)^{-1} \\
\widehat{\mathbf{m}}=\mathbf{m}+K(\mathbf{z}-H \mathbf{m})  \tag{29}\\
\widehat{P}=(I-K H) P .
\end{gather*}
$$

One cycle of the recursive GM-MM-CBMeMBer algorithm can be described as follows.
(1) The GM Mixing and Prediction Stage. Suppose that at time $k-1$ the multi-Bernoulli posterior density $\bar{\pi}_{k-1}\left(Y_{k-1} \mid\right.$ $\left.Z_{1: k-1}\right)=\left\{\left(\bar{r}_{k-1}^{(i)}, \bar{p}_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right)\right)\right\}_{i=1}^{M_{k-1}}$ is given and each $\bar{p}_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right), i=1, \ldots, M_{k-1}$, is composed of GM of the form

$$
\begin{align*}
& \bar{p}_{k-1}^{(i)}\left(\mathbf{x}_{k-1}, n_{k-1}\right) \\
& =\sum_{j=1}^{J_{k-1}^{(i)}\left(n_{k-1}\right)} ब_{k-1}^{(i, j)}\left(n_{k-1}\right)  \tag{30}\\
& \\
& \quad \times \mathcal{N}\left(\mathbf{x}_{k-1} ; \mathbf{m}_{k-1}^{(i, j)}\left(n_{k-1}\right), P_{k-1}^{(i, j)}\left(n_{k-1}\right)\right)
\end{align*}
$$

Then, the mixed and predicted multi-Bernoulli density $\bar{\pi}_{k \mid k-1}\left(Y_{k} \quad \mid \quad Z_{1: k-1}\right)=\left\{\left(\bar{r}_{P, k \mid k-1}^{(i)}, \bar{p}_{P, k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k-1}} \cup$ $\left\{\left(\bar{r}_{\Gamma, k}^{(i)}, \bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{\Gamma, k}}$ can be computed as follows:

$$
\begin{aligned}
& \bar{r}_{P, k \mid k-1}^{(i)}=\bar{r}_{k-1}^{(i)} \sum_{n_{k}} \sum_{n_{k-1}} \sum_{j=1}^{J_{k-1}^{(i)}\left(n_{k-1}\right)} h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right) \\
& \times p_{S, k}\left(n_{k-1}\right) \omega_{k-1}^{(i, j)}\left(n_{k-1}\right) \\
& \bar{p}_{P, k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\sum_{n_{k-1}} \sum_{j=1}^{J_{k-1}^{(i)}\left(n_{k-1}\right)} \omega_{P, k \mid k-1}^{(i, j)}\left(n_{k}, n_{k-1}\right) \\
& \times \mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{P, k \mid k-1}^{(i, j)}\left(n_{k}, n_{k-1}\right)\right. \\
&\left.P_{P, k \mid k-1}^{(i, j)}\left(n_{k}, n_{k-1}\right)\right)
\end{aligned}
$$

$$
\begin{equation*}
\left\{\left(\bar{r}_{\Gamma, k}^{(i)}, \bar{p}_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{\Gamma, k}}=\text { given by the birth model (27) } \tag{31}
\end{equation*}
$$

where

$$
\begin{gather*}
\mathbf{m}_{P, k \mid k-1}^{(i, j)}\left(n_{k}, n_{k-1}\right)=F_{k}\left(n_{k}\right) \mathbf{m}_{k-1}^{(i, j)}\left(n_{k-1}\right), \\
P_{P, k \mid k-1}^{(i, j)}\left(n_{k}, n_{k-1}\right)=F_{k}\left(n_{k}\right) P_{k-1}^{(i, j)}\left(n_{k-1}\right)\left(F_{k}\left(n_{k}\right)\right)^{T} \\
+\mathrm{A}_{k}\left(n_{k}\right) Q_{k}\left(n_{k}\right)\left(\mathrm{A}_{k}\left(n_{k}\right)\right)^{T}, \\
\omega_{P, k \mid k-1}^{(i, j)}\left(n_{k}, n_{k-1}\right) \\
=\left(h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right) p_{S, k}\left(n_{k-1}\right) \omega_{k-1}^{(i, j)}\left(n_{k-1}\right)\right) \\
\times\left(\sum_{n_{k}} \sum_{n_{k-1}}^{J_{k=1}^{(i)} \sum_{j=1}^{\left.I_{k-1}\right)} h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)}\right. \\
\left.\times p_{S, k}\left(n_{k-1}\right) \omega_{k-1}^{(i, j)}\left(n_{k-1}\right)\right)^{-1} . \tag{32}
\end{gather*}
$$

(2) The GM Update Stage. Suppose that at time $k$ the mixed and predicted multi-Bernoulli density $\bar{\pi}_{k \mid k-1}\left(Y_{k} \mid Z_{1: k-1}\right)=$ $\left\{\left(\bar{r}_{k \mid k-1}^{(i)}, \bar{p}_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k \mid k-1}}$ is given and each $\bar{p}_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)$, $i=1, \ldots, M_{k \mid k-1}$, is composed of GM of the form

$$
\begin{aligned}
& \bar{p}_{k \mid k-1}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right) \\
&=\sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) \\
& \times \mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{k \mid k-1}^{(i, j)}\left(n_{k}\right), P_{k \mid k-1}^{(i, j)}\left(n_{k}\right)\right)
\end{aligned}
$$

Then, the multi-Bernoulli approximation of the updated density $\bar{\pi}_{k}\left(Y_{k} \mid Z_{1: k}\right) \approx\left\{\left(\bar{r}_{L, k}^{(i)}, \bar{p}_{L, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{M_{k \mid k-1}} \cup\left\{\left(\bar{r}_{U, k}\left(\mathbf{z}_{k}\right)\right.\right.$, $\left.\left.\bar{p}_{U, k}\left(\mathbf{x}_{k}, n_{k} ; \mathbf{z}_{k}\right)\right)\right\}_{\mathbf{z}_{k} \in Z_{k}}$ can be computed as follows:

$$
\left.\begin{array}{c}
\bar{r}_{L, k}^{(i)}=\bar{r}_{k \mid k-1}^{(i)} \frac{1-\sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right)}{1-\bar{r}_{k \mid k-1}^{(i)} \sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \varpi_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right)}, \\
\bar{r}_{U, k}\left(\mathbf{z}_{k}\right)=\left(\sum_{i=1}^{M_{k \mid k-1}} \frac{\left(1-\bar{r}_{k \mid k-1}^{(i)}\right) \bar{r}_{k \mid k-1}^{(i)} \sum_{n_{k}} \sum_{j=1}^{J_{k k-1}^{(i)}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right) Q_{k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right)}{\left(1-\bar{r}_{k \mid k-1}^{(i)} \sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right)\right)^{2}}\right)
\end{array}\right)
$$

$$
\begin{gather*}
\times\left(\kappa_{k}\left(\mathbf{z}_{k}\right)+\sum_{i=1}^{M_{k \mid k-1}} \frac{\bar{r}_{k \mid k-1}^{(i)} \sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right) \mathscr{Q}_{k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right)}{1-\bar{r}_{k \mid k-1}^{(i)} \sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right)}\right)^{-1} \\
\bar{p}_{L, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=\frac{\sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)}\left(1-p_{D, k}\left(n_{k}\right)\right) \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) \mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{k \mid k-1}^{(i, j)}\left(n_{k}\right), P_{k \mid k-1}^{(i, j)}\left(n_{k}\right)\right)}{1-\sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right)} \\
\bar{p}_{U, k}\left(\mathbf{x}_{k}, n_{k} ; \mathbf{z}_{k}\right)=\sum_{i=1}^{M_{k \mid k-1}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \omega_{U, k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right) \mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{U, k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right), P_{U, k}^{(i, j)}\left(n_{k}\right)\right) \tag{34}
\end{gather*}
$$

where

$$
\begin{align*}
& \mathbb{Q}_{k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right)=\mathcal{N}\left(\mathbf{z}_{k} ; H_{k}\left(n_{k}\right) \mathbf{m}_{k \mid k-1}^{(i, j)}\left(n_{k}\right),\right. \\
& \mathrm{B}_{k}\left(n_{k}\right) R_{k}\left(n_{k}\right)\left(\mathrm{B}_{k}\left(n_{k}\right)\right)^{T} \\
& \left.+H_{k}\left(n_{k}\right) P_{k \mid k-1}^{(i, j)}\left(n_{k}\right)\left(H_{k}\left(n_{k}\right)\right)^{T}\right) \\
& \omega_{U, k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right) \\
& =\left(\frac{\bar{r}_{k \mid k-1}^{(i)}}{1-\bar{r}_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\left(n_{k}\right) \widehat{Q}_{k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right)\right) \\
& \times\left(\sum_{i=1}^{M_{k \mid k-1}} \sum_{n_{k}} \sum_{j=1}^{J_{k \mid k-1}^{(i)}\left(n_{k}\right)} \frac{\bar{r}_{k \mid k-1}^{(i)}}{1-\bar{r}_{k \mid k-1}^{(i)}} \omega_{k \mid k-1}^{(i, j)}\left(n_{k}\right) p_{D, k}\right. \\
& \left.\times\left(n_{k}\right) Q_{k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right)\right)^{-1},  \tag{35}\\
& K_{U, k}^{(i, j)}\left(n_{k}\right)=P_{k \mid k-1}^{(i, j)}\left(n_{k}\right)\left(H_{k}\left(n_{k}\right)\right)^{T} \\
& \times\left(H_{k}\left(n_{k}\right) P_{k \mid k-1}^{(i, j)}\left(n_{k}\right)\left(H_{k}\left(n_{k}\right)\right)^{T}\right. \\
& \left.+\mathrm{B}_{k}\left(n_{k}\right) R_{k}\left(n_{k}\right)\left(\mathrm{B}_{k}\left(n_{k}\right)\right)^{T}\right)^{-1},  \tag{38}\\
& \mathbf{m}_{U, k}^{(i, j)}\left(n_{k} ; \mathbf{z}_{k}\right)=\mathbf{m}_{k \mid k-1}^{(i, j)}\left(n_{k}\right)+K_{U, k}^{(i, j)}\left(n_{k}\right) \\
& \times\left(\mathbf{z}_{k}-H_{k}\left(n_{k}\right) \mathbf{m}_{k \mid k-1}^{(i, j)}\left(n_{k}\right)\right), \\
& P_{U, k}^{(i, j)}\left(n_{k}\right)=\left(I-K_{U, k}^{(i, j)}\left(n_{k}\right) H_{k}\left(n_{k}\right)\right) P_{k \mid k-1}^{(i, j)}\left(n_{k}\right) .
\end{align*}
$$

from the method described in Algorithm 1, the GM multitarget state estimation can be easily obtained as

$$
\begin{array}{r}
\widehat{X}_{k}=\left\{\widehat{\mathbf{x}}_{k}^{(i)}\right\}_{i=1}^{\widehat{T}_{k}} \quad \text { with } \widehat{\mathbf{x}}_{k}^{(i)}=\sum_{n_{k}} \sum_{j=1}^{J_{k}^{(i)}\left(n_{k}\right)} \Phi_{k}^{(i, j)}\left(n_{k}\right) \mathbf{m}_{k}^{(i, j)}\left(n_{k}\right), \\
i=1, \ldots, \widehat{T}_{k} .
\end{array}
$$

Now turn to considering the extension of the GM-MMCBMeMBer filter to nonlinear dynamical and observation models using the EK filtering approximation. Assumptions $B$ and $C$ are still required, but the dynamic and observation processes can be relaxed to the nonlinear models

$$
\begin{aligned}
\mathbf{x}_{k} & =a_{k}\left(\mathbf{x}_{k-1}, \mathbf{w}_{k}\left(n_{k}\right), n_{k}\right), \\
\mathbf{z}_{k} & =u_{k}\left(\mathbf{x}_{k}, \mathbf{v}_{k}\left(n_{k}\right), n_{k}\right),
\end{aligned}
$$

where $a_{k}\left(\cdot, \cdot, n_{k}\right)$ and $u_{k}\left(\cdot, \cdot, n_{k}\right)$ are known model-dependent nonlinear functions, and $\mathbf{w}_{k}\left(n_{k}\right)$ and $\mathbf{v}_{k}\left(n_{k}\right)$ are modeldependent process and observation noise vectors of known statistics.

For the EK-GM-MM-CBMeMBer filter, the closed form expressions for the mixing, prediction, and update of individual Gaussian components are approximated by replacing $F_{k}\left(n_{k}\right), \mathrm{A}_{k}\left(n_{k}\right), H_{k}\left(n_{k}\right), \mathrm{B}_{k}\left(n_{k}\right)$ in the corresponding recursive equations (30)-(35) of the GM-MM-CBMeMBer filter
with the corresponding local linearization of the nonlinear dynamical and observation models

$$
\begin{align*}
& F_{k}^{\mathrm{EK}}\left(n_{k}\right)=\left.\frac{\partial a_{k}\left(\mathbf{x}_{k-1}, \mathbf{w}_{k}\left(n_{k}\right), n_{k}\right)}{\partial \mathbf{x}_{k-1}}\right|_{\substack{\mathbf{x}_{k-1}=\widehat{\mathbf{x}}_{k-1} \\
\mathbf{w}_{k}\left(n_{k}\right)=0}}, \\
& \mathrm{~A}_{k}^{\mathrm{EK}}\left(n_{k}\right)=\left.\frac{\partial a_{k}\left(\mathbf{x}_{k-1}, \mathbf{w}_{k}\left(n_{k}\right), n_{k}\right)}{\partial \mathbf{w}_{k}\left(n_{k}\right)}\right|_{\substack{\mathbf{x}_{k-1}=\widehat{\mathbf{x}}_{k-1} \\
\mathbf{w}_{k}\left(n_{k}\right)=\mathbf{0}}}, \\
& H_{k}^{\mathrm{EK}}\left(n_{k}\right)=\left.\frac{\partial u_{k}\left(\mathbf{x}_{k}, \mathbf{v}_{k}\left(n_{k}\right), n_{k}\right)}{\partial \mathbf{x}_{k}}\right|_{\substack{\mathbf{x}_{k-1}=\widehat{\mathbf{x}}_{k l k-1} \\
\mathbf{v}_{k}\left(n_{k}\right)=\mathbf{0}}},  \tag{39}\\
& \mathrm{B}_{k}^{\mathrm{EK}}\left(n_{k}\right)=\left.\frac{\partial u_{k}\left(\mathbf{x}_{k}, \mathbf{v}_{k}\left(n_{k}\right), n_{k}\right)}{\partial \mathbf{v}_{k}\left(n_{k}\right)}\right|_{\substack{\mathbf{x}_{k-1}=\widehat{\mathbf{x}}_{k l k-1} \\
\mathbf{v}_{k}\left(n_{k}\right)=\mathbf{0}}} .
\end{align*}
$$

Note that the unscented Kalman version for the GM-MM-CBMeMBer filter can be derived by approximating the mean and covariance of individual Gaussian components with a set of sigma points and the unscented transform [20]. Because of the space limitation, the details of the UK-GM-MM-CBMeMBer filter are not presented here.

## 6. Simulations

6.1. Nonlinear Example Using SMC Implementations. In this nonlinear example, we evaluate the performance of the proposed MM-CBMeMBer filter by benchmarking it against the single-model CBMeMBer filters, the MM-PHD filter, and the MM-CPHD filter using the SMC implementations.

Consider a two-dimensional scenario with an unknown and time varying number of the maneuvering targets observed over the region $[-1000,1000] \times[-1000,1000](\mathrm{m})$ for a period of $N=50$ time steps. The sampling interval is $\Delta t=1$ (s). Each of the targets may move at a nearly constant velocity or execute a coordinated turn in the surveillance period. Therefore, the model set designed for this example can be composed of a constant velocity (CV) model and a coordinated turn (CT) model with varying turn rate [27]. The target kinematical state is $\mathbf{x}_{k}=\left[\begin{array}{lllll}x_{k} & \dot{x}_{k} & y_{k} & \dot{y}_{k} & \vartheta_{k}\end{array}\right]^{T}$, where $\left[\begin{array}{ll}x_{k} & y_{k}\end{array}\right]^{T}$ and $\left[\begin{array}{ll}\dot{x}_{k} & \dot{y}_{k}\end{array}\right]^{T}$, respectively, represent the position and the velocity in $x$ and $y$ coordinates and $\vartheta_{k}$ represents the turn rate. For the turn rate $\mathcal{\vartheta}_{k}$, let the anticlockwise direction be positive and the clockwise direction be negative.

The model-dependent dynamics for the individual maneuvering target is given by the linear-Gaussian model

$$
\begin{equation*}
f_{k \mid k-1}\left(\mathbf{x}_{k} \mid \mathbf{x}_{k-1}, n_{k}\right)=\mathscr{N}\left(\mathbf{x}_{k} ; F_{k}\left(n_{k}\right) \mathbf{x}_{k-1}, Q_{k}\left(n_{k}\right)\right) \tag{40}
\end{equation*}
$$

Let $n_{k}=1$ denote the CV model and $n_{k}=2$ the CT model; then

$$
\begin{gather*}
F_{k}\left(n_{k}=1\right)=\left[\begin{array}{ll}
F_{\mathrm{CV}} & \\
& 0
\end{array}\right], \\
F_{k}\left(n_{k}=2\right)=\left[\begin{array}{ll}
F_{\mathrm{CT}}\left(\vartheta_{k-1}\right) & \\
& 1
\end{array}\right], \\
Q_{k}\left(n_{k}=1\right)=\sigma_{w}^{2}\left(n_{k}=1\right)\left[\begin{array}{ll}
Q & \\
& 0
\end{array}\right],  \tag{41}\\
Q_{k}\left(n_{k}=2\right)=\left[\begin{array}{ll}
\sigma_{1, w}^{2}\left(n_{k}=2\right) Q & \\
& \Delta t^{2} \sigma_{2, w}^{2}\left(n_{k}=2\right)
\end{array}\right]
\end{gather*}
$$

with

$$
\begin{gather*}
F_{\mathrm{CV}}=\left[\begin{array}{cccc}
1 & \Delta t & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & \Delta t \\
0 & 0 & 0 & 1
\end{array}\right], \\
F_{\mathrm{CT}}\left(\vartheta_{k-1}\right)=\left[\begin{array}{cccc}
1 & \frac{\Delta t \sin \vartheta_{k-1}}{\vartheta_{k-1}} & 0 & -\frac{1-\Delta t \cos \vartheta_{k-1}}{\vartheta_{k-1}} \\
0 & \Delta t \cos \vartheta_{k-1} & 0 & -\Delta t \sin \vartheta_{k-1} \\
0 & \frac{1-\Delta t \cos \vartheta_{k-1}}{\vartheta_{k-1}} & 1 & \frac{\Delta t \sin \vartheta_{k-1}}{\vartheta_{k-1}} \\
0 & \Delta t \sin \vartheta_{k-1} & 0 & \Delta t \cos \vartheta_{k-1}
\end{array}\right], \\
Q=\left[\begin{array}{cccc}
\frac{\Delta t^{4}}{4} & \frac{\Delta t^{3}}{2} & 0 & 0 \\
\frac{\Delta t^{3}}{2} & \Delta t^{2} & 0 & 0 \\
0 & 0 & \frac{\Delta t^{4}}{4} & \frac{\Delta t^{3}}{2} \\
0 & 0 & \frac{\Delta t^{3}}{2} & \Delta t^{2}
\end{array}\right], \tag{42}
\end{gather*}
$$

where $\sigma_{w}\left(n_{k}\right)$ is the level of the power spectral density of the process noise for model $n_{k}$. In this example, they are given by $\sigma_{w}\left(n_{k}=1\right)=0.1\left(\mathrm{~m} / \mathrm{s}^{2}\right), \sigma_{1, w}\left(n_{k}=2\right)=0.2\left(\mathrm{~m} / \mathrm{s}^{2}\right)$, $\sigma_{2, w}\left(n_{k}=2\right)=1 \times 10^{-3}\left(\mathrm{rad} / \mathrm{s}^{2}\right)$.

The Markovian model transition probability matrix is taken as

$$
\left[h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)\right]=\left[\begin{array}{cc}
0.8 & 0.2  \tag{43}\\
0.2 & 0.8
\end{array}\right] .
$$

At time $k$, the range $\rho_{k}$ and bearing $\varphi_{k}$ measurements of the targets are generated by a sensor located at $\left[\begin{array}{ll}0 & 0\end{array}\right]^{T}$. The measurement noise is independent and identically distributed (IID) zero-mean Gaussian white noise with covariance matrix $R_{k}=\operatorname{diag}\left(\sigma_{\rho}^{2} \sigma_{\varphi}^{2}\right)$, where $\operatorname{diag}(\cdot)$ denotes the diagonal matrix, and $\sigma_{\rho}$ and $\sigma_{\varphi}$ are, respectively, standard deviations (STDs) of the range and bearing measurements. In this example, they are taken as $\sigma_{\rho}=10(\mathrm{~m})$ and
$\sigma_{\varphi}=0.01$ (rad). The single-measurement single-target likelihood density is

$$
g_{k}\left(\mathbf{z}_{k} \mid \mathbf{y}_{k}\right)=\mathcal{N}\left(\mathbf{z}_{k} ;\left[\begin{array}{c}
\sqrt{x_{k}^{2}+y_{k}^{2}}  \tag{44}\\
\arctan \frac{y_{k}}{x_{k}}
\end{array}\right], R_{k}\right)
$$

The detection probability and the survival probability are, respectively, taken as $p_{D, k}\left(\mathbf{x}_{k}, n_{k}\right)=p_{D}=0.95$ and $p_{S, k}\left(\mathbf{x}_{k-1}, n_{k-1}\right)=p_{S}=0.95$ in this example.

The clutter is modeled as a Poisson RFS with the intensity $\kappa_{k}\left(\mathbf{z}_{k}\right)=\lambda_{c, k} f_{c, k}\left(\mathbf{z}_{k}\right)$. In this example, we take $\lambda_{c, k}=20$ and $f_{c, k}(\cdot)=\mathscr{U}(\cdot)$, where $\mathscr{U}(\cdot)$ denotes the density of the uniform distribution over the observation region.

Figure 1 shows the true trajectories for the maneuvering targets and sensor location.

In Figure 1, "○" denotes the locations at which targets are born and " $\square$ " denotes the locations at which targets die. Target 1 is born at 1 s and dies at 30 s . It first moves at a nearly constant velocity from the first second to the 15 th second and then executes a coordinated turn in the anticlockwise direction from the 16th second to the 30th second. Target 2 is born at 1 s and dies at 35 s . It first executes a coordinated turn in the anticlockwise direction from the first second to the 20th second and then moves at a nearly constant velocity from the 21 st second to the 35 th second. Target 3 is born at 10 s and dies at 42 s . It first executes a coordinated turn in the anticlockwise direction from the 10th second to the 30th second and then moves at a nearly constant velocity from the 31st second to the 42 nd second. Target 4 is born at 20 s and dies at 50 s . It first moves at a nearly constant velocity from the 20th second to the 30th second and then executes a coordinated turn in the clockwise direction from the 31st second to the 50th second. The motions of the targets are summarized in Table 1.

The birth process is a multi-Bernoulli RFS with density $\pi_{\Gamma, k}\left(Y_{k}\right)=\left\{\left(r_{\Gamma, k}^{(i)}, p_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)\right)\right\}_{i=1}^{3}$, where $r_{\Gamma, k}^{(1)}=0.04, r_{\Gamma, k}^{(2)}=$ $r_{\Gamma, k}^{(3)}=0.02, p_{\Gamma, k}^{(i)}\left(\mathbf{x}_{k}, n_{k}\right)=h_{\Gamma, k}^{(i)}\left(n_{k}\right) \mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{\Gamma, k}^{(i)}, P_{\Gamma, k}^{(i)}\right)$ with

$$
\begin{align*}
& \mathbf{m}_{\Gamma, k}^{(1)}=\left[\begin{array}{lllll}
-600 & 0 & 800 & 0 & 0
\end{array}\right]^{T}, \\
& \mathbf{m}_{\Gamma, k}^{(2)}=\left[\begin{array}{lllll}
-650 & 0 & -800 & 0 & 0
\end{array}\right]^{T}, \\
& \mathbf{m}_{\Gamma, k}^{(3)}=\left[\begin{array}{lllll}
400 & 0 & -400 & 0 & 0
\end{array}\right]^{T},  \tag{45}\\
& P_{\Gamma, k}^{(1)}=P_{\Gamma, k}^{(2)}=P_{\Gamma, k}^{(3)}=\operatorname{diag}\left(\begin{array}{lllll}
400 & 400 & 400 & 400 & 0.01
\end{array}\right)
\end{align*}
$$

and the distribution of the model births

$$
\left[h_{\Gamma, k}^{(i)}\left(n_{k}\right)\right]=\left[\begin{array}{ll}
0.5 & 0.5 \tag{46}
\end{array}\right]
$$

For the purpose of comparison, we estimate the number and states of the maneuvering targets using the proposed SMC-MM-CBMeMBer filter, the CV model SMCCBMeMBer filter, the CT model SMC-CBMeMBer filter, the SMC-MM-PHD filter, and the SMC-MM-CPHD filter, respectively. At each time step in the SMC implementations of the CBMeMBer-based filters, a maximum of $L_{\max }=1000$ and minimum of $L_{\text {min }}=300$ particles per hypothesized


Figure 1: The true trajectories for the maneuvering targets and sensor location.

Table 1: The motions of the targets.

|  | Born <br> time | Die <br> time | CV motion | CT motion |
| :--- | :---: | :---: | :---: | :---: |
| Target 1 | 1 s | 30 s | $1 \mathrm{~s}-15 \mathrm{~s}$ | $16 \mathrm{~s}-30 \mathrm{~s}$, anticlockwise |
| Target 2 | 1 s | 35 s | $21 \mathrm{~s}-35 \mathrm{~s}$ | $1 \mathrm{~s}-20 \mathrm{~s}$, anticlockwise |
| Target 3 | 10 s | 42 s | $31 \mathrm{~s}-42 \mathrm{~s}$ | $10 \mathrm{~s}-30 \mathrm{~s}$, anticlockwise |
| Target 4 | 20 s | 50 s | $20 \mathrm{~s}-30 \mathrm{~s}$ | $31 \mathrm{~s}-50 \mathrm{~s}$, clockwise |

track are imposed, and pruning of hypothesized tracks is performed with a threshold of $\tilde{r}_{\text {threshold }}=0.001$. At each time step in the SMC implementations of the PHD-based filters, 1000 particles are used to represent one target and $K$-means method [14] is used to cluster the resampled particles to extract the multitarget states. The proposal densities $\alpha_{k}^{(i)}\left(n_{k} \mid\right.$ $\left.n_{k-1}\right), \beta_{k}^{(i)}\left(n_{k}\right), q_{k}^{(i)}\left(\mathbf{x}_{k} \mid \mathbf{x}_{k-1}, n_{k}, Z_{k}\right)$, and $b_{k}^{(i)}\left(\mathbf{x}_{k} \mid n_{k}, Z_{k}\right)$ in (16) and (18) are, respectively, taken as $h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)$, $h_{\Gamma, k}^{(i)}\left(n_{k}\right), f_{k \mid k-1}\left(\mathbf{x}_{k} \mid \mathbf{x}_{k-1}, n_{k}\right)$ and $\mathcal{N}\left(\mathbf{x}_{k} ; \mathbf{m}_{\Gamma, k}^{(i)}, P_{\Gamma, k}^{(i)}\right)$. We now conduct 500 Monte Carlo (MC) simulation experiments on the same clutter intensity and target trajectories, but with independently generated clutter and target-generated measurements in each trial.

The MC averages of the mean and STD of the cardinality distribution for the five methods at each time step are shown along with the true target number in Figure 2, respectively.

Figures 2(a)-2(e) demonstrate that the target number estimates from the SMC-MM-PHD, SMC-MM-CPHD, and SMC-MM-CBMeMBer filters converge to the ground truth, whereas the CV model SMC-CBMeMBer and CT model SMC-CBMeMBer filters produce significant bias in estimating the target number. This is because the SMC-MMPHD, SMC-MM-CPHD, and SMC-MM-CBMeMBer filters can effectively capture the model switching property of the maneuvering targets, so their performance is significantly better than that of the two single-model SMC-CBMeMBer

(a)

(c)

(b)

(d)


| _ | Truth |
| :--- | :--- |
| $\circ$ | Mean |
| ..-- | STD |

(e)

Figure 2: The 500 MC run averages of cardinality statistics versus time for the (a) CV model SMC-CBMeMBer filter, (b) CT model SMCCBMeMBer filter, (c) SMC-MM-PHD filter, (d) SMC-MM-CPHD filter, and (e) SMC-MM-CBMeMBer filter.
filters, which show poor adaptation to target maneuvers and yield larger estimation errors.

Moreover, as plotted in Figures 2(c)-2(e), the STD of the cardinality distribution from the SMC-MM-CBMeMBer filter is lower than that of the SMC-MM-PHD filter, but larger than that of SMC-MM-CPHD filter. In addition, the STDs of the cardinality distributions from the three MM-based filters increase in different degrees at the instances when the maneuver occurs (i.e., 16 (s), 21 (s), and 31 (s)). The STD plots of the SMC-MM-PHD and SMC-MM-CPHD filters seem to fluctuate more obviously than the SMC-MM-CBMeMBer filter. This phenomenon indicates that the performance of the SMC-MM-CBMeMBer filter may be more stable and robust at the maneuver instances than that of the SMC-MM-PHD and SMC-MM-CPHD filters.

The optimal subpattern assignment (OSPA) metric [28], which can jointly capture differences in cardinality and individual elements between two finite sets, is used to evaluate the performance of the five methods. Given the actual and estimated multitarget state sets $X_{k}=\left\{\mathbf{x}_{k}^{(i)}\right\}_{i=1}^{T_{k}}$ and $\widehat{X}_{k}=$ $\left\{\widehat{\mathbf{x}}_{k}^{(i)}\right\}_{i=1}^{\widehat{T}_{k}}$, the OSPA metric of order $p=2$ with cut-off $c$ between the two sets is defined by

$$
\begin{align*}
& \operatorname{OSPA}_{2, k}^{(c)}\left(X_{k}, \widehat{X}_{k}\right) \\
& =\left(\frac { 1 } { \widehat { T } _ { k } } \left(\min _{\pi \in \Pi_{\widehat{T}_{k}}} \sum_{i=1}^{T_{k}} \min \left(c,\left\|\mathbf{x}_{k}^{(i)}-\widehat{\mathbf{x}}_{k}^{(\pi(i))}\right\|_{2}\right)^{2}\right.\right.  \tag{47}\\
& \left.\left.+c^{2}\left(\widehat{T}_{k}-T_{k}\right)\right)\right)^{1 / 2}
\end{align*}
$$

if $T_{k} \leq \widehat{T}_{k}$ and $\operatorname{OSPA}_{2, k}^{(c)}\left(X_{k}, \widehat{X}_{k}\right)=\operatorname{OSPA}_{2, k}^{(c)}\left(\widehat{X}_{k}, X_{k}\right)$ if $T_{k}>$ $\widehat{T}_{k} \cdot \Pi_{\widehat{T}_{k}}$ denotes the set of permutations on $\left\{1,2, \ldots, \widehat{T}_{k}\right\} \cdot\|\cdot\|_{2}$ denotes the 2-norm. In this example, we take $c=100(\mathrm{~m})$.

The MC averages of the OSPA metric for the target position estimates, derived by the five methods, are shown in Figure 3.

The OSPA metric is composed of two components each separately accounting for "localization" and "cardinality" errors. This results in high peaks in OSPA metric at the instances where the estimated number is incorrect. Figure 3 shows that (1) both the single-model SMC-CBMeMBer filters perform significantly worse than the other MM-based filters because of the large cardinality errors produced by the two filters as seen in Figures 2(a) and 2(b); (2) although the SMC-MM-CPHD filter can estimate the target number most accurately, the OSPA metric of the SMC-MM-CBMeMBer filter is smaller than that of the SMC-MM-CPHD filter, which is in turn smaller than that of the SMC-MM-PHD filter. This phenomenon indicates that the SMC-MM-CBMeMBer filter outperforms the SMC-MM-CPHD (and hence SMC-MM-PHD) filter in jointly estimating the multitarget number and states. A reason for this is that the additional errors could be introduced in the clustering processes of the SMC-MM-PHD and SMC-MM-CPHD filters to extract state estimates from the particle population; (3) the OSPA plots of


Figure 3: The 500 MC run averages of OSPA against time.
the three MM-based filters in Figure 3 fluctuate against time due to the varying target number, the target maneuvers, and clutter. However, increase of the OSPA from the SMC-MM-CBMeMBer filter seems to be smallest at the maneuver instances (i.e., 16 (s), 21 (s), and 31 (s)) among the three MM-based methods. This phenomenon also indicates that the performance of the SMC-MM-CBMeMBer filter may be more stable and robust at the maneuver instances than that of the SMC-MM-PHD and SMC-MM-CPHD filters.

For comparing the overall performance of the three MMbased filters, the 500 MC trial averages of the OSPA distance (time-averaged over the duration of the scenario) for the three MM-based filters are shown in Table 2 against the clutter rate from $\lambda_{c, k}=20$ to $\lambda_{c, k}=100$. The result of time-averaging can be viewed as a broad indication of filter performance, although the average is likely to be scenario dependent.

Table 2 shows that the OSPA distances of the three MMbased filters increase with higher $\lambda_{c, k}$. It reflects that the performance of the three MM-based algorithms degrades by different degrees as the $\lambda_{c, k}$ increases. Among the three MMbased algorithms, the SMC-MM-PHD filter always works the worst. The SMC-MM-CBMeMBer filter outperforms the SMC-MM-CPHD filter when $\lambda_{c, k}$ is relatively lower ( $\lambda_{c, k} \leq$ 60). However, as the $\lambda_{c, k}$ increases, the OSPA distance of SMC-MM-CBMeMBer filter increases more rapidly than that of the SMC-MM-CPHD filter. Therefore, as $\lambda_{c, k}$ continues to increase until it reaches $\lambda_{c, k}=80$, the OSPA distance of SMC-MM-CBMeMBer filter is very close to that of the SMC-MM-CPHD filter. When $\lambda_{c, k}$ is relatively higher (i.e., $\lambda_{c, k}=$ 80), the SMC-MM-CPHD filter outperforms the SMC-MMCBMeMBer filter. A possible reason for this is that, compared

TABLE 2: Time-averaged OSPA distance ( $m$ ) in various $\lambda_{c, k}$.

|  | $\lambda_{c, k}=20$ | $\lambda_{c, k}=40$ | $\lambda_{c, k}=60$ | $\lambda_{c, k}=80$ | 74.3 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| SMC-MM-PHD filter | 38.3 | 48.2 | 60.5 | 43.4 |  |
| SMC-MM-CPHD filter | 32.8 | 35.9 | 39.4 | 88.1 |  |
| SMC-MM-CBMeMBer filter | 25.6 | 31.3 | 37.4 | 47.8 |  |

with the SMC-MM-CBMeMBer filter, the advantage of the target number estimate for the SMC-MM-CPHD filter is more obvious as the $\lambda_{c, k}$ increases and it finally leads that the OSPA distance of the SMC-MM-CPHD filter is smaller than that of the SMC-MM-CBMeMBer filter when $\lambda_{c, k}$ is relatively higher (i.e., $\lambda_{c, k}=80$ ).
6.2. Linear-Gaussian Example Using GM Implementations. In this linear-Gaussian example, we evaluate the performance of the proposed MM-CBMeMBer filter by benchmarking it against the single-model CBMeMBer filters, the MM-PHD filter, and the MM-CPHD filter using the GM implementations.

The simulation scenario and true trajectories for the maneuvering targets are the same as those of Example 1. The target kinematical state now turns into $\mathbf{x}_{k}=$ $\left[\begin{array}{llll}x_{k} & \dot{x}_{k} & y_{k} & \dot{y}_{k}\end{array}\right]^{T}$. The model set for this example is designed as follows. Model $n_{k}=1$ is a CV model with linearGaussian dynamics given by $\mathcal{N}\left(\mathbf{x}_{k} ; F_{\mathrm{CV}} \mathbf{x}_{k-1}, \sigma_{\mathrm{CV}, w}^{2} Q\right)$; models $n_{k}=2,3,4,5$ are, respectively, CT models with turn rates of $\vartheta=\pi / 30,-\pi / 30, \pi / 20, \pi / 15(\mathrm{rad} / \mathrm{s})$ with linear-Gaussian dynamics given by $\mathscr{N}\left(\mathbf{x}_{k} ; F_{\mathrm{CT}}(\vartheta) \mathbf{x}_{k-1}, \sigma_{\mathrm{CT}, w}^{2} Q\right)$. In this example, $\sigma_{\mathrm{CV}, w}$ and $\sigma_{\mathrm{CT}, w}$ are given by $\sigma_{\mathrm{CV}, w}=0.1\left(\mathrm{~m} / \mathrm{s}^{2}\right), \sigma_{\mathrm{CT}, w}=$ $0.2\left(\mathrm{~m} / \mathrm{s}^{2}\right)$.

The Markovian model transition probability matrix now turns into

$$
\left[h_{k \mid k-1}\left(n_{k} \mid n_{k-1}\right)\right]=\left[\begin{array}{cccccc}
0.6 & 0.1 & 0.1 & 0.1 & 0.1  \tag{48}\\
0.1 & 0.6 & 0.1 & 0.1 & 0.1 \\
0.1 & 0.1 & 0.6 & 0.1 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.6 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.1 & 0.6
\end{array}\right] .
$$

The $x$-position and $y$-position measurements $\mathbf{z}_{k}=$ $\left[\begin{array}{ll}x_{k} & y_{k}\end{array}\right]^{T}$ of the maneuvering targets are generated by the linear-Gaussian single-measurement single-target likelihood density given by $\mathcal{N}\left(\mathbf{z}_{k} ; H_{k} \mathbf{x}_{k}, R_{k}\right)$ with

$$
H_{k}=\left[\begin{array}{llll}
1 & 0 & 0 & 0  \tag{49}\\
0 & 0 & 1 & 0
\end{array}\right]
$$

and $R_{k}=\operatorname{diag}\left(\sigma_{x}^{2} \sigma_{y}^{2}\right)$. In this example, they are taken as $\sigma_{x}=\sigma_{y}=8(\mathrm{~m})$, and the kinematical state independent survival and detection probabilities are taken as $p_{D, k}\left(n_{k}\right)=$ $p_{D}=0.95$ and $p_{S, k}\left(n_{k-1}\right)=p_{S}=0.95$.

The experiment settings of the clutter and birth model are also the same as those of Example 1 except that the $\mathbf{m}_{\Gamma, k}^{(i)}, P_{\Gamma, k}^{(i)}$, $i=1,2,3$, and $\left[h_{\Gamma, k}^{(i)}\left(n_{k}\right)\right]$ turn into

$$
\begin{gather*}
\mathbf{m}_{\Gamma, k}^{(1)}=\left[\begin{array}{llll}
-600 & 0 & 800 & 0
\end{array}\right]^{T}, \\
\mathbf{m}_{\Gamma, k}^{(2)}=\left[\begin{array}{llll}
-650 & 0 & -800 & 0
\end{array}\right]^{T}, \\
\mathbf{m}_{\Gamma, k}^{(3)}=\left[\begin{array}{llll}
400 & 0 & -400 & 0
\end{array}\right]^{T},  \tag{50}\\
P_{\Gamma, k}^{(1)}=P_{\Gamma, k}^{(2)}=P_{\Gamma, k}^{(3)}=\operatorname{diag}\left(\begin{array}{llll}
400 & 400 & 400 & 400
\end{array}\right), \\
{\left[h_{\Gamma, k}^{(i)}\left(n_{k}\right)\right]=\left[\begin{array}{lllll}
0.2 & 0.2 & 0.2 & 0.2 & 0.2
\end{array}\right] .}
\end{gather*}
$$

For the purpose of comparison, we estimate the number and states of the maneuvering targets using the proposed GM-MM-CBMeMBer filter, the CV model GM-CBMeMBer filter, the CT model GM-CBMeMBer filter with turn rate of $\mathcal{\vartheta}=\pi / 20(\mathrm{rad} / \mathrm{s})$, (this turn rate seems to be most suitable for the scenario among the above four turn rates), the GM-MM-PHD filter, and the GM-MM-CPHD filter, respectively. At each time step in the GM implementations of the CBMeMBer-based filters, pruning of hypothesized tracks is performed with a threshold of $\tilde{r}_{\text {threshold }}=0.001$. In addition, the pruning and merging of Gaussian components are performed for each hypothesized track using a weight threshold of $10^{-5}$, a merging threshold of $4(\mathrm{~m})$, and a maximum of $J_{\max }=100$ components, which are also used in the GM implementations of the PHD-based filters.

The MC averages of the mean and STD of the cardinality distribution for the five methods at each time step are shown along with the true target number in Figure 4, respectively.

Similar to the SMC implementations, Figures 4(a)-4(e) demonstrate that the GM implementations of the three MMbased filters are unbiased in the target number estimates, whereas the GM implementations of the two single-model GM-CBMeMBer filters are significantly biased. Moreover, the GM-MM-CBMeMBer filter has a lower STD of the estimated cardinality than the GM-MM-PHD filter but has a larger STD than the GM-MM-CPHD filter. Again, The STD plots of the GM-MM-PHD and GM-MM-CPHD filters seem to fluctuate more obviously than the GM-MM-CBMeMBer filter at the maneuver instances (i.e., 16 (s), 21 (s), and 31 (s)).

The MC averages of the OSPA metric for the target position estimates, derived by the five methods, are shown in Figure 5.

In contrast to the SMC case, Figure 5 shows that (1) the rather poor performance of the two single-model GMCBMeMBer filters can be expected as the direct results of their significant cardinality biase as seen in Figures 4(a) and


FIGURE 4: The 500 MC run averages of cardinality statistics versus time for the (a) CV model GM-CBMeMBer filter, (b) CT model GMCBMeMBer filter, (c) GM-MM-PHD filter, (d) GM-MM-CPHD filter, and (e) GM-MM-CBMeMBer filter.

Table 3: Time-averaged OSPA distance ( $m$ ) in various $\lambda_{c, k}$.

|  | $\lambda_{c, k}=20$ | $\lambda_{c, k}=40$ | $\lambda_{c, k}=60$ | $\lambda_{c, k}=80$ | $\lambda_{c, k}=100$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| GM-MM-PHD filter | 22.8 | 28.2 | 34.5 | 40.9 | 47.1 |
| GM-MM-CPHD filter | 20.0 | 24.8 | 30.3 | 36.1 | 41.9 |
| GM-MM-CBMeMBer filter | 22.6 | 27.7 | 34.1 | 40.6 | 46.7 |



Figure 5: The 500 MC run averages of OSPA against time.

4(b); (2) the OSPA metric of the GM-MM-CBMeMBer filter is similar to that of the GM-MM-PHD filter but is larger than that of the GM-MM-CPHD filter. This is because that, like the MM-CBMeMBer filter, the GM implementations of the MM-PHD and MM-CPHD filters also allow state estimates to be extracted from the posterior intensity in a much more efficient and reliable manner than particle clustering in the SMC-based approach. As a result, the GM-MM-CPHD filter, which has the lowest STD of the estimated cardinality, performs best among the three MM-based filters. Although the GM-MM-CBMeMBer filter has a lower STD of the estimated cardinality than the GM-MM-PHD filter, the performance of the two filters is similar. A reason for this is that the GM-MM-PHD filter may have more of an advantage than the GM-MM-CBMeMBer filter in the relatively high signal to noise ratio (SNR) of this scenario.

Although the GM-MM-CPHD filter outperforms the proposed GM-MM-CBMeMBer filter, it can be only used in the linear-Gaussian condition. In the nonlinear nonGaussian conditions, both the MM-CPHD filter and MMCBMeMBer filter must be implemented by the SMC method. In this case, the GM-MM-CBMeMBer filter outperforms the GM-MM-CPHD filter significantly, which is shown in Section 6.1.

The 500 MC trial averages of the OSPA distance (timeaveraged over the duration of the scenario) for the three

MM-based filters are shown in Table 3 against the clutter rate from $\lambda_{c, k}=20$ to $\lambda_{c, k}=100$.

Similar to the SMC implementations, Table 3 shows that the OSPA distances of the GM implementations of the three MM-based filters increase with higher $\lambda_{c, k}$. However, in various $\lambda_{c, k}$, the GM-MM-CPHD filter always has the best performance among the three MM-based algorithms, and the GM-MM-CBMeMBer filter has the similar performance with the GM-MM-PHD filter.

## 7. Conclusions and Future Work

An MM-CBMeMBer filter, which is a multiple-model extension to the CBMeMBer filter, is proposed for tracking multiple maneuvering targets in clutter. The SMC and GM implementations of the proposed filter are, respectively, presented for generic models and for linear-Gaussian models. Then, the EK and UK filtering approximations for the GM-MM-CBMeMBer filter in nonlinear condition are described briefly. Simulation results show that (1) the proposed MMCBMeMBer filter significantly outperforms the single-model CBMeMBer filters in tracking multiple maneuvering targets; (2) under relatively low clutter density, the SMC-MM-CBMeMBer filter outperforms the SMC-MM-PHD and SMC-MM-CPHD filters; (3) the performance of the GM-MM-CBMeMBer filter is similar to that of the GM-MM-PHD filter and hence is inferior to that of GM-MM-CPHD filter.

The future work is focused on the following three aspects. First, the track labeling problem in the proposed approach needs to be considered. Second, practical data need to be used for the performance evaluation of the proposed approaches. Third, the multiple-sensor versions of the CBMeMBer and MM-CBMeMBer filters need to be proposed for improving the performance of the single-sensor CBMeMBer and MMCBMeMBer filters.

## Acknowledgments

This research work was supported by the National Key Fundamental Research \& Development Programs (973) of China (2013CB329405), Foundation for Innovative Research Groups of the National Natural Science Foundation of China (61221063), Natural Science Foundation of China (61203221, 61004087), Ph.D. Programs Foundation of Ministry of Education of China (20100201120036), China Postdoctoral Science Foundation (2011M501442, 20100481338), and Fundamental Research Funds for the Central University.

## References

[1] I. R. Goodman, R. P. S. Mahler, and H. T. Nguyen, Mathematics of Data Fusion, Kluwer Academic, Norwood, Mass, USA, 1997.
[2] R. Mahler, Statistical Multisource Multitarget Information Fusion, Artech House, Norwood, Mass, USA, 2007.
[3] R. Mahler, "Multi-target Bayes filtering via first-order multitarget moments," IEEE Transactions on Aerospace and Electronic Systems, vol. 39, no. 4, pp. 1152-1178, 2003.
[4] R. Mahler, "PHD filters of higher order in target number," IEEE Transactions on Aerospace and Electronic Systems, vol. 43, no. 4, pp. 1523-1543, 2007.
[5] T. Zajic and R. Mahler, "A particle-systems implementation of the PHD multitarget tracking filter," in The International Society for Optical Engineering, vol. 5096 of Proceedings of SPIE, pp. 291299, April 2003.
[6] H. Sidenbladh, "Multi-target particle filtering for the probability hypothesis density," in Proceedings of the 6th International Conference on Information Fusion, pp. 800-806, Cairns, Australia, 2003.
[7] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," IEEE Transactions on Aerospace and Electronic Systems, vol. 41, no. 4, pp. 1224-1245, 2005.
[8] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," IEEE Transactions on Signal Processing, vol. 54, no. 11, pp. 4091-4104, 2006.
[9] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," IEEE Transactions on Signal Processing, vol. 55, no. 7, pp. 3553-3567, 2007.
[10] H. Zhang, Z. Jing, and S. Hu, "Gaussian mixture CPHD filter with gating technique," Signal Processing, vol. 89, no. 8, pp. 15211530, 2009.
[11] D. Clark, I. T. Ruiz, Y. Petillot, and J. Bell, "Particle PHD filter multiple target tracking in sonar image," IEEE Transactions on Aerospace and Electronic Systems, vol. 43, no. 1, pp. 409-416, 2007.
[12] E. Maggio, M. Taj, and A. Cavallaro, "Efficient multitarget visual tracking using random finite sets," IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 8, pp. 1016-1027, 2008.
[13] F. Lian, C. Han, W. Liu, J. Liu, and J. Sun, "Unified cardinalized probability hypothesis density filters for extended targets and unresolved targets," Signal Processing, vol. 92, no. 7, pp. 17291744, 2012.
[14] D. E. Clark and J. Bell, "Multi-target state estimation and track continuity for the particle PHD filter," IEEE Transactions on Aerospace and Electronic Systems, vol. 43, no. 4, pp. 1441-1453, 2007.
[15] W. Liu, C. Han, F. Lian, and H. Zhu, "Multitarget state extraction for the PHD filter using MCMC approach," IEEE Transactions on Aerospace and Electronic Systems, vol. 46, no. 2, pp. 864-883, 2010.
[16] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 409423, 2009.
[17] R. Hoseinnezhad, B.-N. Vo, B.-T. Vo, and D. Suter, "Bayesian integration of audio and visual information for multi-target tracking using a CB-member filter," in Proceedings of the 36th

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11), pp. 2300-2303, Prague, Czech Republic, May 2011.
[18] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking-part V: multiple-model methods," IEEE Transactions on Aerospace and Electronic Systems, vol. 41, no. 4, pp. 1255-1321, 2005.
[19] W. Mei, G. Shan, and C. Wang, "Practical development of the second-order extended Kalman filter for very long range radar tracking," Signal Processing, vol. 91, no. 5, pp. 1240-1248, 2011.
[20] S. Sadhu, S. Mondal, M. Srinivasan, and T. K. Ghoshal, "Sigma point Kalman filter for bearing only tracking," Signal Processing, vol. 86, no. 12, pp. 3769-3777, 2006.
[21] K. Punithakumar, T. Kirubarajan, and A. Sinha, "Multiplemodel probability hypothesis density filter for tracking maneuvering targets," IEEE Transactions on Aerospace and Electronic Systems, vol. 44, no. 1, pp. 87-98, 2008.
[22] S. A. Pasha, B.-N. Vo, H. D. Tuan, and W.-K. Ma, "A Gaussian mixture PHD filter for jump Markov system models," IEEE Transactions on Aerospace and Electronic Systems, vol. 45, no. 3, pp. 919-936, 2009.
[23] C. Ouyang, H.-B. Ji, and Z.-Q. Guo, "Extensions of the SMCPHD filters for jump Markov systems," Signal Processing, vol. 92, no. 6, pp. 1422-1430, 2012.
[24] W. Li and Y. Jia, "Gaussian mixture PHD filter for jump Markov models based on best-fitting Gaussian approximation," Signal Processing, vol. 91, no. 4, pp. 1036-1042, 2011.
[25] R. Georgescu and P. Willett, "The multiple model CPHD tracker," IEEE Transactions on Signal Processing, vol. 60, no. 4, pp. 1741-1751, 2012.
[26] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," IEEE Transactions on Signal Processing, vol. 49, no. 3, pp. 613-624, 2001.
[27] X. R. Li and V. P. Jikov, "A survey of maneuvering target tracking: dynamic models," in Signal and Data Processing of Small Targets, vol. 4048 of Proceedings of SPIE, pp. 212-235, Orlando, Fla, USA, April 2000.
[28] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," IEEE Transactions on Signal Processing, vol. 56, no. 8, pp. 3447-3457, 2008.

## Research Article

# Sign Inference for Dynamic Signed Networks via Dictionary Learning 

Yi Cen, Rentao Gu, and Yuefeng Ji<br>State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China<br>Correspondence should be addressed to Yuefeng Ji; jyf@bupt.edu.cn

Received 19 July 2013; Accepted 15 October 2013
Academic Editor: Dewei Li
Copyright © 2013 Yi Cen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Mobile online social network (mOSN) is a burgeoning research area. However, most existing works referring to mOSNs deal with static network structures and simply encode whether relationships among entities exist or not. In contrast, relationships in signed mOSNs can be positive or negative and may be changed with time and locations. Applying certain global characteristics of social balance, in this paper, we aim to infer the unknown relationships in dynamic signed mOSNs and formulate this sign inference problem as a low-rank matrix estimation problem. Specifically, motivated by the Singular Value Thresholding (SVT) algorithm, a compact dictionary is selected from the observed dataset. Based on this compact dictionary, the relationships in the dynamic signed mOSNs are estimated via solving the formulated problem. Furthermore, the estimation accuracy is improved by employing a dictionary self-updating mechanism.


## 1. Introduction

Over the past few years, a number of mobile applications that allow users to enjoy networking have emerged. Correspondingly, there has been a proliferation in mobile online social networks (mOSNs). With the ubiquitous use of mobile devices and a rapid shift of technology, it is worthy to investigate the mOSNs from a privacy or security standpoint $[1,2]$. The related applications are also extensive such as authentication and recommendation online. In this context, researches about mobile online networks where two opposite kinds of relationships can occur have become common; people not only form links to indicate friendship, support, or approval but signify disapproval or distrust of the opinions of others. It is natural to model such networks as signed networks, where the sign of a link weight can be either positive or negative, representing the status of a relationship. Analogous to traditional social networks analysis, the relationships in signed mOSNs can be represented as a graph, where nodes denote the objects (e.g., people or mobile terminals) and signed edges denote the relationships or links (e.g., a communication made between two people). The link structure of the resulting graph can be exploited to detect
underlying groups of objects, predict missing links, and handle many other tasks [3-17].

One of the most fundamental theories that are applicable to signed social networks is social structural balance $[5,6,16]$. Structural balance corresponds to the possibility of exactly dividing the signed graph into two adversary subcommunities such that all edges within each subcommunity have positive weights while all edges joining agents of different communities have negative weights. Obviously, graphs of nonnegative weights are a special case of structural balance, in which one of the two subcommunities is empty. Since the assumption that structural balance exists in a real signed network might be too extreme, a concept called weak structural balance further generalizes structural balance by discussing the multiadversary-subcommunities partition of the signed graph [7].

Structural balance and weak structural balance have been shown to be valid to analyze signed networks. For instance, the sign inference problem, which aims to infer the unknown relationship between two objects, can be solved by mining balance information of signed networks from local and global perspectives [8-10, 12-17]. With the help of the result inferred, it is possible to predict the relationships so that legitimate
participants can eliminate networking security vulnerabilities. Nevertheless, most of these state-of-the-art methods for sign inference problem are mainly considered from a static point of view, and dynamic scenarios are rarely taken into account. Therefore, it is necessary to establish a rational dynamic network model to infer the sign of relationships.

Actually, there exist several inherent qualities of mOSNs that are challenging to reliably sense the global states of relationships for the large networks in practice [2]. First, in contrast to traditional social networks, the observations of relationships in mOSNs are closely associated with the geographical environment, as well as the relative locations and signal coverage of mobile terminals/network access points. Due to these spatial constraints, such observations, which seem linearly related to the global data of relationships (i.e., linearly sampled from the global data), are bound to miss a significant number of values. Consequently, they are not sufficient to unambiguously infer the true status by the traditional solutions of linear-inverse problem in general. Second, in mOSNs different relations between entities may appear at different times. Accordingly, observations of the networks vary during a time period long enough. These dynamic interactions over time essentially introduce time dimension to the problem of mining, the potential relationship structures. Third, despite maintaining the dynamic performance, the underlying relationships in reality always display some "redundancy" attributed to the gradual/periodic variation [3], the relative stability, and so forth. Owing to the aforementioned characteristics of mOSNs, the mass redundant data generated in variant scenarios will result in resource challenge. Hence, although many observers collect features for at least part of the networks, there are still serious impediments to reliable large-scale or network-wide data processing. After these aspects of mOSNs are learned, it is reasonable to organize the entire relationship dataset in the form of tensor coincident with its spatiotemporal structure. Meanwhile, efficient relationship inference approaches associated with the tensor model are required to overcome the obstacles of this data processing.

The aim of this paper is to develop algorithms for the sign inference in signed mOSNs in global and spatiotemporal evolvement perspectives. In particular, we assume that the signed mOSN possesses an underlying dynamic weakly balanced complete network structure. Suppose that we are given an incomplete networking observation tensor (or 3dimensional array), which consists of the adjacency matrices corresponding to the snapshots of the underlying dynamic weakly balanced complete network at times $T_{0}, T_{0}+1, \ldots, T_{0}+$ $T-1, T_{0}+T$. Then the sign inference task is to estimate the sign patterns of all possible links in the dynamic complete network at time $T_{0}+T$. Utilizing the low-rank property of the weak structural balance and the features extracted from the observation tensor, we consider the inference via the incomplete relationship data as an underdetermined linear-inverse problem and develop an approach via a lowrank matrix reconstruction to solve this problem. Moreover, we regard the observation tensor as the training data set and choose a dictionary from it to improve the validity and efficiency of our inference approach. The dictionary
selection method is designed by reducing the size of an overcomplete feature set extracted from the training dataset. Also, a dictionary self-updating mechanism is introduced to improve accuracy of the inference.

Here are the key contributions we make in this paper.
(i) A dictionary selection approach based on group sparsity has been designed to generate a set containing minimal sizes of features to increase computational efficiency. Specifically, the observation tensor is considered to be the raw materials for feature extraction.
(ii) The sign inference problem referring to the weakly balanced mOSNs is formulated as a low-rank matrix reconstruction from the selected dictionary. Under certain mild conditions, a low-rank matrix reconstruction algorithm is applied to solve the sign inference problem, and it turns out to be much more accurate and efficient than other inference methods in the literature. A dictionary self-updating mechanism is also introduced to adjust the dynamic characteristics of the network and improve the sensing accuracy.

The rest of this paper is organized as follows. In Section 2, we build the model of the dynamic signed network. Some basics of balance theory are also reviewed for the sake of integrality. In Section 3, we first extract the initial candidate feature pool from the observation tensor and propose a dictionary selection approach. Then we propose our low-rank matrix reconstruction method to solve the sign inference problem. The implementation details of the dictionary selfupdating procedure are also proposed. In Section 4, we conduct numerical experiments which demonstrate the validity of our network model for sign inference and justify the performance of our methods as well. Finally, we present our conclusions in Section 5.

## 2. Background and Preliminaries

2.1. Dynamic Signed Network Structure. Formally, a dynamic undirected signed network is represented as a dynamic graph $\mathscr{G}=(\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ is the vertex set of size $n$ and $\mathscr{E}$ is the edge set varying over time. A network snapshot denoted by $\delta_{t}=\left(\mathscr{V}, \mathscr{E}_{t}, \mathbf{A}^{(t)}\right)$ presents the connections of $\mathscr{G}$ observed at time $t$. Here, $\mathscr{E}_{t}$ is the subset of $\mathscr{E}$ and $\mathbf{A}^{(t)} \in\{-1,0,1\}^{n \times n}$ is the adjacency matrix of $\mathcal{S}_{t}$ with the signed weights

$$
a_{i j}^{(t)}= \begin{cases}1 & \text { if } i \text { and } j \text { have positive relationship, }  \tag{1}\\ -1 & \text { if } i \text { and } j \text { have negattive relationship, } \\ 0 & \text { if relationship between } i \text { and } j \text { is unknown. }\end{cases}
$$

Particularly, for each time $t$, a zero entry in $\mathbf{A}^{(t)}$ is treated as an unknown relationship based on the acknowledgement that some potential attitudes exist between any two entities, even if the relationship itself is not observed. From this viewpoint, we can assume that there exists an underlying dynamics complete signed network $\overline{\mathscr{G}}$, in which only some partial relationships are observed at times


Figure 1: Illustration of the adjacency tensor, the cube units symbolize the data of relationships: (a) the adjacency tensor of the observed network and (b) the adjacency tensor of the underlying complete network.
$T_{0}, T_{0}+1, \ldots, T_{0}+T-1$, respectively. Correspondingly, we let $\overline{\mathscr{A}} \in\{-1,1\}^{n \times n \times T}$ denote the three-dimensional tensor that contains relationship information between all pairs of entities in $\overline{\mathscr{G}}$. Thus, the observation tensor $\mathscr{A}$ consisting of a series of network snapshots can be represented as

$$
\mathscr{A}_{i, j, t}=a_{i j}^{(t)}= \begin{cases}\overline{\mathscr{A}}_{i, j, t}, & (i, j, t) \in \Omega  \tag{2}\\ 0, & \text { otherwise }\end{cases}
$$

where $\Omega$ is the index set of the observed entries. Let $\mathscr{P}_{\Omega}$ be the orthogonal projection operator onto the span of tensors vanishing outside $\Omega$ so that the ( $i, j, t$ )th component of $\mathscr{P}_{\Omega}(\mathscr{X})$ is equal to $\mathscr{X}_{i, j, t}$ when $(i, j, t) \in \Omega$ and zero otherwise. Then we have $\mathscr{P}_{\Omega}(\overline{\mathscr{A}})=\mathscr{A}$ (shown in Figure 1) and $\mathscr{P}_{\Omega_{t}}\left(\overline{\mathbf{A}}^{(t)}\right)=\mathbf{A}^{(t)}$ for each time slice $t$, where $\bigcap_{t} \Omega_{t}=\emptyset$ and $\bigcup_{t} \Omega_{t}=\Omega$.

While the above kind of signed networks is called homogeneous, that is, relationships of the networks are between the same kinds of entities, a signed network can also be heterogeneous. In a heterogeneous signed network, there can be more than one kind of entities, and relationships between same or different entities can be positive and negative, such as YouTube with two kinds of entities-users and videos. Moreover, this three-dimensional network adjacency tensor can increase dimensions (e.g., spatial dimension, etc.) to adapt to a wider range of scenarios. In this paper, we mainly focus our attention on three-dimensional homogeneous signed networks.
2.2. Weak Structural Balance. Structural balance theory was first formulated by Heider [18] in order to understand the structure in a network of individuals whose mutual relationships are characterized in terms of friendship and hostility. Formally, a triad is considered balanced if the product of the signs in the triad is positive; that is, it contains an even number of negative edges. This is in agreement with principles such as "a friend of my friend is more likely to be my friend" and "an enemy of my friend is more likely to be my
enemy" [6]. The configurations of balanced and unbalanced triads are shown in Figure 2. One possible weakness of this theory is that the defined balance relationships might be too strict. In this perspective, by extending the fundamental beliefs in real networks, weak structural balance is proposed as a way of eliminating the assumption that "the enemy of my enemy is my friend" [7]. Equivalently, the case that "the enemy of my enemy is my enemy" is permitted. Therefore, the local structure of weak balance posits that only triads with exactly two positive edges are implausible and that all other kinds of triads should be permissible (also illustrated in Figure 2).

The formal definition of weakly balanced networks is as follows.

Definition 1 (weakly balanced networks [7]). A (possibly incomplete) network is weakly balanced if and only if it is possible to obtain a weakly balanced complete network by filling the missing edges in its adjacency matrix. Furthermore, in terms of patterns of global structure, a complete network is weakly balanced if and only if the vertex set can be divided into $r$ clusters, $r \geq 1$, such that all the edges within clusters are positive and all the edges between clusters are negative.

There exists the literature discussing the approaches of clustering and sign prediction with respect to signed networks. Ideas derived from local balance of signed networks can be successfully used to yield algorithms for sign inference [ 9,10$]$. Meanwhile, several works analyze the social interrelations from global perspective of structural balance [8, 13-15, 17]. In particular, it is shown in [8] that the adjacency matrix of weakly balanced networks has a "low-rank" structure, and the sign prediction methods based on low-rank modeling were proposed as well.

Theorem 2 (low-rank structure of signed networks [8]). The adjacency matrix $\overline{\mathbf{A}} \in\{1,-1\}^{n \times n}$ of a complete $r$-weakly balanced network has rank 1, if $r \leq 2$, and has rank $r$ for all $r>2$.


Figure 2: Signed undirected connectivity configurations mentioned in Section 2.2: (i) (a) and (b) are balanced triads, (c) and (d) are unbalanced triads, and (ii) (a), (b), and (d) are weakly balanced triads.

Actually, since the global viewpoint of weak balance stated in Definition 1 obeys clustering characteristics presented in Theorem 2, for $\overline{\mathbf{A}}$, there exists an invertible matrix $\mathbf{P}$ such that

$$
\mathbf{P} \overline{\mathrm{A}} \mathbf{P}^{T}=\left(\begin{array}{lllll}
\mathbf{1}_{n_{1}} & & & &  \tag{3}\\
& \mathbf{1}_{n_{2}} & & -1 & \\
& & \mathbf{1}_{n_{3}} & & \\
& -1 & & \ddots & \\
& & & & \mathbf{1}_{n_{r}}
\end{array}\right),
$$

where $\mathbf{1}_{n_{i}}$ on the primary diagonal is an $n_{i}$-order square matrix whose entries are all $1\left(\sum_{i=1}^{r} n_{i}=n\right)$ and the other entries of $\mathbf{P} \overline{\mathbf{A}} \mathbf{P}^{T}$ are all -1 . The $n_{i}$-order square matrix indicates the $i$ th cluster.

Notation. For $\mathbf{X} \in R^{m \times n}$, let the mixed norm $\|\mathbf{X}\|_{2,1}=$ $\sum_{i=1}^{m}\left\|\mathbf{X}_{i .}\right\|_{2} ;$ the soft-thresholding operator $\mathscr{D}_{\tau}(\cdot): \mathbf{X} \in$ $R^{m \times n} \mapsto \mathbf{Y} \in R^{m \times n}$ is also defined obeying

$$
\mathbf{Y}_{i \cdot}= \begin{cases}0 & \text { if }\left\|\mathbf{X}_{i}\right\|_{2} \leq \tau  \tag{4}\\ \left(1-\frac{\tau}{\left\|\mathbf{X}_{i}\right\|_{2}}\right) \mathbf{X}_{i,} & \text { otherwise }\end{cases}
$$

where $\mathbf{X}_{i}$. and $\mathbf{Y}_{i}$. denote the $i$ th row of $\mathbf{X}$ and $\mathbf{Y}$, respectively [15]. The invertible vectorization is denoted by $\operatorname{vec}(\cdot)$ : $R^{m \times n} \mapsto R^{m n}$.

Let $\delta_{\mu, L}^{1,1}\left(R^{m \times n}\right)$ be the class of convex functions with Lipschitz gradient [19]. A continuous differentiable function $f(\mathbf{Y})$ belongs to $\mathcal{S}_{\mu, L}^{1,1}\left(R^{m \times n}\right)$ for some $0 \leq \mu \leq L$ if for any $\mathbf{X}, \mathbf{Y} \in R^{m \times n}$ we have both of the following:

$$
\begin{gather*}
\|\nabla f(\mathbf{X})-\nabla f(\mathbf{Y})\|_{F} \leq L\|\mathbf{X}-\mathbf{Y}\|_{F}  \tag{5}\\
\langle\nabla f(\mathbf{X})-\nabla f(\mathbf{Y}), \mathbf{X}-\mathbf{Y}\rangle \geq \mu\|\mathbf{X}-\mathbf{Y}\|_{F}^{2}
\end{gather*}
$$

## 3. Sign Inference via Dictionary Learning

In this section, we focus on a solution of the sign inference to estimate connection statuses via dictionary learning. As the preparation, we propose a large-scale dictionary selection method to generate the dictionary for inferring. Assume that
we are given a (usually incomplete) network observation tensor $\mathscr{A}$ sampled from an underlying dynamic weakly balanced complete network $\overline{\mathscr{G}}$ with the adjacency tensor $\overline{\mathscr{A}}$. As the description in Section 1, it is reasonable to suppose that most relationships between entities have their own stability in a long period of time in practice and subsequently the change in the scale of each subcommunity is limited. Apparently, this implies the strong dependence retained among the observed data. Combining these assumptions with the lowrank characteristic of weakly balanced complete networks, we extract an initial feature pool from the observation tensor $\mathscr{A}$ and propose a dictionary selection method to compress the scale of the feature pool in Section 3.1. The corresponding algorithm is presented, respectively, in Section 3.2. With the trained dictionary, we propose our sign inference approach and dictionary updating mechanism in Section 3.3, which are also inspired by the low-rank characteristic of weakly balanced complete networks.

The method we propose to handle the dictionary selection is motivated by the Singular Value Thresholding (SVT) algorithm, which is a simple and efficient algorithm for nuclear norm minimization problems proposed by Cai et al. [20]. Our basic idea is to obtain the optimal solution of the trace norm minimization problem by solving its dual problem whose objective function can be shown to be continuously differentiable with Lipschitz continuous gradient. Specifically, we prove that the optimal solution of the primary problem can be readily obtained from the optimal solution of the dual problem. We first provide a brief review of the standard SVT algorithm.

Considering the problem

$$
\begin{align*}
& \min _{\mathbf{X}} \quad \tau\|\mathbf{X}\|_{*}+\frac{1}{2}\|\mathbf{X}\|_{F}^{2}  \tag{6}\\
& \text { subject to } \quad \mathscr{P}_{\Omega}(\mathbf{X})=\mathscr{P}_{\Omega}(\mathbf{M}),
\end{align*}
$$

Cai et al. [20] give a theoretical analysis that, when $\tau \rightarrow \infty$, the optimal solution of problem (6) converges to that of the standard problem:

$$
\begin{array}{ll}
\min _{\mathbf{X}} & \|\mathbf{X}\|_{*}  \tag{7}\\
\text { subject to } \quad \mathscr{P}_{\Omega}(\mathbf{X})=\mathscr{P}_{\Omega}(\mathbf{M}) .
\end{array}
$$

Given that $\tau>0$, the SVT algorithm operates as a linear Bregman iteration scheme. Furthermore, by defining the Lagrangian function of problem (6) as

$$
\begin{equation*}
\mathscr{L}(\mathbf{X}, \mathbf{Y})=\tau\|\mathbf{X}\|_{*}+\frac{1}{2}\|\mathbf{X}\|_{F}^{2}+\left\langle\mathbf{Y}, \mathscr{P}_{\Omega}(\mathbf{M}-\mathbf{X})\right\rangle \tag{8}
\end{equation*}
$$

where $\mathbf{Y}$ is the Lagrangian dual variable, we can derive its dual function as

$$
\begin{equation*}
f(\mathbf{Y})=\inf _{\mathbf{X}} \mathscr{L}(\mathbf{X}, \mathbf{Y}) \tag{9}
\end{equation*}
$$

Cai et al. show that SVT indeed optimizes the dual function $f(\mathbf{Y})$ via the gradient ascent method.
3.1. Large-Scale Dictionary Selection. We address how to select the dictionary given an initial candidate feature pool in this subsection. To this end, we first extract an initial candidate feature pool from $\mathscr{A}$, which is sampled from $\overline{\mathscr{A}}$. Since $\overline{\mathscr{A}}$ consists of the adjacency matrices $\overline{\mathbf{A}}^{(t)}\left(t=T_{0}, T_{0}+\right.$ $\left.1, \ldots, T_{0}+T-1\right)$, the matrix $\mathbf{A}^{(t)}$ in $\mathscr{A}$ can retain the information of $\overline{\mathbf{A}}^{(t)}$ more or less. Thus, we reserve the group of $\mathbf{A}^{(t)}$ with relatively higher sample rate to extract features. We use singular value decomposition (SVD) to express each $\mathbf{A}^{(t)}$ as a series of orthogonal bases in Hilbert space; that is,

$$
\begin{equation*}
\mathbf{A}^{(t)}=\sum_{r=1}^{r_{t}} \sigma_{r}^{(t)} \mathbf{u}_{r}^{(t)}\left(\mathbf{v}_{r}^{(t)}\right)^{T}=\sum_{r=1}^{r_{t}} \sigma_{r}^{(t)} \mathbf{U}_{r}^{(t)} \tag{10}
\end{equation*}
$$

where $\mathbf{u}_{r}^{(t)}$ and $\mathbf{v}_{r}^{(t)}$ are singular vectors of $\mathbf{A}^{(t)}$ with eigenvalue $\sigma_{r}^{(t)}, 1 \leq r \leq r_{t}$. Without loss of generality, we sort $\sigma_{1}^{(t)} \geq \sigma_{2}^{(t)} \geq$ $\cdots \geq \sigma_{r_{t}}^{(t)}$ the singular values of $\mathbf{A}^{(t)}$ in descending order, and set

$$
\begin{equation*}
L_{t}=r_{t}-\arg \max _{r=1, \ldots, r_{t}-1}\left(\sigma_{r}^{(t)} \geq \sigma_{r+1}^{(t)}\right) \tag{11}
\end{equation*}
$$

Then, due to the low-rank property of the weakly balanced complete adjacency matrix, we keep the group of $\mathbf{U}_{r}^{(t)}$ corresponding to the $L_{t}$ largest $\sigma_{r}^{(t)}$ as the features. By this procedure, we extract an initial candidate feature pool as $\left\{\mathbf{U}_{r}^{(t)}: T_{0} \leq t \leq T_{0}+T-1,1 \leq r \leq L_{t}\right\}$, where each matrix $\mathbf{U}_{r}^{(t)} \in R^{n \times n}$ denotes a feature. Equivalently, we can discuss $\mathscr{Q}=\left\{\operatorname{vec}\left(\mathbf{U}_{r}^{(t)}\right): T_{0} \leq t \leq T_{0}+T-1,1 \leq r \leq L_{t}\right\}$ and form the matrix $\Phi=\left[\operatorname{vec}\left(\mathbf{U}_{1}\right), \operatorname{vec}\left(\mathbf{U}_{2}\right), \ldots, \operatorname{vec}\left(\mathbf{U}_{S}\right)\right]$ for convenience, where $\operatorname{vec}\left(\mathbf{U}_{s}\right)=\operatorname{vec}\left(\mathbf{U}_{r}^{(t)}\right), 1 \leq r \leq L_{t}, 1 \leq s \leq S=\sum_{t} L_{t}$, and $T_{0} \leq t \leq T_{0}+T-1$.

Due to massive data of the initial feature pool $\Phi$, we hope to find an optimal subset to form the dictionary $\Psi=$ $\left[\operatorname{vec}\left(\mathbf{U}_{1}\right), \operatorname{vec}\left(\mathbf{U}_{2}\right), \ldots, \operatorname{vec}\left(\mathbf{U}_{K}\right)\right]$ such that the set $\Phi$ can be well reconstructed by $\Psi$ and the size of $\Psi$ is as small as possible. To achieve this goal, we select $\Psi$ such that the rest of the features in $\Phi$ can be well reconstructed using it. Analogous to the optimization problem in [21], the basic problem is formulated as follows:

$$
\begin{equation*}
\min _{\mathbf{X}} \quad\|\mathbf{X}\|_{2,1} \tag{12}
\end{equation*}
$$

$$
\text { subject to } \quad \Phi \mathbf{X}=\Phi
$$

where $\Phi \in R^{N \times S}\left(N=n^{2}\right), \mathbf{X} \in R^{S \times S}$, and $\|\mathbf{X}\|_{2,1}=$ $\sum_{i=1}^{S}\left\|\mathbf{X}_{i \cdot} \cdot\right\|_{2}$. Apparently, $\|\mathbf{X}\|_{2,1}$ enforces the group sparsity on the variable $\mathbf{X}$ and the optimal solution usually contains zero rows. This means that not all features in $\Phi$ are necessary to be selected to reconstruct any data sample.

Motivated by SVT, we have the equivalent problem of (12) as follows:

$$
\begin{align*}
& \min _{\mathbf{X}} \quad\|\mathbf{X}\|_{2,1}+\frac{1}{2}\|\mathbf{X}\|_{F}^{2}  \tag{13}\\
& \text { subject to } \quad \Phi \mathbf{X}=\Phi
\end{align*}
$$

The Lagrangian function of problem (13) is defined as

$$
\begin{equation*}
\mathscr{L}(\mathbf{X}, \mathbf{Y})=\tau\|\mathbf{X}\|_{2,1}+\frac{1}{2}\|\mathbf{X}\|_{F}^{2}+\langle\mathbf{Y}, \Phi-\Phi \mathbf{X}\rangle \tag{14}
\end{equation*}
$$

and its dual function is

$$
\begin{equation*}
f(\mathbf{Y})=\inf _{\mathbf{X}} \mathscr{L}(\mathbf{X}, \mathbf{Y}) . \tag{15}
\end{equation*}
$$

We first examine the properties of the dual function $f(\mathbf{Y})$ and then show how to achieve the optimal solution of the problem (13) from its dual optimum directly. As the mixed norm $\|\mathbf{X}\|_{2,1}$ is not differentiable, it is difficult to optimize the dual function $f(\mathbf{Y})$ directly. However, we can obtain a useful property of the dual function $f(\mathbf{Y})$ as follows.

Theorem 3. For all $\tau \geq 0$, the dual function $f(\mathbf{Y})$ is continuously differentiable with Lipschitz continuous gradient at most M. Furthermore, the primal optimal $\widehat{\mathbf{X}}$ of the problem (13) is given by

$$
\begin{equation*}
\widehat{\mathbf{X}}=\mathscr{D}_{\tau}(\Phi \widehat{\mathbf{Y}}) \tag{16}
\end{equation*}
$$

when the dual optimal $\widehat{\mathbf{Y}}$ of the problem (13) is obtained.
The proof of Theorem 3 is based on the following results.
Lemma 4. For each $\tau \geq 0$ and $\mathbf{Y} \in R^{m \times n}$, one has

$$
\begin{equation*}
\mathscr{D}_{\tau}(\mathbf{Y})=\arg \min _{\mathbf{X}} \tau\|\mathbf{X}\|_{2,1}+\frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_{F}^{2} \tag{17}
\end{equation*}
$$

As a matter of fact, considering the following optimization problem:

$$
\begin{equation*}
\min _{x \in R} \tau|x|+\frac{1}{2}(x-y)^{2} \tag{18}
\end{equation*}
$$

it is easy to show that the unique solution admits a closed form called the soft-thresholding operator, following a terminology introduced by Donoho and Johnstone [22]; it can be written that

$$
\widehat{y}= \begin{cases}0 & \text { if }|x| \leq \tau  \tag{19}\\ \left(1-\frac{\tau}{|x|}\right) x & \text { otherwise }\end{cases}
$$

Thus, from a generalized view, one has Lemma 4.
Also, the following result can be deduced based on the properties of Moreau-Yosida regularization [23].

Lemma 5. For any $\mathbf{X}, \mathbf{Y} \in R^{m \times n}$, one has

$$
\begin{equation*}
\left\|\mathscr{D}_{\tau}(\mathbf{X})-\mathscr{D}_{\tau}(\mathbf{Y})\right\|_{F}^{2} \leq\left\langle\mathscr{D}_{\tau}(\mathbf{X})-\mathscr{D}_{\tau}(\mathbf{Y}), \mathbf{X}-\mathbf{Y}\right\rangle . \tag{20}
\end{equation*}
$$

It follows that $\mathscr{D}_{\tau}(\mathbf{Y})$ is globally Lipschitz continuous with modulus 1.

Proof of Theorem 3. Since

$$
\begin{align*}
f(\mathbf{Y})= & \inf _{\mathbf{X}} \mathscr{L}(\mathbf{X}, \mathbf{Y}) \\
= & \inf _{\mathbf{X}}\left(\tau\|\mathbf{X}\|_{2,1}+\frac{1}{2}\|\mathbf{X}\|_{F}^{2}+\langle\mathbf{Y}, \Phi-\Phi \mathbf{X}\rangle\right) \\
= & \inf _{\mathbf{X}}\left(\tau\|\mathbf{X}\|_{2,1}+\frac{1}{2}\|\mathbf{X}\|_{F}^{2}+\langle\mathbf{Y}, \Phi\rangle-\left\langle\Phi^{T} \mathbf{Y}, \mathbf{X}\right\rangle\right) \\
= & \inf _{\mathbf{X}}\left(\tau\|\mathbf{X}\|_{2,1}+\frac{1}{2}\left\|\mathbf{X}-\Phi^{T} \mathbf{Y}\right\|_{F}^{2}\right)  \tag{21}\\
& +\langle\mathbf{Y}, \Phi\rangle-\frac{1}{2}\left\|\Phi^{T} \mathbf{Y}\right\|_{F}^{2} \\
= & g(\mathbf{Y})+\langle\mathbf{Y}, \Phi\rangle-\frac{1}{2}\left\|\Phi^{T} \mathbf{Y}\right\|_{F}^{2}
\end{align*}
$$

and $g(\mathbf{Y})$ is the Moreau-Yosida regularization of the mixed norm $\|\cdot\|_{2,1}$, using the well-known properties of MoreauYosida regularization [23], we get the results that $g(\mathbf{Y})$ is a globally continuously differentiable convex function. Moreover, $\nabla \boldsymbol{g}(\mathbf{Y})=\Phi\left(\Phi^{T} \mathbf{Y}-\mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}\right)\right)$ and $\nabla \boldsymbol{g}(\mathbf{Y})$ is continuously differentiable with Lipschitz continuous gradient $\rho$; that is, for any $\mathbf{Y}_{1}, \mathbf{Y}_{2} \in R^{N \times S}$,

$$
\begin{align*}
\left\|\nabla g\left(\mathbf{Y}_{1}\right)-\nabla g\left(\mathbf{Y}_{2}\right)\right\|_{F} & \leq\left\|\Phi^{T}\left(\mathbf{Y}_{1}-\mathbf{Y}_{2}\right)\right\|_{F}  \tag{22}\\
& \leq \rho\left\|\left(\mathbf{Y}_{1}-\mathbf{Y}_{2}\right)\right\|_{F}
\end{align*}
$$

where $\rho=\sup _{\|\mathbf{Z}\|_{F}=1, \mathbf{Z} \in R^{N \times S}}\left\|\Phi^{T} \mathbf{Z}\right\|_{F}$. Then the gradient of $f(\mathbf{Y})$ can be obtained as follows:

$$
\begin{align*}
\nabla f(\mathbf{Y}) & =\nabla g(\mathbf{Y})+\Phi-\Phi \Phi^{T} \mathbf{Y} \\
& =\Phi\left(\Phi^{T} \mathbf{Y}-\mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}\right)\right)+\Phi-\Phi \Phi^{T} \mathbf{Y}  \tag{23}\\
& =\Phi-\Phi \mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}\right) .
\end{align*}
$$

It follows that, for any $\mathbf{Y}_{1}, \mathbf{Y}_{2} \in R^{N \times S}$,

$$
\begin{aligned}
\| \nabla & f_{2}\left(\mathbf{Y}_{1}\right)-\nabla f_{2}\left(\mathbf{Y}_{2}\right) \|_{F} \\
& =\left\|\Phi-\Phi \mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}_{1}\right)-\Phi+\Phi \mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}_{2}\right)\right\|_{F} \\
& =\left\|\Phi\left(\mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}_{1}\right)-\mathscr{D}_{\tau}\left(\Phi^{T} \mathbf{Y}_{2}\right)\right)\right\|_{F} \\
& \leq\left\|\Phi \Phi^{T}\left(\mathbf{Y}_{1}-\mathbf{Y}_{2}\right)\right\|_{F} \\
& \leq M\left\|\mathbf{Y}_{1}-\mathbf{Y}_{2}\right\|_{F},
\end{aligned}
$$

where the first inequality follows from (20) and $M=$ $\sup _{\|\mathbf{Z}\|_{F}=1, \mathbf{Z} \in R^{N \times S}}\left\|\Phi \Phi^{T} \mathbf{Z}\right\|_{F}$. When the dual optimal $\widehat{\mathbf{Y}}$ is obtained, by using the result of (21), we can get

$$
\begin{align*}
\widehat{\mathbf{X}} & =\arg \min _{\mathbf{X}} \mathscr{L}(\mathbf{X}, \mathbf{Y}) \\
& =\arg \min _{\mathbf{X}}\left(\tau\|\mathbf{X}\|_{2,1}+\frac{1}{2}\left\|\mathbf{X}-\Phi^{T} \mathbf{Y}\right\|_{F}^{2}\right)  \tag{25}\\
& =\mathscr{D}_{\tau}(\Phi \widehat{\mathbf{Y}}) .
\end{align*}
$$

This concludes the proof.
Since $f(\mathbf{Y})$ is the dual function of the objective function (13), $f(\mathbf{Y})$ is concave. Let

$$
\begin{align*}
q(\mathbf{Y})= & -f(\mathbf{Y}) \\
=- & \left(\tau\left\|\mathscr{D}_{\tau}(\Phi \mathbf{Y})\right\|_{2,1}\right.  \tag{26}\\
& \left.+\frac{1}{2}\left\|\mathscr{D}_{\tau}(\Phi \mathbf{Y})\right\|_{F}^{2}+\left\langle\mathbf{Y}, \Phi-\Phi \mathscr{D}_{\tau}(\Phi \mathbf{Y})\right\rangle\right),
\end{align*}
$$

which is convex. Thus, the following holds for any $\mathbf{Y}_{1}, \mathbf{Y}_{2} \in$ $R^{N \times S}$ :

$$
\begin{equation*}
\left\langle q\left(\mathbf{Y}_{1}\right)-q\left(\mathbf{Y}_{2}\right), \mathbf{Y}_{1}-\mathbf{Y}_{2}\right\rangle \geq 0 . \tag{27}
\end{equation*}
$$

It is also easy to show that $q(\mathbf{Y})$ belongs to the class $\mathcal{S}_{0, M}^{1,1}\left(R^{N \times S}\right)$ and

$$
\begin{equation*}
\nabla q(\mathbf{Y})=-\Phi\left(\mathbf{I}-\mathscr{D}_{\tau}\left(\Phi^{T} \widehat{\mathbf{Y}}\right)\right) \tag{28}
\end{equation*}
$$

where $\mathbf{I} \in R^{S \times S}$ is the identity matrix. Therefore, we can solve problem (13) by minimizing the objective function $q(\mathbf{Y})$; that is,

$$
\begin{equation*}
\min _{\mathbf{Y}} q(\mathbf{Y}) . \tag{29}
\end{equation*}
$$

Therefore, the dictionary $\Psi$ is selected by the optimal solution $\widehat{\mathbf{Y}}$; that is, the $i$ th column of $\Phi$ is chosen to be the atom of $\Psi$ if $\left\|\widehat{\mathbf{Y}}_{i \cdot}\right\|_{2} \neq 0$. The optimization algorithm is presented in the next subsection.
3.2. Optimization Methods. In this subsection, we develop an efficient optimization algorithm to solve the dual problem (29). Because the objective function $q(\mathbf{Y})$ is continuously differentiable with Lipschitz continuous gradient, it is feasible to utilize gradient-based optimization methods to achieve the optimal solution for their simplicity and low complexity within each iteration. However, classical gradient-based methods for functions with Lipschitz continuous gradient converge at a rate of $O(1 / N)$, where $N$ is the number of iterations during optimization [19]. In fact, this is too slow especially when dealing with large-scale datasets. Note that Nesterov showed in his work [24] that an accelerated gradient algorithm can be constructed such that $O\left(1 / N^{2}\right)$, the lower bound on the convergence rate for gradient-based methods [25], is achieved when minimizing unconstrained
smooth functions. With this consideration, in the following we propose an accelerated thresholding algorithm to solve these smooth convex optimization problems using Nesterov's method with an adaptive line search scheme [19, 26].

We recall Nesterov's method with an adaptive line search scheme as follows. Take the unconstrained smooth convex minimization problem $\min _{\mathbf{y} \in R^{n}} q(\mathbf{y})$, for instance, where $q(\mathbf{y})$ belongs to $\mathcal{S}_{\mu, L}^{1,1}\left(R^{n}\right), \mu \geq 0$, and $L<+\infty$. Nesterov's method for this problem utilizes two sequences: $\left\{\mathbf{y}_{l}\right\}$ and $\left\{\mathbf{s}_{l}\right\}, \mathbf{y}_{l}, \mathbf{s}_{l} \in$ $R^{n}$. The searching point $\mathbf{s}_{l}$ satisfies

$$
\begin{equation*}
\mathbf{s}_{l}=\mathbf{y}_{l}+\beta_{l}\left(\mathbf{y}_{l}-\mathbf{y}_{l-1}\right), \tag{30}
\end{equation*}
$$

where $\beta_{l}$ is a tuning parameter. The approximate solution $\mathbf{y}_{l+1}$ can be computed as a gradient step of $\mathbf{s}_{l}$ as

$$
\begin{equation*}
\mathbf{y}_{l+1}=\mathbf{s}_{l}-\frac{1}{L_{l}} \nabla q\left(\mathbf{s}_{l}\right) \tag{31}
\end{equation*}
$$

where $1 / L_{l}$ is the step size. Starting from an initial point $y_{0}$, $\mathbf{s}_{l}$ and $\mathbf{y}_{l+1}$ can be computed recursively according to (30) and (31) and can arrive at the optimal solution $\widehat{\mathbf{y}}$. Although it has been shown that Nesterov's method is a very powerful optimization technique for class $\mathcal{S}_{\mu, L}^{1,1}\left(R^{n}\right)$ [19], how to choose $\beta_{l}$ and $1 / L_{l}$ in each iteration is a critical issue in Nesterov's method. When they are set properly, the sequence $\left\{\mathbf{y}_{l}\right\}$ can converge to the optimal $\hat{\mathbf{y}}$ at a certain convergence rate. As a well-known scheme for setting $\beta_{l}$ and $L_{l}$, Nesterov's constant scheme assumes $\beta_{l}$ and $L_{l}$ to be constant [19], while Nemirovski's line search scheme requires $L_{l}$ to monotonically increase, and $\beta_{l}$ is independent of $L_{l}$ [27]. Both of the settings result in slow convergence.

To overcome this drawback, an adaptive line search scheme for Nesterov's method is proposed in [26]. Under the assumption that $\tilde{\mu}$, the low bound of $\mu$, is known in advance, this scheme is built upon the estimate sequence [19] defined as follows.

Definition 6 (estimate sequence [19]). A pair of sequences $\left\{\phi_{l}(\mathbf{y})\right\}$ and $\left\{\lambda_{l} \geq 0\right\}$ is called an estimate sequence of function $q(\mathbf{y})$ if $\lim _{k \rightarrow \infty} \lambda_{k}=0$ and $\phi_{l}(\mathbf{y}) \leq\left(1-\lambda_{l}\right) q(\mathbf{y})+\lambda_{l} \phi_{o}(\mathbf{y})$, for all $\mathbf{y} \in R^{n}$.

The estimate sequence defined in Definition 6 has the following important property.

Theorem 7 (see [19]). $\operatorname{Let}\left\{\phi_{l}(\mathbf{y})\right\}$ and $\left\{\lambda_{k} \geq 0\right\}$ be an estimate sequence. For any sequence $\left\{\mathbf{y}_{l}\right\}, q\left(\mathbf{y}_{l}\right)-\widehat{q} \leq \lambda_{k}\left(\phi_{0}(\widehat{\mathbf{y}})-\widehat{q}\right) \rightarrow 0$ if $q\left(\mathbf{y}_{l}\right) \leq \widehat{\phi}_{k} \equiv \min _{\mathbf{y} \in R^{n}} \phi_{k}(\mathbf{y})$, where $\hat{q}$ is the optimal objective function value.

We further specify the estimation sequence in [19]:

$$
\begin{equation*}
\phi_{l}(\mathbf{y})=\widehat{\phi}_{l}+\frac{\gamma_{l}}{2}\left\|\mathbf{y}-\mathbf{v}_{l}\right\|^{2} \tag{32}
\end{equation*}
$$

where the sequences $\left\{\gamma_{l}\right\},\left\{\mathbf{v}_{l}\right\}$, and $\left\{\hat{\phi}_{l}\right\}$ satisfy

$$
\begin{align*}
& \mathbf{v}_{l+1}= \frac{1}{\gamma_{l+1}}\left(\left(1-\alpha_{l}\right) \gamma_{l} \mathbf{v}_{l}+\tilde{\mu} \alpha_{l} \mathbf{s}_{l}-\alpha_{l} \nabla q\left(\mathbf{s}_{l}\right)\right) \\
& \hat{\phi}_{l+1}=\left(1-\alpha_{l}\right) \gamma_{l}+\tilde{\mu} \alpha_{l} \\
&=\left(1-\alpha_{l}\right) \widehat{\phi}_{l}+\alpha_{l} q\left(\mathbf{s}_{l}\right)  \tag{33}\\
&-\frac{\alpha_{l}^{2}}{2 \gamma_{l+1}}\left\|\nabla q\left(\mathbf{s}_{l}\right)\right\|^{2}+\frac{\alpha_{l}\left(1-\alpha_{l}\right) \gamma_{l}}{\gamma_{l+1}} \\
& \times\left(\frac{\tilde{\mu}}{2}\left\|\mathbf{v}_{l}-\mathbf{s}_{l}\right\|^{2}+\left\langle\nabla q\left(\mathbf{s}_{l}\right), \mathbf{v}_{l}-\mathbf{s}_{l}\right\rangle\right)
\end{align*}
$$

Then Algorithm 2 in [26] is proposed by modifying Nemirovski's line search scheme with the adaptive parameters of this sequence, which satisfy Theorem 7.

Note that Theorem 3 indicates that the objective function $q(\mathbf{Y})$ satisfies the conditions of using Nesterov's method with an adaptive line search scheme. Therefore we directly extend Algorithm 2 in [26] to the high-dimensional scenarios to solve (29). The complete procedures are summarized in Algorithm 1.

In Algorithm 1, the while loop from Step 4 to Step 13 is designed to choose a proper step size to satisfy Step 8. As the Lipschitz gradient of $q(\mathbf{Y})$ is $M, L_{l}$ is upper bounded by $2 M$ since Step 8 always holds when $L_{l} \geq M$ [27]. In Step 14, we initialize $L_{l+1}=h(\theta) L_{l}$, where

$$
h(\theta)= \begin{cases}1, & 1 \leq \theta \leq 5  \tag{34}\\ 0.8, & \theta>5\end{cases}
$$

and $\theta>1$ due to the condition in Step 8 [26]. Apparently, when $\theta$ is large, $L_{l+1}$ can be adjusted to avoid the step size $1 / L_{l}$ becoming too small, which may slow down the convergence rate.
3.3. Sign Inference and Dictionary Update Mechanism. This subsection details how to use the dictionary to solve the sign inference problem. Actually, this problem bears similarity to the sign prediction problem in the static signed networks or the unsigned networks varying periodically $[3,8,11,12]$. In this paper, we intend to infer the unknown relationship between a pair of entities $i$ and $j$ based on partial relationship observations of the entire dynamic network at time $T_{0}+T$. We expect to accomplish this task with the help of the dictionary constructed by the relationship data for times $T_{0}$ through $T_{0}+T-1$. As aforementioned, there exists strong dependence between the connection status at time $T_{0}+T$ and the history relationship dataset in the dynamic network. We formulate the sign inference problem as follows:

$$
\begin{equation*}
\widehat{\mathbf{x}}=\arg \min _{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\Psi \mathbf{x}\|_{2}^{2}+\|\mathbf{x}\|_{1}, \tag{35}
\end{equation*}
$$

where $\Psi$ is the dictionary and $\mathbf{y}$ is the invertible vectorization of the matrix $\mathbf{A}^{\left(T_{0}+T\right)}$ observed at time $T_{0}+T$; that is, $\mathbf{y}=\operatorname{vec}\left(\mathbf{A}^{\left(T_{0}+T\right)}\right)$. Because $\overline{\mathbf{A}}^{\left(T_{0}+T\right)}=\sum_{r} \sigma_{r}^{\left(T_{0}+T\right)} \mathbf{U}_{r}^{\left(T_{0}+T\right)}$ by using SVD and subsequently
(1) Input: $\widetilde{\mu}, \alpha_{-1}=0.5, \mathbf{Y}_{-1}=\mathbf{Y}_{0}, L_{-1}=L_{0}, \gamma_{0} \geq \widetilde{\mu}, \lambda_{0}=1$
(2) Output: $\mathbf{Y}_{N}$
(3) for $l=1,2, \ldots, N$ do
(4)
while true do
compute $\alpha_{k} \in(0,1)$ as the root of $L_{l} \alpha_{l}^{2}=\left(1-\alpha_{l}\right) \gamma_{l}+\alpha_{l} \tilde{\mu}$,
$\gamma_{l+1}=\left(1-\alpha_{l}\right) \gamma_{l}+\alpha_{l} \tilde{\mu}, \beta_{l}=\frac{\left(1-\alpha_{l-1}\right) \gamma_{l}}{\left(\gamma_{l}+L_{l} \alpha_{l}\right) \alpha_{l-1}}$;
(6) compute $\mathbf{S}_{l}=\mathbf{Y}_{l}+\beta_{l}\left(\mathbf{Y}_{l}-\mathbf{Y}_{l-1}\right)$
(7) compute $\mathbf{Y}_{l+1}=\mathbf{S}_{l}-\left(1 / L_{l}\right) \nabla q\left(\mathbf{S}_{l}\right)$
if $q\left(\mathbf{Y}_{l+1}\right) \leq q\left(\mathbf{S}_{l}\right)-\left(1 / 2 L_{l}\right)\left\|\nabla q\left(\mathbf{S}_{l}\right)\right\|_{F}^{2}$ then goto Step 14
else
$L_{l}=2 L_{l}$
end if
end while
(14) set $\theta=2 L_{l} \frac{q\left(\mathbf{S}_{l}\right)-q\left(\mathbf{Y}_{l+1}\right)}{\left\|\nabla q\left(\mathbf{S}_{l}\right)\right\|_{F}^{2}}, L_{l+1}=h(\theta) L_{l}$
(15) set $\lambda_{l+1}=\left(1-\alpha_{l}\right) \lambda_{l}$
(16) end for

Algorithm 1: Adaptive line search scheme for dictionary selection.
$\operatorname{vec}\left(\overline{\mathbf{A}}^{\left(T_{0}+T\right)}\right)=\sum_{r} \sigma_{r}^{\left(T_{0}+T\right)} \operatorname{vec}\left(\mathbf{U}_{r}^{\left(T_{0}+T\right)}\right)$, we will estimate $\overline{\mathbf{A}}^{\left(T_{0}+T\right)}$ in the form of vector and transform the low-rank matrix reconstruction problem into a traditional $l_{1}$-norm minimization problem in compressive sensing. We solve (35) by applying backtracking-based adaptive orthogonal matching pursuit (BAOMP) method, which incorporates a simple backtracking technique to detect the previously chosen atoms' reliability and then deletes the unreliable atoms at each iteration [28]. Then we force that $a_{i j}^{\left(T_{0}+T\right)}$, the element of the resulting matrix, is equal to 1 if $a_{i j}^{\left(T_{0}+T\right)}>1$ or equal to -1 if $a_{i j}^{\left(T_{0}+T\right)} \leq 0$, to ensure the elements coinciding with the value setting of relationships.

Furthermore, assume that we are given a sequence input samples $\mathbf{Y}=\left[\mathbf{y}^{\left(T_{0}+T\right)}, \mathbf{y}^{\left(T_{0}+T+1\right)}, \ldots, \mathbf{y}^{(\widetilde{T})}\right]$, where $\mathbf{y}^{(t)}=$ $\operatorname{vec}\left(\mathbf{A}^{(t)}\right), T_{0}+T \leq t \leq \widetilde{T}$, the task of the sign inference becomes to reconstruct the complete adjacency matrices $\overline{\mathbf{A}}^{(t)}$ one by one. Since the $\overline{\mathbf{A}}^{(t)}$ may contain some features which are not included in dictionary, it is necessary to add these features into the dictionary to increase the accuracy of the inference. However, the inferred matrix is not the original matrix exactly and consequently the unobserved relationships are not really known. In contrast, the observed adjacency matrix $\mathbf{A}^{(t)}$ retains all existing relationships. For this reason, we only use $\mathbf{A}^{(t)}$ to extract the features rather than the optimal solution of (35). We apply the extracting approach in Section 3.2 and add the complementary features into the dictionary. Note that this operation will continuously increase the scale of the dictionary while the samples keep inputting for inference; the dictionary selection approach proposed in Section 3.2 will be applied to compact the dictionary once the size of the dictionary exceeds a predetermined bound.

## 4. Numerical Experiment

In this section, we perform experiments on synthetic networks and show that our low-rank model and dictionary learning method outperform other methods on the task of the sign inference for dynamic signed networks. To ensure that our results are reliable, we conduct all experiments 20 times and average out the results from all of the trials.

To construct synthetic networks, we first consider a weakly balanced complete network $\overline{\mathscr{G}}$ whose adjacency tensor is $\overline{\mathscr{A}}$. The slide of $\overline{\mathscr{A}}$ at time $t$ is an adjacency matrix $\overline{\mathbf{A}}^{(t)}$ in the form of (3). In addition, only a few patterns of $\overline{\mathbf{A}}^{(t)}$ exist in $\overline{\mathscr{A}}$. The observation tensor $\mathscr{A}$ is formed by sampling some entries from $\overline{\mathscr{A}}$. Concretely, we let the adjacency tensor $\overline{\mathscr{A}}$ of $\overline{\mathscr{G}}$ consist of $50250 \times 250$ matrices of complete 4 weakly balanced structure. For the network $\overline{\mathscr{G}}$, four clusters are generated randomly. The size of each cluster is larger than 20 and the sum of the sizes is 250 . We further assume that only a part of network relationships is observed by uniform sampling with probability $p \in(0,1)$. It results in $n^{2} p$ entries being randomly sampled from $\overline{\mathbf{A}}^{(t)}$, where $p$ is the fraction of observed entries. We choose a set of matrices whose lost rates are from 0.05 to 0.55 and apply the approach proposed in Section 3.2 to select the dictionary $\Psi$.

With the dictionary $\Psi$ and the given observed matrix $\mathbf{A}^{(t)}$ at time $t \geq T_{0}+T$, the task of the sign inference is achieved by solving (35). We use BAOMP to estimate the complete matrix $\overline{\mathbf{A}}^{(t)}$ and compare the performance of our approach to two state-of-the-art methods, alternating least square (ALS) [29] and singular value projection (SVP) [30], for the sign inference problem. Different from accuracy defined by the relative error on the observed set in [8], we utilize the


Figure 3: Accuracy of sign inference algorithms on synthetic datasets. In general, we can see that dictionary learning outperforms ALS and SVP.


Figure 4: An example of the sign inference. (a) illustrates the original matrix. Given the matrix with $98 \%$ lost-rate, (b) is the result inferred by dictionary learning method. The similarity of inferred matrix is 0.9347 .
similarity between the inferred matrix and the original one to indicate the accuracy of estimation. The definition of the similarity is $\left|\left\langle\overline{\mathbf{A}}^{(t)}, \widehat{\mathbf{A}}^{(t)}\right\rangle\right| /\left\|\overline{\mathbf{A}}^{(t)}\right\|_{F}\left\|\widehat{\mathbf{A}}^{(t)}\right\|_{F}$. We vary the lostrate of the original matrix $\overline{\mathbf{A}}^{(t)}$ from 0.5 to 0.999 and plot the inference accuracy in Figure 3 (lost-rate: $0.5,0.6,0.7,0.8$, $0.9,0.95,0.96,0.97,0.98,0.99,0.995$, and 0.999). Apparently, dictionary learning outperforms ALS and SVP. To present our result more clearly, we also use a visual expression in which the white pixels represent 1 and the black pixels represent -1 . Figure 4 shows one example of the sign inference and we find that relationships and the clusters can almost be accurately estimated by our inference approach.

## 5. Conclusion

In this paper, we establish a low-rank tensor model for the dynamic weakly balanced signed networks. With this model, we first extract the feature pool and propose an approach to extract the compact dictionary from pool. To improve the performance of the selection approach, we derive the corresponding dual problem and introduce an accelerated thresholding algorithm to solve the dual problem. Consequently, the optimal solution of the primary problem can be readily obtained from optimizing the dual problem. In addition, combined with the compact dictionary generation method, the sign inference approach is provided for estimating
missing relationships of the dynamic weakly balanced signed networks at a certain time slice. Also, the approach is endowed with the function of the dictionary updating if relationship statuses change.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was jointly supported by National 973 Program of China (no. 2013CB329204), National 863 Program of China (no. 2011AA01A104), National Natural Science Foundation of China (no. 61100206), Research Fund for Doctoral Program of Higher Education of China (no. 20120005130001), China, and Fund of State Key Laboratory of Information Photonics and Optical Communications (Beijing University of Posts and Telecommunications), China.

## References

[1] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: challenges and opportunities," IEEE Network, vol. 24, no. 4, pp. 13-18, 2010.
[2] N. Jabeur, S. Zeadally, and B. Sayed, "Mobile social networking applications," Communications of the ACM, vol. 56, no. 3, pp. 71-79, 2013.
[3] D. M. Dunlavy, T. G. Kolda, and E. Acar, "Temporal link prediction using matrix and tensor factorizations," ACM Transactions on Knowledge Discovery from Data, vol. 5, no. 2, article 10, 2011.
[4] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," Nature, vol. 453, no. 7191, pp. 98-101, 2008.
[5] F. Harary, "On the notion of balance of a signed graph," The Michigan Mathematical Journal, vol. 2, no. 2, pp. 143-146, 1953.
[6] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," Psychological Review, vol. 63, no. 5, pp. 277-293, 1956.
[7] J. A. Davis, "Clustering and structural balance in graphs," Human Relations, vol. 20, no. 2, pp. 181-187, 1967.
[8] C. J. Hsieh, K. Y. Chiang, and I. S. Dhillon, "Low rank modeling of signed networks," in Proceedings of the Knowledge Discovery and Data Mining Conference (KDD '12), 2012.
[9] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in Proceedings of the 19th International World Wide Web Conference (WWW '10), pp. 641-650, April 2010.
[10] K.-Y. Chiang, N. Natarajan, A. Tewari, and I. S. Dhillon, "Exploiting longer cycles for link prediction in signed networks," in Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM '11), pp. 1157-1162, October 2011.
[11] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," Journal of the American Society for Information Science and Technology, vol. 58, no. 7, pp. 1019-1031, 2007.
[12] L. Lü and T. Zhou, "Link prediction in complex networks: a survey," Physica A, vol. 390, no. 6, pp. 1150-1170, 2011.
[13] M. Szell, R. Lambiotte, and S. Thurner, "Multirelational organization of large-scale social networks in an online world," Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 31, pp. 13636-13641, 2010.
[14] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: mining a social network with negative edges," in Proceedings of the 18th International World Wide Web Conference (WWW '09), pp. 741-750, ACM, 2009.
[15] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10 ), pp. 1361-1370, ACM, April 2010.
[16] J. A. Davis, "Structural balance, mechanical solidarity, and interpersonal relations," The American Journal of Sociology, vol. 68, no. 4, pp. 444-462, 1963.
[17] A. Srinivasan, "Local balancing influences global structure in social networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 5, pp. 17511752, 2011.
[18] F. Heider, "Attitudes and cognitive organization," The Journal of Psychology, vol. 21, no. 1, pp. 107-112, 1946.
[19] Y. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course, Kluwer Academic, Boston, Mass, USA, 2003.
[20] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," SIAM Journal on Optimization, vol. 20, no. 4, pp. 1956-1982, 2010.
[21] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11), pp. 3449-3456, June 2011.
[22] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," Journal of the American Statistical Association, vol. 90, no. 432, pp. 1200-1224, 1995.
[23] J. B. Hiriart-Urruty and C. Lemarechal, Convex Analysis and Minimization Algorithms, vol. 2, Springer, Heidelberg, Germany, 2nd edition, 1996.
[24] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O\left(1 / k^{2}\right)$," Soviet Mathematics Doklady, vol. 27, no. 2, pp. 372-376, 1983.
[25] A. Nemirovsky and D. Yudin, Informational Complexity and Efficient Methods for Solution of Convex Extremal Problems, John Wiley \& Sons, New York, NY, USA, 1983.
[26] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09), pp. 547-555, July 2009.
[27] A. Nemirovski, "Efficient Methods in Convex Programming," Lecture Notes, 1994.
[28] H. Huang and A. Makur, "Backtracking-based matching pursuit method for sparse signal reconstruction," IEEE Signal Processing Letters, vol. 18, no. 7, pp. 391-394, 2011.
[29] Y. Koren, R. M. Bell, and C. Volinsky, "Matrixfactorization techniques for recommender systems," IEEE Computer, vol. 42, no. 8, pp. 30-37, 2009.
[30] P. Jain, R. Meka, and I. Dhillon, "Guaranteed rank minimization via singular value projection," in Proceedings of the 24th Annual Conference on Neural Information Processing Systems (NIPS '10), pp. 934-945, December 2010.

# Adaptive Correction Forecasting Approach for Urban Traffic Flow Based on Fuzzy c-Mean Clustering and Advanced Neural Network 

He Huang, ${ }^{1}$ Qifeng Tang, ${ }^{2}$ and Zhen Liu ${ }^{1}$<br>${ }^{1}$ Shanghai Urban-Rural Construction and Transportation Development Institute, Shanghai 300032, China<br>${ }^{2}$ Department of Automation, Shanghai Jiao Tong University, Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240, China

Correspondence should be addressed to Qifeng Tang; tqf001@qq.com
Received 5 July 2013; Revised 8 October 2013; Accepted 9 October 2013
Academic Editor: Baocang Ding
Copyright © 2013 He Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Forecasting of urban traffic flow is important to intelligent transportation system (ITS) developments and implementations. The precise forecasting of traffic flow will be pretty helpful to relax road traffic congestion. The accuracy of traditional single model without correction mechanism is poor. Summarizing the existing prediction models and considering the characteristics of the traffic itself, a traffic flow prediction model based on fuzzy $c$-mean clustering method (FCM) and advanced neural network (NN) was proposed. FCM can improve the prediction accuracy and robustness of the model, while advanced NN can optimize the generalization ability of the model. Besides these, the output value of the model is calibrated by the correction mechanism. The experimental results show that the proposed method has better prediction accuracy and robustness than the other models.


## 1. Introduction

Real-time forecasting of traffic flow is an important issue in advanced traffic management [1]. The traffic simulation is correspondingly needed to make these forecasting models reliable way, which aim to influence travel behavior, reduce traffic congestion, improve mobility, and enhance air quality. Traffic forecasting models can be used to provide urban traffic control centers with an automated tool for anticipating the congestion that may arise on road facilities and its expected duration [2].

The urban traffic flow forecasting models rely on historical and current flow data. The problem of traffic flow forecasting belongs to a standard time series prediction task and the purpose is to fetch the function which can relates future values of traffic flow to previous and current measurement of traffic flow [3]. A variety of forecasting techniques has been applied to forecast the urban traffic flow. In [4], DanechPajouh and Aron designed a layered statistical method with a mathematical clustering technique to group the traffic flow data and a separately tuned linear regression model for each cluster. The ARIMA model, initially developed by Kim et
al., is one of the most popular approaches in traffic flow forecasting [5-7]. However, the limitation of ARIMA models is that their natural tendency, concentrated on the mean values of the past series data, seems unable to capture the rapid varying process changes underlying of traffic flow. The artificial neural network (ANN) is widely applied in traffic flow forecasting. Yin et al. [8] developed a fuzzy-neural model (FNM) to predict traffic flow in an urban street network. The empirical results showed that the FNM model provides more accurate forecasting results than the BPNN model. These researches are committed to improve the performance of the algorithms. However, there are many factors which can affect the traffic flow, the traditional single model can hardly improve the prediction accuracy and no online correction mechanism was considered. This motivates the paper.

In this paper, the traffic flow forecasting model has 3 techniques: first, the input data of the model is divided into several categories according to FCM, and different categories have different model. Second, a training model based on a well-defined part-connected neural network (NN) was proposed and the cooperative quantum-particle swarm evolutionary algorithm (CQGAPSO) is used to train the
model. Last, the error between predicted value and real value is used to compensate the output of the model. These methods can improve the accuracy and generalization of the forecasting model can also overcome the model mismatch.

This paper is organized as follows. The forecasting methodology is introduced in Section 2. Cases of studying of urban traffic flow forecasting are given in Section 3. Conclusions are finally made in Section 4.

## 2. Forecasting Methodology

2.1. The Framework of the Proposed Method. According to the change rule of traffic flow time series, there is an essential linkage between the future and the previous flow [9]. Thus, the previous traffic flow value can be used to forecast the future flow. Set $f(t)$ as the traffic flow at time $t, f(t-1)$ as the value at time $t-1$. In this paper, $f(t), f(t-1), \ldots$, and $f(t-s)$ are the input values of the model at time $t$ and $f(t+p)$ is the predicted value at time $t+p$. The input values are denoted as $x(i)$ and the predicted value is denoted $y(i)$. The traffic flow forecasting model is made to build the relationship between $y(i)$ and $x(i)$. Therefore, once the relationship is obtained, the model can be used to predict the future traffic flow based on the real-time measured data in practice.

In the previous studies, the single prediction model mentioned above was adopted to forecast the urban traffic flow. However, it is not universally applicable for all the traffic scenarios. Since the urban traffic system is an unstable system, which exhibits significant variation in different periods, it is necessary to establish different prediction models to forecast the future traffic state accurately. According to the measured data from the float car, Guo et al. [10] analyzed the degree of traffic congestion on different days in a week. The results showed that the traffic congestion of Monday is more serious than the other days, especially in the morning peak hour, and the most serious traffic congestion of evening peak hour occurred in Friday. Moreover, the degree of traffic congestion during commuting time on the weekend is less than the degree on weekdays. It can be concluded that by observing the traffic flow data, the travel modes and travel demand are different on each day of a week, and the data characteristic of the same day for every week is similar. Therefore, in order to improve the accuracy of prediction for traffic flow or travel speed on the road, it is necessary to classify the traffic flow pattern and apply a suitable model to forecast each pattern. This classification would guarantee that each prediction model has a good performance in a particular period. As urban traffic flow system is a complicated process influenced by many factors, it is believed that using the multimodel method to predict the traffic flow is appropriate.

From the analysis made above, in this paper, for the sake of modeling, the historical traffic data should be divided into seven classes corresponding to each day of a week. Besides, considering the widely variation of traffic flow from morning to night, especially in the rush hour, using a single model to describe a complex nonlinear object usually results in low accuracy and poor generalization. So we use FCM to process the data and choose the reasonable clustering number by


Figure 1: The framework of the proposed method.


Figure 2: The architecture of FCM.
the experiments and use the approach based on multiple-input-single-output three-layer feedforward neural network with switches to model each cluster. Meanwhile, in order to overcome the model mismatch, the adaptive correction mechanism is added to our approach. The framework of the proposed method is illustrated in Figure 1.
2.2. Fuzzy c-Means Clustering. The model of forecasting traffic flow is a multiinput single-output system; the training sample set can be expressed as $D=\left\{f_{i}(t+p),\left[f_{i}(t) f_{i}(t-\right.\right.$ $\left.\left.1), \ldots, f_{i}(t-s)\right] \mid s=1,2 \ldots m, i=1,2, \ldots, n\right\}$. Here, $n$ is the sample number of training set, $m$ is the number of input variables; $\left[f_{i}(t) f_{i}(t-1), \ldots, f_{i}(t-s)\right]$ denotes the $i$ th input vector. Suppose $D$ is divided into $c$ clusters $\left\{D_{1}, D_{2}, \ldots D_{c}\right\}$; thus, $c$ submodels $\left\{M_{1}, M_{2}, \ldots M_{c}\right\}$ should be built for each $D_{i}$, and the result of the FCM can be expressed as membership matrix $U=\left[u_{i j}\right]_{i=1,2 \ldots c, j=1,2 \ldots n}=$ $\left\{U_{i} \mid i=1,2 \ldots c\right\}$. $u_{i j}$ denotes the degree of the element $x_{j}$ in training sample set belonging to the $i$ th cluster. The value of $u_{i j}$ is between 0 and 1 . The architecture of FCM method is shown in Figure 2 [11, 12].

Clustering number $c$ is a very important parameter. Here, we do experiments to choose the appropriate clustering number $c$. Let $c$ increase from 2 to a constant. Then, make models separately based on FCM and calculate the mean square error and the maximum error according to (1). Last, we can obtain the best clustering number $c$.

Consider

$$
\begin{align*}
& \text { MSE }=\left(\frac{1}{n_{1}} \sum_{i=1}^{n_{1}}\left(y_{i}-\widetilde{y_{i}}\right)^{2}\right)^{0.5},  \tag{1}\\
& \text { MAXE }=\max _{i=1}^{n_{1}}\left(\left|y_{i}-\widetilde{y_{i}}\right|\right) .
\end{align*}
$$

Table 1: The result of the FCM.

| Clustering number $c$ | Monday 1 |  | Wednesday 1 |  | Sunday 1 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | MSE | MAXE | MSE | MAXE | MSE | MAXE |
| 1 | 14.0515 | 57.5568 | 14.2659 | 61.2561 | 14.1235 | 57.3078 |
| 2 | 13.3489 | 51.3476 | 14.0024 | 53.1487 | 13.2149 | 51.0947 |
| 3 | 10.0456 | 43.1834 | 13.4820 | 48.4621 | 9.8952 | 42.1001 |
| 4 | 11.8439 | 44.9576 | 11.2106 | 45.2548 | 11.7541 | 44.8259 |
| 5 | 13.2563 | 47.5963 | 13.5279 | 49.3247 | 13.1589 | 46.2985 |
| Clustering number $c$ | Monday 2 |  | Wednesday 2 |  | Sunday 2 |  |
|  | MSE | MAXE | MSE | MAXE | MSE | MAXE |
| 1 | 15.1125 | 60.9547 | 14.3762 | 59.0143 | 14.4321 | 60.6465 |
| 2 | 14.5876 | 54.2154 | 14.1144 | 59.9821 | 14.2547 | 61.6435 |
| 3 | 11.5248 | 40.1257 | 13.5520 | 50.7234 | 13.6464 | 50.9542 |
| 4 | 12.5487 | 42.2037 | 11.3017 | 47.0984 | 12.3014 | 48.2549 |
| 5 | 13.6587 | 44.1023 | 13.6975 | 51.2459 | 13.4164 | 50.8216 |

2.3. The Forecasting Model Based on Neural Network with Switches. In the architecture of FCM method, each model needs a modeling tool. NN, SVM, and Kalman filtering are always used to forecast the traffic flow. Here, we adopt an advanced NN, the multiple-input-single-output three-layer feedforward neural network with switches was proposed and well defined in [13]. A multiple-input-single-output (MISO) three-layer feedforward neural work with switches is shown in Figure 3.

Various methods were proposed to train the NN with switches [13-15]. In those methods, the population was partitioned to parameters and structure population. The parameters population was composed of the weight of the links, while the structure population was composed of the link switches. This model could eliminate some ill effects of approximation ability caused by redundant structure of NN.
2.4. The Adaptive Correction Mechanism. The traffic flow is the measurable variable, and the real-time data is used to predict the future traffic flow [16]. For example, at current time $t$, we can obtain the real value $f(t)$ from the sensors and the predicted value $\tilde{f}(t)$ by forecasting the model. Here is an error $e=f(t)-\tilde{f}(t)$ because of the model mismatch. At time $k$, the model should forecast the traffic flow at time $k+p$; the error can be used to compensate the initial predicted value $f^{\prime}(t+p)$ according to (2). $h$ is the offline correction coefficient.

Consider

$$
\begin{equation*}
\tilde{f}(t+p)=f^{\prime}(t+p)+h \cdot e \tag{2}
\end{equation*}
$$

The training set $D$ can be used to fetch $h$; to fetch $h$ is to find the relationship between $f(t)-\tilde{f}(t)$ and $f(t+p)-\widetilde{f}(t+p)$, and here, $t=1,2 \ldots m, m$ is the sample number of training set. $h$ can be calculated by least square method (SLM).

When the model is forecasting the traffic flow online, the correction coefficient $h$ should be refreshed in real time. For example, at current time $t$, we can calculate $\Delta h$ using a small piece of historical data to obtain the relationship between $f(t-i-p)-\widetilde{f}(t-i-p)$ and $f(t-i)-\widetilde{f}(t-i)$. Here, $i$ is a small positive integer. The online correction coefficient $\Delta h$ can be obtained by SLM and (2) should be modified.

$\square$ Switches
Figure 3: The structure of three-layer feedforward NN with switches.

Consider

$$
\begin{equation*}
\tilde{f}(t+p)=f^{\prime}(k+p)+(h+\Delta h) \cdot e . \tag{3}
\end{equation*}
$$

## 3. Experimental Results

In order to explain the effectiveness of the proposed method, we choose the data gathered from Shanghai north-south highway including from August to October. The historical data on August and September is used to build the training set, while the data on October is used to build the testing set. There is a large difference of traffic flow every day in a week, thus we build different models for every day. Here, we use the first two Monday, Wednesday and Sunday on October to verify the proposed model.

Table 2: The result of adding the correction mechanism.

| $h+\Delta h$ | Monday 1 |  | Wednesday 1 |  | Sunday 1 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | MSE | MAXE | MSE | MAXE | MSE | MAXE |
| $h+\Delta h=0$ | 10.0456 | 43.1843 | 11.2106 | 45.2548 | 9.8952 | 42.1001 |
| $-0.1 \leq \Delta h \leq 0.1$ | 9.9758 | 42.1285 | 10.2654 | 44.1657 | 9.7561 | 40.6548 |
| $-0.2 \leq \Delta h \leq 0.2$ | 15.2648 | 59.2154 | 16.2299 | 61.2147 | 14.3215 | 57.6519 |
| $-0.3 \leq \Delta h \leq 0.3$ | 99.2154 | 70.2165 | 105.2647 | 85.2594 | 99.0147 | 69.2589 |
| $h+\Delta h$ | Monday 2 |  | Wednesday 2 |  | Sunday 2 |  |
|  | MSE | MAXE | MSE | MAXE | MSE | MAXE |
| $h+\Delta h=0$ | 11.5248 | 40.1257 | 11.3017 | 47.0984 | 12.3014 | 48.2549 |
| $-0.1 \leq \Delta h \leq 0.1$ | 10.4529 | 38.2489 | 10.3594 | 46.2813 | 11.2497 | 47.9523 |
| $-0.2 \leq \Delta h \leq 0.2$ | 16.5489 | 58.2146 | 17.2016 | 64.0525 | 16.2018 | 58.2687 |
| $-0.3 \leq \Delta h \leq 0.3$ | 103.4269 | 88.2159 | 106.2184 | 86.3468 | 101.4512 | 98.1264 |

Table 3: The comparison of 3 different models.

|  | Monday 1 |  | Wednesday 1 |  | Sunday 1 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | MSE | MAXE | MSE | MAXE | MSE | MAXE |
| Model (a) | 13.4956 | 46.7109 | 14.3495 | 48.3459 | 12.2304 | 46.0239 |
| Model (b) | 10.0456 | 43.1834 | 11.2106 | 45.2548 | 9.8952 | 42.1001 |
| Model (c) | 9.9758 | 42.1285 | 10.2654 | 44.1657 | 9.7561 | 40.6548 |
|  | Monday 2 |  | Wednesday 2 |  | Sunday 2 |  |
|  | MSE | MAXE | MSE | MAXE | MSE | MAXE |
| Model (a) | 15.2430 | 49.4545 | 14.4506 | 51.4539 | 16.1356 | 55.2341 |
| Model (b) | 11.5248 | 40.1257 | 11.3017 | 47.0984 | 12.3014 | 48.2549 |
| Model (c) | 10.4529 | 38.2489 | 10.3594 | 46.2813 | 11.2497 | 47.9523 |

The number of training sample is 2800 and the testing sample number is 650 . There is 2 minutes between each data. Based on the experience, we choose 3 as the dimension of input data. On request, we should predict the traffic flow after 10 minutes. Thus the width of the prediction $p$ is 5 . We totally do 3 experiments: (1) the traditional single model; (2) the multimodels based on FCM; (3) the multimodels based on FCM and adaptive correction mechanism.

First, all the data should be filtered before modeling and NN with switches is used as the modeling tool. Then we should determine the Clustering number $c$ by FCM, "CQGAPSO" algorithm is used to train the NN model and the parameter of "CQGAPSO" algorithm is given in [17]. The hidden nodes number is 6 . The training accuracy is $1 \times 10^{-5}$ and the iteration times of training the NN are 2000. The experiments are implemented for 50 times. Table 1 gives the result of FCM.

Form Table 1, we can find MSE and MAXE get better after an initial increase in growth of clustering number $c$. However, if $c$ continues to grow, MSE and MAXE will get worse. That is because with the increasing of the clustering number, the generalization ability of the model gets poorer. The best clustering number $c$ is at the turning point. Then the model should be added the adaptive correction mechanism. In order to obtain an appropriate correction coefficient, $h$ is a fixed number which is calculated offline while $\Delta h$ is a changed number which calculated online and we should limit the scope of $\Delta h$. Table 2 gives the result of adding the adaptive correction mechanism. From Table 2, we can find if
the adaptive correction mechanism parameter value is $-0.1 \leq$ $\Delta h \leq 0.1$, MSE and MAXE is the best. If the scope of $\Delta h$ is very wide, MSE and MAXE will get worse because the compensation value is too large. Table 3 gives the comparison of every approach. Model (a) is the traditional single model, model (b) is the model (a) with FCM, model (c) is the model (b) with the correction mechanism. We can find the reasonable clustering number $c$ and correction mechanism can improve the forecasting ability.

The Comparison of every approach is illustrated in Figure 3. Figure 4(a) is the traditional single model, Figure 4(b) is the model with FCM, Figure 4(c) is the model with FCM and the correction mechanism. In Figure 4(a), the predictive curve is smooth and cannot track exactly especially at the peak value because the approximation capability of the traditional single model is limited. In Figure 4(b), we can get some submodels by FCM and multimodel can improve the forecasting ability. Without the correction mechanism, the model error cannot be corrected in real time. In Figure 4(c), we use the correction mechanism and it compensates the initial forecasting value with the model error value. From Table 3 and Figure 4, we can find that the predictive accuracy is better than model (a) and (b).

## 4. Conclusions

Aiming at solving the problem of forecasting urban traffic flow, this paper proposes a forecasting model by the use of FCM and correction mechanism. The experimental results


Figure 4: Continued.


Figure 4: the result of forecasting the traffic flow.
indicate that the proposed method can perform better than other methods and show the application prospect.

## Acknowledgments

This paper is supported by the National Science Foundation of China (Grant nos. 71361130012, 61221003), the Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20120073110017), and the State Key Laboratory of Synthetical Automation for Process Industries.

## References

[1] M.-W. Li, W.-C. Hong, and H.-G. Kang, "Urban traffic flow forecasting using Gauss-SVR with cat mapping, cloud model and PSO hybrid algorithm," Neurocomputing, vol. 99, no. 1, pp. 230-240, 2013.
[2] T. Pohlmann and B. Friedrich, "A combined method to forecast and estimate traffic demand in urban networks," Transportation Research C, vol. 31, pp. 131-144, 2013.
[3] J. Abdi, B. Moshiri, B. Abdulhai, and A. K. Sedigh, "Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm," Engineering Applications of Artificial Intelligence, vol. 25, no. 5, pp. 1022-1042, 2012.
[4] M. Danech-Pajouh and M. Aron, "Athena: a method for shortterm inter-urban motorway traffic forecasting," Recherche Transports Sécurité, vol. 6, pp. 11-16, 1991.
[5] H.-W. Kim, J.-H. Lee, Y.-H. Choi, Y.-U. Chung, and H. Lee, "Dynamic bandwidth provisioning using ARIMA-based traffic
forecasting for Mobile WiMAX," Computer Communications, vol. 34, no. 1, pp. 99-106, 2011.
[6] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, Wiley, 1976.
[7] Y. Kamarianakis and P. Prastacos, "Space-time modeling of traffic flow," Computers and Geosciences, vol. 31, no. 2, pp. 119-133, 2005.
[8] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," Transportation Research Part C, vol. 10, no. 2, pp. 85-98, 2002.
[9] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," Expert Systems with Applications, vol. 36, no. 3, pp. 6164-6173, 2009.
[10] J. Guo, M. Liu, L. Yu, J. Guan, S. Guo, and X. Zhang, "Development and applications of macroscopic measurement of traffic congestion in Beijing," in Proceedings of the 7th China Annual Conference on Intelligent Transportation Systems, Beijing, China, 2007.
[11] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," Computers and Geosciences, vol. 10, no. 23, pp. 191-203, 1984.
[12] Z. Liu and Q. Liu, "Studying cost-sensitive learning for multiclass imbalance in Internet traffic classification," The Journal of China Universities of Posts and Telecommunications, vol. 19, no. 6, pp. 63-72, 2012.
[13] H. K. Lam, S. H. Ling, F. H. F. Leung, and P. K. S. Tam, "Tuning of the structure and parameters of neural network using an improved genetic algorithm," in Proceedings of the 27th Annual

Conference of the IEEE Industrial Electronics Society (IECON '2001), pp. 25-30, December 2001.
[14] J.-T. Tsai, J.-H. Chou, and T.-K. Liu, "Tuning the structure and parameters of a neural network by using hybrid Taguchi-genetic algorithm," IEEE Transactions on Neural Networks, vol. 17, no. 1, pp. 69-80, 2006.
[15] G. Yu, G. Li, Y. Bai, and X. Jin, "Tuning of the structure and parameters of wavelet neural network using improved chaotic PSO," in Proceedings of the 26th Chinese Control Conference (CCC '07), pp. 228-232, July 2007.
[16] Q. Tang, D. Li, Y. Xi, and D. Yin, "Soft-sensing design based on semiclosed-loop framework," Chinese Journal of Chemical Engineering, vol. 20, no. 6, pp. 1213-1218, 2012.
[17] Q. F. Tang, The cooperative quantum-particle swarm algorithm and its application in the energy utilization optimization of the steam network [M.E. thesis], East China University of Science\&Technology, China, 2011, (Chinese).

# Robust Fault Detection of Linear Uncertain Time-Delay Systems Using Unknown Input Observers 

Saeed Ahmadizadeh, ${ }^{1}$ Jafar Zarei, ${ }^{2}$ and Hamid Reza Karimi ${ }^{3}$<br>${ }^{1}$ Electrical Engineering Department, Iran University of Science and Technology, Narmak, Tehran 16846, Iran<br>${ }^{2}$ Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz 71557-13876, Iran<br>${ }^{3}$ Department of Engineering, Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway<br>Correspondence should be addressed to Hamid Reza Karimi; hamid.r.karimi@uia.no

Received 18 July 2013; Accepted 13 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 Saeed Ahmadizadeh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper deals with the problem of fault detection for linear uncertain time-delay systems. The proposed method for Luenberger observers is developed for unknown input observers (UIOs), and a novel procedure for the design of residual based on UIOs is presented. The design procedure is carried out based on the $H_{\infty}$ model matching approach which minimizes the difference between generated residuals by the optimal observer and those by the designed observer in the presence of uncertainties. The optimal observer is designed for the ideal system and works so that the fault effect is maximized while the exogenous disturbances and noise effects are minimized. This observer can give disturbance decoupling in the presence of noise and uncertainties for linear uncertain time-delay systems. The developed method is applied to a numerical example, and the simulation results show that the proposed approach is able to detect faults reliably in the presence of modeling errors, disturbances, and noise.


## 1. Introduction

Fault detection and isolation (FDI) is an essential and challenging problem in many industrial applications. Among the various reported methods, much attention has been paid to model based approaches in the field of control engineering in recent years. For example, fault detection problem for discrete-time Markov jump systems and switched systems is investigated in [1, 2], respectively. The problem of fault reconstruction for a class of descriptor linear systems using sliding mode observers is presented in [3]. The sliding mode observers have been designed such that the actuator fault can be reconstructed using output measurements. The datadriven scheme for FDI is presented in [4] which exploits an adaptive residual generator and a bank of isolation observers. The designed scheme obtains observer parameters without identification of complete process model.

However, model based approaches are based on some idealized assumptions, one of which is that the mathematical model of the plant is a faithful replica of the plant dynamics [5]. As the mathematical model of a plant hardly represents its complete behavior, due to the existence of model uncertainty, noise, and unknown disturbances, it is essential to design a fault diagnosis system to take these effects into consideration. Motivated by the abovementioned issues, a robust fault detection scheme is exploited to design fault detection systems so that high sensitivity to faults as well as low sensitivity to uncertainties and perturbation can be obtained. Optimization techniques are widely used to solve this problem. One of the commonly used approaches to design such FDI scheme is representing the design procedure by $H_{\infty}$ and $H_{-}$ performance indexes. The main advantage of this approach is that it can be solved by linear matrix inequality (LMI) [6]. In $[7,8]$, a two-step FDI design methodology is presented.

In this methodology, the optimal fault detection filter (FDF) has firstly been designed, neglecting the existence of model uncertainty. Next, the FDF, which is used as residual generator, has been obtained via $H_{\infty}$ model matching technique. The same approach is used in [9]; however, different dynamics is considered for the FDF.

Time delay is an inherent characteristic of many industrial systems; therefore, robust FDI problem for LTI systems with time delay received great attention over recent decades, and numerous articles have been presented. One approach is to solve the formulated design procedure using the eigenstructure assignment approach in which the residual signal is thoroughly decoupled from delay-free unknown input. Then the effect of the unknown input is minimized using $H_{\infty}$ norm [10]. In the presence of uncertainty, the same approach as delay-free case can be employed to obtain a robust FDI system. Indeed, solving the $H_{\infty}$ model matching problem results in achieving an FDF which acts the same as the optimal one [11-13]. Although the same approach is considered in these references, solving procedures are completely different owing to the difference between the dynamic of the filter and system. In [14], the problem of robust FDF design for the class of linear systems has been investigated. The system is subjected to mixed neutral and discrete time-varying delays and some nonlinear perturbations. The Luenberger type observer has been utilized to design FDF such that the residual signals effectively show fault occurrence.

Another approach commonly used to robust FDI scheme is to employ the unknown input observers (UIO), in which the residual is designed to be sensitive to faults but insensitive to unknown disturbances. Although the UIO has been widely used in estimation problems in both time delay and delayfree systems [15-18], there are few references that handle the problem of designing robust FDI [19, 20]. In [20], a design procedure has been proposed for delay-free systems so that perturbations and exogenous signals have less effect on the residual signal and the fault has a detectable effect on the residual; however, the problem has not been presented in $H_{\infty}$ model matching technique. Motivated by this consideration, a robust FDF design using UIO for uncertain systems with time delay is presented and solved using $H_{\infty}$ model matching approach. In contrast to our previous work [18], we are concerned to design a robust FDF for the case in which the dynamic characteristic of fault signal is known. For this purpose, a two-step design procedure is developed. In the first step, the optimal fault detection based on UIO is designed for the system without uncertainty. Next, the UIO-based fault detection filter is approached to optimal one in the sense of $H_{\infty}$ norm. It is demonstrated through simulation that the presented fault detection observer is robust against uncertainty and sensitive enough to the faults.

Notation. The notations used throughout the paper are fairly standard. I and $\mathbf{0}$ represent identity matrix and zero matrix; the superscript " $T$ " stands for matrix transposition. $\|\cdot\|$ refers to the Euclidean vector norm or the induced matrix 2-norm. $\operatorname{diag}\{\cdot\}$ represents a block diagonal matrix.

The notation $P>0$ means that $P$ is real symmetric and positive definite; the symbol $*$ denotes the elements below the main diagonal of a symmetric block matrix.

## 2. Problem Statement

Many different industrial systems such as mechanical, electrical, meteorological, chemical, economic, and biological systems include time delay. In many studies linearized model of these systems around point of operation is considered. However, there are always some discrepancies between the real dynamics of the system and linearized model. These differences arise from systems uncertainty, as a consequence of neglecting dynamics, and changes in system parameters. Therefore, the following linear uncertain system with additive disturbances and time delay is considered to represent the described model:

$$
\begin{gather*}
\dot{x}(t)=(A+\Delta A(t)) x(t)+\left(A_{d}+\Delta A_{d}(t)\right) x(t-\tau) \\
+(B+\Delta B(t)) u(t)+E d(t)+F_{x} f(t)+\operatorname{Rn}(t),  \tag{1}\\
y(t)=C x(t)+F_{y} f(t)+D n(t),
\end{gather*}
$$

where $x(t) \in \Re^{n}$ is the state vector, $y(t) \in \Re^{p}$ is the output vector, $u(t) \in \mathfrak{R}^{q}$ is the input vector, $d(t) \in \mathfrak{R}^{m}$ is an unknown scalar function representing the disturbance that belongs to $L_{2}^{m}(0, \infty), f(t) \in \mathfrak{R}^{f}$ denotes the faults, and $n(t) \in \Re^{r}$ represents the noise. Note that $\Delta A(t), \Delta B(t)$, and $\Delta A_{d}(t)$ are the norm bounded time-varying uncertainties of the matrices $A, B$, and $A_{d}$, respectively, and $\tau \geq 0$ is a constant delay. It is assumed that the characteristics of uncertainty matrices belong to

$$
\begin{gather*}
\Omega_{1}=\left\{\Delta A(t) \mid \Delta A(t)=M_{1} \Sigma_{1}(t) N_{1}, \Sigma_{1}^{T}(t) \Sigma_{1}(t) \leq I\right\} \\
\Omega_{2}=\left\{\Delta B(t) \mid \Delta B(t)=M_{2} \Sigma_{2}(t) N_{2}, \Sigma_{2}^{T}(t) \Sigma_{2}(t) \leq I\right\} \\
\Omega_{3}=\left\{\Delta A_{d}(t) \mid \Delta A_{d}(t)=M_{3} \Sigma_{3}(t) N_{3}, \Sigma_{3}^{T}(t) \Sigma_{3}(t) \leq I\right\} \tag{2}
\end{gather*}
$$

where $M_{i}$ and $N_{i}$ are predefined matrices. It is supposed that all over the paper the dimensions of matrices are compatible if they are not explicitly mentioned.

The dynamic characteristic of fault signal can be described by [21]

$$
\begin{gather*}
\dot{\theta}(t)=A_{\theta} \theta(t), \quad t \geq t_{f}, \\
\theta(t)=0, \quad t \in\left[0, t_{f}\right]  \tag{3}\\
\theta\left(t_{f}\right)=\theta_{0}, \\
f(t)=F_{\theta} \theta(t),
\end{gather*}
$$

where $t_{f}$ is the time when a fault occurs and $A_{\theta}$ and $F_{\theta}$ are known matrices with appropriate dimensions. The initial time, $t_{f}$, and initial state, $\theta_{0}$, are supposed to be unknown. The dynamic equation (3) represents any fault with known
dynamic characteristic and unknown amplitude and phase [21].

Unknown input observer for the class of time-delay system (1) is considered as [22]

$$
\begin{gather*}
\dot{z}(t)=F z+G z(t-\tau)+H u(t)+K_{1 z} y(t)+K_{2 z} y(t-\tau), \\
\hat{x}(t)=z(t)+L_{1} y(t), \tag{4}
\end{gather*}
$$

where $\widehat{x}(t)$ is the estimated state vector. The dynamic of $\widehat{x}(t)$ is governed by

$$
\begin{align*}
\dot{\hat{x}}(t)= & F \widehat{x}+G \widehat{x}(t-\tau)+H u(t)+L_{1} \dot{y}(t)  \tag{5}\\
& +L_{2} y(t)+L_{3} y(t-\tau),
\end{align*}
$$

where $F, G, H$, and $L_{1}$ are the observer matrices and $L_{2}=$ $K_{1 z}-F L_{1}, L_{3}=K_{2 z}-G L_{1}$.

The observer matrices will be designed such that the disturbance and input are decoupled from the estimation error defined by $e(t)=x(t)-\widehat{x}(t)$. When UIO-based filter defined by (4) is applied to the system described in (1), the state estimation error will be

$$
\begin{align*}
\dot{e}(t)= & F e(t)+G e(t-\tau)+\left(\left(I-L_{1} C\right) A-L_{2} C-F\right) x(t) \\
& +\left(\left(I-L_{1} C\right) A_{d}-L_{3} C-G\right) x(t-\tau) \\
& +\left(\left(I-L_{1} C\right) B-H\right) u(t) \\
& +\left(I-L_{1} C\right) E d(t) \\
& +\left(\left(I-L_{1} C\right) F_{x} F_{\theta}-L_{2} F_{y} F_{\theta}-L_{1} F_{y} F_{\theta} A_{\theta}\right) \theta(t) \\
& -L_{3} F_{y} F_{\theta} \theta(t-\tau)+\left(\left(R-L_{1} C R\right)-L_{2} D\right) n(t) \\
& -L_{3} D n(t-\tau)-L_{1} D \dot{n}(t) \\
& +\left(I-L_{1} C\right) \Delta A(t) x(t)+\left(I-L_{1} C\right) \Delta A_{d}(t) x(t-\tau) \\
& +\left(I-L_{1} C\right) \Delta B(t) u(t) . \tag{6}
\end{align*}
$$

In the absence of uncertainties and faults, it is shown that the observer, defined by (4), is UIO for the predefined system by (1) if the following conditions are satisfied [22].

## Condition 1:

$$
\begin{equation*}
\dot{e}(t)=F e(t)+G e(t-\tau) \text { is asymptotically stable. } \tag{7}
\end{equation*}
$$

Condition 2:

$$
\begin{equation*}
F=\left(I-L_{1} C\right) A-L_{2} C . \tag{8}
\end{equation*}
$$

Condition 3:

$$
\begin{equation*}
G=\left(I-L_{1} C\right) A_{d}-L_{3} C . \tag{9}
\end{equation*}
$$

Condition 4:

$$
\begin{equation*}
H=\left(I-L_{1} C\right) B . \tag{10}
\end{equation*}
$$

Condition 5:

$$
\begin{equation*}
\left(I-L_{1} C\right) E=\mathbf{0} \tag{11}
\end{equation*}
$$

where $\mathbf{0}$ denotes a null matrix with compatible dimension. Using these relationships, and considering definitions in (12), the state estimation error dynamic (6) is transformed to (13):

$$
T=\left(I-L_{1} C\right)
$$

$$
\begin{gather*}
\bar{F}=\left[\left(I-L_{1} C\right) F_{x} F_{\theta}-L_{2} F_{y} F_{\theta}-L_{1} F_{y} F_{\theta} A_{\theta}-L_{3} F_{y} F_{\theta}\right] \\
\bar{R}=\left[-L_{1} D \quad\left(I-L_{1} C\right) R-L_{2} D-L_{3} D\right] \\
\bar{\theta}=\left[\theta^{T}(t) \theta^{T}(t-\tau)\right]^{T} \\
\bar{n}=\left[\dot{n}^{T}(t) n^{T}(t) n^{T}(t-\tau)\right]^{T}  \tag{12}\\
\dot{e}(t)= \\
 \tag{13}\\
+ \\
+T \Delta A(t)+G e(t-\tau)+\bar{F} \bar{\theta}+\bar{R} \bar{n} \\
\\
+T \Delta B(t) u(t) .
\end{gather*}
$$

In order to use the UIO for fault detection purposes, a residual signal should be defined. Difference between measured output and estimated output is usually considered as a residual signal. In current work, a more general form for residual reference signal is considered as follows:

$$
\begin{gather*}
\hat{y}(t)=C \widehat{x}(t), \\
r(t)=V_{1}(y(t)-\widehat{y}(t))+V_{2}(y(t-\tau)-\widehat{y}(t-\tau)) \\
=V_{1} C e(t)+V_{2} C e(t-\tau)+K_{1} \bar{\theta}(t)+K_{2} \bar{n}(t),  \tag{14}\\
K_{1}=\left[\begin{array}{lll}
V_{1} F_{y} F_{\theta} & V_{2} F_{y} F_{\theta}
\end{array}\right] \\
K_{2}=\left[\begin{array}{lll}
\mathbf{0} & V_{1} D & V_{2} D
\end{array}\right] .
\end{gather*}
$$

The goal of robust fault detection problem is to minimize the performance index defined in (15) for all classes of model uncertainties belonging to $\Omega_{i}$. In general, this performance index is minimized using $H_{\infty}$ model matching approach which minimizes the difference between the residual signal $(r(t))$ and reference residual signal $\left(r_{f}(t)\right)$
in the presence of the worst case disturbance signals. This performance index has been minimized by the following steps.

Step 1. The ideal residual signal generator system has been designed for the system without uncertainty defined in (16). The residual signal shows the maximum sensitivity to the fault signal while it has the minimum sensitivity to disturbance, noise, and unknown inputs.

Step 2. The residual signal generator system has been designed such that the performance index (17) is minimized and the overall system (19) is asymptotically stable:

$$
\begin{gather*}
J_{r}=\min _{\left(\Delta A, \Delta B, \Delta A_{d}\right) \in \Omega_{i}} \frac{\left\|G_{r[d \bar{n}}\right\|_{\infty}}{\left\|G_{r \bar{f}}\right\|_{\infty}},  \tag{15}\\
\dot{e}_{f}(t)=F^{*} e_{f}(t)+G^{*} e_{f}(t-\tau)+\bar{F}^{*} \bar{\theta}(t)+\bar{R}^{*} \bar{n}(t), \\
r_{f}(t)=V_{1}^{*} C e_{f}(t)+V_{2}^{*} C e_{f}(t-\tau)+K_{1}^{*} \bar{\theta}(t)+K_{2}^{*} \bar{n}(t),  \tag{16}\\
J_{r_{e}}=\sup _{\left(\Delta A, \Delta B, \Delta A_{d}\right) \in \Omega_{i}} \frac{\left\|r_{e}\right\|_{2}}{\|\omega\|_{2}}<\gamma, \tag{17}
\end{gather*}
$$

where

$$
\begin{gather*}
r_{e}(t)=r(t)-r_{f}(t) \\
\omega=\left[\begin{array}{lll}
u^{T} & \bar{\theta}^{T} & d^{T} \\
\bar{n}^{T}
\end{array}\right]^{T}  \tag{18}\\
\dot{\zeta}(t)=(\widetilde{A}+\Delta \widetilde{A}) \zeta(t)+\left(\widetilde{A}_{d}+\Delta \widetilde{A}_{d}\right) \zeta(t-\tau) \\
+\left(\widetilde{B}_{\omega_{1}}+\Delta \widetilde{B}_{\omega_{1}}\right) \omega(t)  \tag{19}\\
r_{e}(t)=\widetilde{C}_{1} \zeta(t)+\widetilde{C}_{2} \zeta(t-\tau)+\widetilde{D} \omega(t),
\end{gather*}
$$

where

$$
\begin{aligned}
& \zeta(t)=\left[\begin{array}{lll}
e^{T}(t) & e_{f}^{T}(t) & x^{T}(t)
\end{array}\right]^{T}, \\
& \widetilde{A}=\left[\begin{array}{ccc}
F & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & F^{*} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & A
\end{array}\right], \quad \widetilde{A}_{d}=\left[\begin{array}{ccc}
G & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & G^{*} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & A_{d}
\end{array}\right], \\
& \widetilde{B}_{\omega_{1}}=\left[\begin{array}{cccc}
\mathbf{0} & \bar{F} & \mathbf{0} & \bar{R} \\
\mathbf{0} & \bar{F}^{*} & \mathbf{0} & \bar{R}^{*} \\
B & K_{3} & E & K_{4}
\end{array}\right], \quad \Delta \widetilde{A}=\left[\begin{array}{ccc}
\mathbf{0} & \mathbf{0} & T \Delta A \\
\mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \Delta A
\end{array}\right],
\end{aligned}
$$

$$
\begin{gather*}
\Delta \widetilde{A}_{d}=\left[\begin{array}{ccc}
\mathbf{0} & \mathbf{0} & T \Delta A_{d} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \Delta A_{d}
\end{array}\right], \quad \Delta \widetilde{B}_{\omega_{1}}=\left[\begin{array}{cccc}
T \Delta B & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\Delta B & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right], \\
\widetilde{C}_{1}=\left[\begin{array}{lll}
V_{1} C & -V_{1}^{*} C & \mathbf{0}
\end{array}\right], \quad \widetilde{C}_{2}=\left[\begin{array}{lll}
V_{2} C & -V_{2}^{*} C & \mathbf{0}
\end{array}\right], \\
\widetilde{D}=\left[\begin{array}{ll}
\mathbf{0} & K_{1}-K_{1}^{*} \\
\mathbf{0} & K_{2}-K_{2}^{*}
\end{array}\right], \\
K_{3}=\left[\begin{array}{lll}
F_{x} F_{\theta} & \mathbf{0}
\end{array}\right], \quad K_{4}=\left[\begin{array}{lll}
\mathbf{0} & R & \mathbf{0}
\end{array}\right] . \tag{20}
\end{gather*}
$$

Furthermore, using (2) it is easy to see that $\Delta \widetilde{A}, \Delta \widetilde{A}_{d}, \Delta \widetilde{B}_{\omega_{1}}$ can be expressed by

$$
\begin{gather*}
\Delta \widetilde{A}=\widetilde{M}_{1} \Sigma_{1}(t) \widetilde{N}_{1}=\left[\begin{array}{c}
T M_{1} \\
\mathbf{0} \\
M_{1}
\end{array}\right] \Sigma_{1}(t)\left[\begin{array}{lll}
\mathbf{0} & \mathbf{0} & N_{1}
\end{array}\right], \\
\Delta \widetilde{A}_{d}=\widetilde{M}_{2} \Sigma_{2}(t) \widetilde{N}_{2}=\left[\begin{array}{c}
T M_{2} \\
\mathbf{0} \\
M_{2}
\end{array}\right] \Sigma_{2}(t)\left[\begin{array}{lll}
\mathbf{0} & \mathbf{0} & N_{2}
\end{array}\right],  \tag{21}\\
\Delta \widetilde{B}_{\omega_{1}}=\widetilde{M}_{3} \Sigma_{3}(t) \widetilde{N}_{3}=\left[\begin{array}{c}
T M_{3} \\
\mathbf{0} \\
M_{3}
\end{array}\right] \Sigma_{3}(t)\left[\begin{array}{llll}
N_{3} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right] .
\end{gather*}
$$

Before developing theorems that are utilized in designed procedure, the following lemmas, which are useful to prove the theorems, are introduced.

Lemma 1 (see [22]). Condition 5 is solvable if and only if the following relation holds:

$$
\begin{equation*}
\operatorname{rank}(C E)=m, \quad m \leq p \tag{22}
\end{equation*}
$$

The general solution of condition 5 can be calculated by

$$
\begin{align*}
L_{1} & =E(C E)^{+}+Y\left[I-C E(C E)^{+}\right]  \tag{23}\\
& =\Theta_{1}+Y \Theta_{2}
\end{align*}
$$

where $Y$ is an arbitrary matrix with an appropriate dimension.
Lemma 2. Suppose that $M, N$, and $\Sigma(t)$ are compatible and $\Sigma^{T}(t) \Sigma(t) \leq I$; then there exists a scalar $\varepsilon>0$ such that the following equation holds:

$$
\begin{equation*}
M \Sigma N+(M \Sigma N)^{T} \leq \varepsilon M M^{T}+\varepsilon^{-1} N^{T} N \tag{24}
\end{equation*}
$$

(a) Reference Model Selection (Step 1). Reference model selection is an important key to the design of robust fault detection filter for linear uncertain time-delay systems. To this end, the analogues procedure as that for fault detection in [6] is extended for delay systems considering the UIO as the
fault detection filter. According to (16), the reference residual signal can be written as sum of two signals, $r_{f \bar{n}}(t)$ and $r_{f \bar{\theta}}(t)$. The reference model should be chosen such that the effect of exogenous signals on the reference residual signal is minimized while the effect of fault signal is maximized. These two tasks are described mathematically by

$$
\begin{align*}
& \left\|T\left(r_{f \bar{n}}, \bar{n}\right)\right\|_{\infty} \leq \alpha \\
& \left\|T\left(r_{f \bar{f}}, \bar{f}\right)\right\|_{-} \geq \beta \tag{25}
\end{align*}
$$

where $T(\cdot, \cdot)$ is the transfer function between two signals. The following two theorems provide conditions which ensure the
asymptotic stability of (16). They also provide the conditions that increase the sensitivity of the residual signal from faults and decrease the sensitivity of residual signal from noise.

Theorem 3. For given $\alpha>0$, if there exist symmetric positive matrices $P, Q, V_{1}^{*}, V_{2}^{*}, \Phi_{1}^{*}, \Phi_{2}^{*}$, and $\Phi_{3}^{*}$ such that the following LMI holds:

$$
\left[\begin{array}{cccc}
P F^{*}+F^{* T} P+Q & P G^{*} & P \bar{R}^{*} & C^{T} V_{1}^{* T}  \tag{26}\\
* & -Q & \mathbf{0} & C^{T} V_{2}^{* T} \\
* & * & -\alpha^{2} I & K_{2}^{* T} \\
* & * & * & -I
\end{array}\right]<0
$$

where

$$
\begin{gather*}
P F^{*}=P\left(A-\Theta_{1} C A\right)-\Phi_{1}^{*}\left(\Theta_{2} C A\right)-\Phi_{2}^{*} C, \\
P G^{*}=P\left(A_{d}-\Theta_{1} C A_{d}\right)-\Phi_{1}^{*}\left(\Theta_{2} C A_{d}\right)-\Phi_{3}^{*} C,  \tag{27}\\
P \bar{R}^{*}=\left[-P\left(\Theta_{1} D\right)-\Phi_{1}^{*}\left(\Theta_{2} D\right) P\left(R-\Theta_{1} C R\right)-\Phi_{1}^{*}\left(\Psi_{2} C R\right)-\Phi_{2}^{*} D-\Phi_{3}^{*} D\right], \\
K_{2}^{*}=\left[\begin{array}{lll}
\mathbf{0} & \left.V_{1}^{*} D V_{2}^{*} D\right],
\end{array}\right.
\end{gather*}
$$

then the system (28) is asymptotically stable and $\left\|T\left(r_{f \bar{n}}, \bar{n}\right)\right\|_{\infty} \leq \alpha$. Furthermore, the UIO matrices are obtained from conditions 2 to 5 , and $Y^{*}=P^{-1} \Phi_{1}, L_{2}=P^{-1} \Phi_{2}, L_{3}=$ $P^{-1} \Phi_{3}$,

$$
\begin{gather*}
\dot{e}_{f \bar{n}}(t)=F^{*} e_{f \bar{n}}(t)+G^{*} e_{f \bar{n}}(t-\tau)+\bar{R}^{*} \bar{n}(t),  \tag{28}\\
r_{f \bar{n}}(t)=V_{1}^{*} C e_{f \bar{n}}(t)+V_{2}^{*} C e_{f \bar{n}}(t-\tau)+K_{2}^{*} \bar{n}(t) .
\end{gather*}
$$

Proof. Condition $\left\|T\left(r_{f \bar{n}}, \bar{n}\right)\right\|_{\infty} \leq \alpha$ is equivalent to $J_{r_{f \bar{n}}}$ : $\int_{0}^{\infty}\left(r_{f \bar{n}}^{T}(t) r_{f \bar{n}}(t)-\alpha^{2} \bar{n}^{T}(t) \bar{n}(t)\right) d t \geq 0$. Now Consider the

Lyapunov-Krasovskii function which is defined as $V(t)=$ $e_{f \bar{n}}^{T}(t) P e_{f \bar{n}}(t)+\int_{t-\tau}^{t} e_{f \bar{n}}^{T}(s) Q e_{f \bar{n}}(s) d s$. Then we have

$$
\begin{align*}
J_{r_{f \bar{n}}}= & \int_{0}^{\infty}\left(r_{f \bar{n}}^{T}(t) r_{f \bar{n}}(t)-\alpha^{2} \bar{n}^{T}(t) \bar{n}(t)+\dot{V}(t)\right) d t  \tag{29}\\
& +V(0)-V(\infty)
\end{align*}
$$

Assume $r_{f \bar{n}}(t)=\mathbf{0}$ for $t \in[-\tau, 0]$. Since $V(\infty)>0$, it can be concluded that

$$
\begin{equation*}
J_{r_{f \bar{n}}} \leq \int_{0}^{\infty}\left(r_{f \bar{n}}^{T}(t) r_{f \bar{n}}(t)-\alpha^{2} \bar{n}^{T}(t) \bar{n}(t)+\dot{V}(t)\right) d t \tag{30}
\end{equation*}
$$

Taking derivative from $V(t)$ and considering (28) yield

$$
J_{r_{f \bar{n}}} \leq \int_{t-\tau}^{t}\left[\begin{array}{c}
e_{f \bar{n}}(t)  \tag{31}\\
e_{f \bar{n}}(t-\tau) \\
\bar{n}(t)
\end{array}\right]^{\left[\begin{array}{ccc}
P F^{*}+F^{* T} P+Q+C^{T} V_{1}^{* T} V_{1}^{*} C & P G^{*}+C^{T} V_{1}^{* T} V_{2}^{*} C & P \bar{R}^{*}+C^{T} V_{1}^{* T} K_{2}^{*} \\
* & C^{T} V_{2}^{* T} V_{2}^{*} C-Q & C^{T} V_{2}^{* T} K_{2}^{*} \\
* & * & -\alpha^{2} I+K_{2}^{* T} K_{2}^{*}
\end{array}\right]}\left[\begin{array}{c}
e_{f \bar{n}}(t) \\
e_{f \bar{n}}(t-\tau) \\
\bar{n}(t)
\end{array}\right] .
$$

Hence, $\Xi<0$ implies $J_{r_{f \bar{n}}}<0$. Using the Schur complement theorem (26) is concluded from (31). Indeed, the inequality (26), without considering (27), includes nonlinear terms of $P Y^{*}, P L_{2}^{*}$, and $P L_{3}^{*}$ which lead the LMI to be infeasible. To overcome this problem, define $\Phi_{1}^{*}=P Y^{*}$,
$\Phi_{2}^{*}=P L_{2}^{*}$, and $\Phi_{3}^{*}=P L_{3}^{*}$. Using conditions 2, 3 and (23) it can be shown that (27) makes the LMI feasible. It completes the proof.

Theorem 4. For given $\beta>0$, if there exists symmetric positive matrices $P, Q, V_{1}^{*}, V_{2}^{*}, \Phi_{1}^{*}, \Phi_{2}^{*}$, and $\Phi_{3}^{*}$ such that the following LMI holds:

$$
\left[\begin{array}{cccc}
P F^{*}+F^{* T} P-Q+2 \varphi_{1}\left(V_{1}^{*}, V_{1 c}^{n}\right) & -P G^{*} & -P \bar{F}^{*} & C^{T} V_{1}^{* T}  \tag{32}\\
* & -Q+2 \varphi_{2}\left(V_{2}^{*}, V_{2 c}^{n}\right) & \mathbf{0} & C^{T} V_{2}^{* T} \\
* & * & \beta^{2} F_{\theta}^{T} F_{\theta}+2 \varphi_{3}\left(K_{1}^{*}, K_{1}^{n}\right) & K_{1}^{* T} \\
* & * & * & -I
\end{array}\right]<0,
$$

where

$$
\begin{gather*}
P F^{*}=P\left(A-\Theta_{1} C A\right)-\Phi_{1}^{*}\left(\Theta_{2} C A\right)-\Phi_{2}^{*} C, \\
P G^{*}=P\left(A_{d}-\Theta_{1} C A_{d}\right)-\Phi_{1}^{*}\left(\Theta_{2} C A_{d}\right)-\Phi_{3}^{*} C, \\
P \bar{F}^{*}=\left[P\left(F_{x} F_{\theta}-\Theta_{1} C F_{x} F_{\theta}\right)-\Phi_{1}^{*}\left(\Theta_{2} C F_{x} F_{\theta}\right)-\Phi_{2}^{*} F_{y} F_{\theta}-P\left(\Psi_{1} F_{y} F_{\theta} A_{\theta}\right)-\Phi_{1}^{*}\left(\Psi_{2} F_{y} F_{\theta} A_{\theta}\right)-\Phi_{3}^{*} F_{y} F_{\theta}\right],  \tag{33}\\
K_{1}^{*}=\left[V_{1}^{*} F_{y} F_{\theta} V_{2}^{*} F_{y} F_{\theta}\right], \\
V_{1 c}^{n}=V_{1}^{* n-1} C, \quad V_{2 c}^{n}=V_{2}^{* n-1} C, \quad K_{11}^{n}=V_{1}^{* n-1} F_{y} F_{\theta}, K_{12}^{n}=V_{2}^{* n-1} F_{y} F_{\theta}, \quad \text { for } n=1,2, \ldots, \\
K_{1}^{n}=\left[\begin{array}{ll}
K_{11}^{n} & \left.K_{12}^{n}\right]
\end{array}\right. \\
\varphi_{1}\left(V_{1}^{*}, V_{1 c}^{n}\right)=\left(V_{1 c}^{n}\right)^{T} V_{1 c}^{n}-\left(V_{1 c}^{n}\right)^{T} V_{1}^{*} C-C^{T} V_{1}^{* T} V_{1 c}^{n},  \tag{34}\\
\varphi_{2}\left(V_{2}^{*}, V_{2 c}^{n}\right)=\left(V_{2 c}^{n}\right)^{T} V_{2 c}^{n}-\left(V_{2 c}^{n}\right)^{T} V_{2}^{*} C-C^{T} V_{2}^{* T} V_{2 c}^{n}, \\
\varphi_{3}\left(K_{1}^{*}, K_{1}^{n}\right)=\left(K_{1}^{n}\right)^{T} K_{1}^{n}-\left(K_{1}^{n}\right)^{T} V_{1}^{*} F_{y} F_{\theta}-F_{\theta}^{T} F_{y}^{T} V_{1}^{* T} K_{1}^{n},
\end{gather*}
$$

then the system (35) is asymptotically stable, and $\left\|T\left(r_{f \bar{\theta}}, \bar{f}\right)\right\|_{-} \geq \beta$. Moreover, the UIO matrices are obtained from conditions 2 to 5, and $Y^{*}=P^{-1} \Phi_{1}, L_{2}=P^{-1} \Phi_{2}, L_{3}=$ $P^{-1} \Phi_{3}$,

$$
\begin{gather*}
\dot{e}_{f \bar{\theta}}(t)=F^{*} e_{f \bar{\theta}}(t)+G^{*} e_{f \bar{\theta}}(t-\tau)+\bar{F}^{*} \bar{\theta}(t), \\
r_{f \bar{\theta}}(t)=V_{1}^{*} C e_{f \bar{\theta}}(t)+V_{2}^{*} C e_{f \bar{\theta}}(t-\tau)+K_{2}^{*} \bar{\theta}(t) . \tag{35}
\end{gather*}
$$

Proof. Condition $\left\|T\left(r_{f \bar{f}}, \bar{f}\right)\right\|_{-} \geq \beta$ is equivalent to $J_{r_{f \bar{f}}}$ : $\int_{0}^{\infty}\left(r_{f \bar{\theta}}^{T}(t) r_{f \bar{\theta}}(t)-\beta^{2} \bar{f}^{T}(t) \bar{f}(t)\right) d t \geq 0$. Now consider the Lyapunov-Krasovskii function which is defined as $V(t)=$

$$
\begin{align*}
& e_{f \bar{\theta}}^{T}(t) P e_{f \bar{\theta}}(t)+\int_{t-\tau}^{t} e_{f \bar{\theta}}^{T}(s) Q e_{f \bar{\theta}}(s) d s \text {. Then we have } \\
& J_{r_{f \bar{\theta}}}=\int_{0}^{\infty}\left(r_{f \bar{\theta}}^{T}(t) r_{f \bar{\theta}}(t)-\beta^{2} \bar{f}^{T}(t)\right.  \tag{36}\\
& \quad \times \bar{f}(t)-\dot{V}(t)) d t-V(0)+V(\infty)
\end{align*}
$$

Assume $r_{f \bar{\theta}}(t)=\mathbf{0}$ for $t \in[-\tau, 0]$. Since $V(\infty)>0$, we have

$$
\begin{equation*}
J_{r_{f \bar{\theta}}} \geq \int_{0}^{\infty}\left(r_{f \bar{\theta}}^{T}(t) r_{f \bar{\theta}}(t)-\beta^{2} \bar{f}^{T}(t) \bar{f}(t)-\dot{V}(t)\right) d t \tag{37}
\end{equation*}
$$

Taking derivative from $V(t)$ and considering (35) yield

$$
J_{r_{f \bar{\theta}}}>\int_{t-\tau}^{t}\left[\begin{array}{c}
e_{f \bar{\theta}}(t)  \tag{38}\\
e_{f \overline{\bar{\theta}}}(t-\tau) \\
\bar{\theta}(t)
\end{array}\right]^{\left[\begin{array}{ccc}
-P F^{*}-F^{*^{T}} P-Q+C^{T} V_{1}^{* T} V_{1}^{*} C & -P G^{*}+C^{T} V_{1}^{* T} V_{2}^{*} C & -P \bar{F}^{*}+C^{T} V_{1}^{* T} K_{1}^{*} \\
* & C^{T} V_{2}^{* T} V_{2}^{*} C+Q & C^{T} V_{2}^{* T} K_{1}^{*} \\
* & * & -\beta^{2} F_{\theta}^{T} F_{\theta}+K_{1}^{* T} K_{1}^{*}
\end{array}\right]}\left[\begin{array}{c}
e_{f \bar{\theta}}(t) \\
e_{f} \bar{\theta}(t-\tau) \\
\bar{\theta}(t)
\end{array}\right] .
$$

Hence, $\Xi>0$ implies $J_{r_{f \bar{\theta}}}>0 . \Xi>0$ is equivalent to

$$
\left[\begin{array}{ccc}
P F^{*}+F^{*^{T}} P+Q-C^{T} V_{1}^{* T} V_{1}^{*} C & -P G^{*}+C^{T} V_{1}^{* T} V_{2}^{*} C & -P \bar{F}^{*}+C^{T} V_{1}^{* T} K_{1}^{*}  \tag{39}\\
* & -Q-C^{T} V_{2}^{* T} V_{2}^{*} C & C^{T} V_{2}^{* T} K_{1}^{*} \\
* & * & \beta^{2} F_{\theta}^{T} F_{\theta}-K_{1}^{* T} K_{1}^{*}
\end{array}\right]<0 .
$$

Then, (39) can be written as

$$
\left[\begin{array}{ccc}
P F^{*}+F^{*^{T}} P+Q-2 C^{T} V_{1}^{* T} V_{1}^{*} C & -P G^{*}+C^{T} V_{1}^{* T} V_{2}^{*} C & -P \bar{F}^{*}+C^{T} V_{1}^{* T} K_{1}^{*}  \tag{40}\\
* & -Q-2 C^{T} V_{2}^{* T} V_{2}^{*} C & C^{T} V_{2}^{* T} K_{1}^{*} \\
* & * & \beta^{2} F_{\theta}^{T} F_{\theta}-2 K_{1}^{* T} K_{1}^{*}
\end{array}\right]+\left[\begin{array}{llll}
C^{T} V_{1}^{* T} & C^{T} V_{2}^{* T} & K_{1}^{* T}
\end{array}\right] I\left[\begin{array}{c}
V_{1}^{*} C \\
V_{2}^{*} C \\
K_{1}^{*}
\end{array}\right]<0 .
$$

Making use of Lemma 2 with $\varepsilon=1$, we have

$$
\begin{align*}
& -C^{T} V_{1}^{* T} V_{1}^{*} C \leq\left(V_{1 c}^{n}\right)^{T} V_{1 c}^{n}-\left(V_{1 c}^{n}\right)^{T} V_{1}^{*} C-C^{T} V_{1}^{* T} V_{1 c}^{n} \\
& -C^{T} V_{2}^{* T} V_{2}^{*} C \leq\left(V_{2 c}^{n}\right)^{T} V_{2 c}^{n}-\left(V_{2 c}^{n}\right)^{T} V_{2}^{*} C-C^{T} V_{2}^{* T} V_{2 c}^{n}  \tag{41}\\
& -K_{1}^{* T} K_{1}^{*} \leq\left(K_{1}^{n}\right)^{T} K_{1}^{n}-\left(K_{1}^{n}\right)^{T} V_{1}^{*} F_{y} F_{\theta}-F_{\theta}^{T} F_{y}^{T} V_{1}^{* T} K_{1}^{n}
\end{align*}
$$

Applying Schur complement to (40) and changing variables $\varphi_{1}\left(V_{1}^{*}, V_{1 c}^{n}\right), \varphi_{2}\left(V_{2}^{*}, V_{2 c}^{n}\right)$, and $\varphi_{3}\left(K_{1}^{*}, K_{1}^{n}\right)$, the LMI (32) is obtained. To overcome the infeasibility of (33), the same variables as those selected in Theorem 3 are used. It completes the proof.

Corollary 5. The system is asymptotically stable and satisfies (16) if there exists symmetric positive matrices $P, Q, V_{1}^{*}, V_{2}^{*}$, $\Phi_{1}^{*}$, $\Phi_{2}^{*}$, and $\Phi_{3}^{*}$ such that the LMIs (26) and (32) hold.

Remark 6. It is desired to obtain a reference residual system which has maximum sensitivity to the fault as well as the minimum sensitivity to the exogenous signal. This aim can be formulated by performance index defined by inf $\alpha / \beta$. To this end, an iterative optimization method presented in [6] is
developed for the proposed structure. The procedures of this method are as follows.
(1) Choose appropriate values of $\alpha$ and $\beta$.
(2) Solve the LMI (26), and find a feasible solution for $P$, $Q, V_{1}^{*}, V_{2}^{*}, \Phi_{1}^{*}, \Phi_{2}^{*}$, and $\Phi_{3}^{*}$ matrices.
(3) Set $V_{1 c}^{n}=V_{1}^{* n-1} C, V_{2 c}^{n}=V_{2}^{* n-1} C, K_{11}^{n}=V_{1}^{* n-1} F_{y} F_{\theta}$, and $K_{12}^{n}=V_{2}^{* n-1} F_{y} F_{\theta}$. Then, solve (26) and (32) by increasing $n$ to find a feasible solution for $P, \mathrm{Q}, V_{1}^{*}$, $V_{2}^{*}, \Phi_{1}^{*}, \Phi_{2}^{*}$, and $\Phi_{3}^{*}$.
(4) Increase $\beta$ and decrease $\alpha$ and go to step 2 . Continue this procedure until the feasible solution cannot be found for LMIs (26) and (32).
(b) Robust UIO Design (Step 2). As mentioned before, the residual signal generator system is obtained by minimizing (17). To this end, Theorem 8 is presented which guarantees that the overall system (19) is asymptotically stable and performance index (17) is minimized. Before presenting Theorem 8, the following theorem is presented which helps prove Theorem 8.
Theorem 7. For a given $\gamma>0$ and the system (37), if there exist symmetric positive matrices $P, Q$ and constants $\varepsilon_{1}, \varepsilon_{2}$, and $\varepsilon_{3}$ such that the LMI (44) holds, then the system (37) is asymptotically stable and $\|v(t)\|_{2} \leq \gamma\|u(t)\|_{2}$ :

$$
\begin{align*}
& \dot{\chi}(t)=(\widetilde{A}+\Delta \widetilde{A}) \chi(t)+\left(\widetilde{A}_{d}+\Delta \widetilde{A}_{d}\right) \chi(t-\tau)+\left(\widetilde{B}_{u}+\Delta \widetilde{B}_{u}\right) u(t),  \tag{42}\\
& v(t)=\widetilde{C}_{1} \chi(t)+\widetilde{C}_{2} \chi(t-\tau)+\widetilde{D} u(t),  \tag{43}\\
& {\left[\begin{array}{ccccccc}
P \widetilde{A}+\widetilde{A}^{T} P+Q+\varepsilon_{1}^{-1} \widetilde{N}_{1}^{T} \widetilde{N}_{1} & P \widetilde{A}_{d} & P \widetilde{B}_{u} & \widetilde{C}_{1}^{T} & P \widetilde{M}_{1} & P \widetilde{M}_{2} & P \widetilde{M}_{3} \\
* & -Q+\varepsilon_{2}^{-1} \widetilde{N}_{2}^{T} \widetilde{N}_{2} & \mathbf{0} & \widetilde{C}_{2}^{T} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & -\gamma^{2} I+\varepsilon_{3}^{-1} \widetilde{N}_{3}^{T} \widetilde{N}_{3} & \widetilde{D}^{T} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & * & -I & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & * & * & -\varepsilon_{1}^{-1} I & \mathbf{0} & \mathbf{0} \\
* & * & * & * & * & -\varepsilon_{2}^{-1} I & \mathbf{0} \\
* & * & * & * & * & * & -\varepsilon_{3}^{-1} I
\end{array}\right]<\mathbf{0} .} \tag{44}
\end{align*}
$$

Proof. Define the following Lyapunov-Krasovskii function:

$$
\begin{equation*}
V(t)=\chi^{T}(t) P \chi(t)+\int_{t-\tau}^{t} \chi^{T}(s) Q \chi(s) d s \tag{45}
\end{equation*}
$$

The performance index $\|v(t)\|_{2} \leq \gamma\|u(t)\|_{2}$ can be written as

$$
\begin{align*}
J_{v}=\int_{0}^{\infty} & \left(v^{T}(t) v(t)-\gamma^{2} u^{T}(t) u(t)+\dot{V}(t)\right) d t  \tag{46}\\
& +V(0)-V(\infty) .
\end{align*}
$$

$$
\begin{align*}
J_{r_{e}} \leq & \int_{0}^{\infty}\left[\begin{array}{c}
\chi(t) \\
\chi(t-\tau) \\
u(t)
\end{array}\right]^{T} \underbrace{\left[\begin{array}{c}
P(\widetilde{A}+\Delta \widetilde{A})+(\widetilde{A}+\Delta \widetilde{A})^{T} P+Q+\widetilde{C}_{1}^{T} \widetilde{C}_{1} \\
* \\
*
\end{array}\right.}_{\Xi} \begin{array}{c}
P\left(\widetilde{A}_{d}+\Delta \widetilde{A}_{d}\right)+\widetilde{C}_{1}^{T} \widetilde{C}_{2} \\
-Q+\widetilde{C}_{2}^{T} \widetilde{C}_{2}
\end{array}  \tag{48}\\
& \times\left[\begin{array}{c}
* \\
\left.\widetilde{B}_{u}+\Delta \widetilde{B}_{u}\right)+\widetilde{C}_{1}^{T} \widetilde{D} \\
\mathbf{0}(t-\tau) \\
u(t)
\end{array}\right] d t .
\end{align*}
$$

Taking derivative from (45) and considering (37) yield
Assume $\chi(t)=\mathbf{0}$ for $t \in[-\tau, 0]$. Since $V(\infty)>0$, we have

$$
\begin{equation*}
J_{v} \leq \int_{0}^{\infty}\left(v^{T}(t) v(t)-\gamma^{2} u^{T}(t) u(t)+\dot{V}(t)\right) d t \tag{47}
\end{equation*}
$$

$\Xi<0$ implies $J_{r_{e}}<0$. The $\Xi<0$ can be written as

$$
\begin{align*}
& {\left[\begin{array}{ccc}
P \widetilde{A}+\widetilde{A}^{T} P+Q & P \widetilde{A}_{d} & P \widetilde{B}_{u} \\
* & -Q & \mathbf{0} \\
* & * & -\gamma^{2} I
\end{array}\right]} \\
& \quad+\left[\begin{array}{c}
\widetilde{C}_{1}^{T} \\
\widetilde{C}_{2}^{T} \\
\widetilde{D}^{T}
\end{array}\right] I\left[\begin{array}{lll}
\widetilde{C}_{1} & \widetilde{C}_{2} & \widetilde{D}
\end{array}\right]  \tag{49}\\
& \quad+\left[\begin{array}{ccc}
P \Delta \widetilde{A}+\Delta \widetilde{A}^{T} P & P \Delta \widetilde{A}_{d} & P \Delta \widetilde{B}_{u} \\
* & \mathbf{0} & \mathbf{0} \\
* & * & \mathbf{0}
\end{array}\right]<0
\end{align*}
$$

Using Lemma 2, one can write the following inequality:

$$
\begin{aligned}
& {\left[\begin{array}{ccc}
P \Delta \widetilde{A}+\Delta \widetilde{A}^{T} P & P \Delta \widetilde{A}_{d} & P \Delta \widetilde{B}_{u} \\
* & \mathbf{0} & \mathbf{0} \\
* & * & \mathbf{0}
\end{array}\right]} \\
& \leq\left[\begin{array}{ccc}
P \widetilde{M}_{1} & P \widetilde{M}_{2} & P \widetilde{M}_{3} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right]\left[\begin{array}{ccc}
\varepsilon_{1} I & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \varepsilon_{2} I & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \varepsilon_{3} I
\end{array}\right] \\
& \quad \times\left[\begin{array}{ccc}
P \widetilde{M}_{1} & P \widetilde{M}_{2} & P \widetilde{M}_{3} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right]+\left[\begin{array}{ccc}
\widetilde{N}_{1}^{T} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \widetilde{N}_{2}^{T} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \widetilde{N}_{3}^{T}
\end{array}\right] \\
& \quad \times\left[\begin{array}{ccc}
\varepsilon_{1} I & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \varepsilon_{2} I & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \varepsilon_{3} I
\end{array}\right]^{-1}\left[\begin{array}{ccc}
\widetilde{N}_{1} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \widetilde{N}_{2} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \widetilde{N}_{3}
\end{array}\right]
\end{aligned}
$$

Considering (49), (50) and using Schur complement (48) lead to (44). It completes the proof.

Theorem 8. For a given $\gamma>0$, if there exist symmetric positive matrices $P_{1}, P_{2}, P_{3}, Q_{1}, Q_{2}$, and $Q_{3}$, matrices $\Phi_{1}, \Phi_{2}$, and $\Phi_{3}$, and constants $\varepsilon_{1}, \varepsilon_{2}$, and $\varepsilon_{3}$ such that LMI $\left[s_{i j}\right]_{14 \times 14}<0$ holds, then the overall system (19) is asymptotically stable, and $J_{r_{e}}<\gamma$. The observer matrices are calculated by considering (4), (5), and conditions 2 to 5, and $Y=P_{1}^{-1} \Phi_{1}, L_{2}=P_{1}^{-1} \Phi_{2}$, and $L_{3}=$ $P_{1}^{-1} \Phi_{3}$. The LMI coefficients are defined as
$s_{1,1}=P_{1} F+\left(P_{1} F\right)^{T}+Q_{1}, \quad s_{1,4}=P_{1} G, \quad s_{1,8}=P_{1} \bar{F}$,
$s_{1,10}=P_{1} \bar{R}, \quad s_{1,11}=C^{T} V_{1}^{T}, \quad s_{1,12}=P_{1}\left(T M_{1}\right)$,
$s_{1,13}=P_{1}\left(T M_{2}\right), \quad s_{1,14}=P_{1}\left(T M_{3}\right)$,
$s_{2,2}=P_{2} F^{*}+\left(P_{2} F^{*}\right)^{T}+Q_{2}, \quad s_{2,5}=P_{2} G^{*}, \quad s_{2,8}=P_{2} \bar{F}^{*}$,
$s_{2,10}=P_{2} \bar{R}^{*}, \quad s_{2,11}=-C^{T} V_{1}^{* T}$,
$s_{3,3}=P_{3} A+\left(P_{3} A\right)^{T}+Q_{3}+\varepsilon_{1}^{-1} N_{1}^{T} N_{1}, \quad s_{3,6}=P_{3} A_{d}$,
$s_{3,7}=P_{3} B, \quad s_{3,8}=P_{3} K_{3}, \quad s_{3,9}=P_{3} E, \quad s_{3,10}=P_{3} K_{4}$,
$s_{3,12}=P_{3}\left(M_{1}\right), \quad s_{3,13}=P_{3}\left(M_{2}\right), \quad s_{3,14}=P_{3}\left(M_{3}\right)$,
$s_{4,4}=-Q_{1}, \quad s_{4,11}=C^{T} V_{2}^{T}$,
$s_{5,5}=-Q_{2}, \quad s_{5,11}=C^{T} V_{2}^{* T}, \quad s_{6,6}=-Q_{3}+\varepsilon_{2}^{-1} N_{2}^{T} N_{2}$,
$s_{7,7}=-\gamma I+\varepsilon_{3}^{-1} N_{3}^{T} N_{3}, \quad s_{8,8}=-\gamma I, \quad s_{8,11}=K_{1}^{T}-K_{1}^{* T}$,
$s_{9,9}=-\gamma I, \quad s_{10,10,}=-\gamma I, \quad s_{10,11}=K_{2}^{T}-K_{2}^{* T}$,
$s_{11,11}=-I, \quad s_{12,12}=-\varepsilon_{1}^{-1} I$,

$$
\begin{align*}
& s_{13,13}=-\varepsilon_{2}^{-1} I, \quad s_{14,14}=-\varepsilon_{3}^{-1} I, \\
& \text { otherwise } \quad s_{i, j}=\mathbf{0} \tag{51}
\end{align*}
$$

$$
\begin{gather*}
P_{1} T=P_{1}\left(I-\Theta_{1} C\right)-\Phi_{1}\left(\Theta_{2} C\right), \\
P_{1} F=P_{1}\left(A-\Theta_{1} C A\right)-\Phi_{1}\left(\Theta_{2} C A\right)-\Phi_{2} C, \\
P_{1} G=P_{1}\left(A_{d}-\Theta_{1} C A_{d}\right)-\Phi_{1}\left(\Theta_{2} C A_{d}\right)-\Phi_{3} C,  \tag{52}\\
P_{1} \bar{F}=\left[P_{1}\left(F_{x} F_{\theta}-\Theta_{1} C F_{x} F_{\theta}\right)-\Phi_{1}\left(\Theta_{2} C F_{x} F_{\theta}\right)-\Phi_{2} F_{y} F_{\theta}-P_{1}\left(\Theta_{1} F_{y} F_{\theta} A_{\theta}\right)-\Phi_{1}\left(\Theta_{2} F_{y} F_{\theta} A_{\theta}\right)-\Phi_{3} F_{y} F_{\theta}\right], \\
P_{1} \bar{R}=\left[P_{1}\left(R-\Theta_{1} C R\right)-\Phi_{1}\left(\Theta_{2} C R\right)-\Phi_{2} D-\Phi_{3} D-P_{1}\left(\Theta_{1} D\right)-\Phi_{1}\left(\Theta_{2} D\right)\right] .
\end{gather*}
$$

Proof. In Theorem 7 assume that $P=\operatorname{diag}\left(\begin{array}{lll}P_{1} & P_{2} & P_{3}\end{array}\right)$ and $Q=\operatorname{diag}\left(Q_{1} \quad Q_{2} \quad Q_{3}\right)$. Then, using system dynamic (19) it is straight forward to see that $s_{i j}$ are the same as (37). Without considering (52), the inequality (51) includes nonlinear terms of $P Y, P L_{2}$, and $P L_{3}$ which lead the LMI to be infeasible. To overcome this problem, define $\Phi_{1}=P_{1} Y, \Phi_{2}=P_{1} L_{2}$, and $\Phi_{3}=P_{1} L_{3}$. Using conditions 2,3 and (23) it can be seen that (52) makes the obtained LMI feasible. It completes the proof.

Remark 9. It should be noted that the present work differs from [18] in the following perspectives.
(a) The results in [18] are obtained without considering dynamic characteristic for fault; however, the current results are achieved by considering dynamic characteristic that is modeled by (3). Hence, the design procedure in [18] is not applicable for the current case.
(b) The residual signal $r(t)$ is constructed based on (14) which uses both estimation error and delay in estimation error; however, the residual signal in [18] is constructed using estimation error. Since two design parameters $V_{1}$ and $V_{2}$ appear in the LMIs, the obtained LMIs are more flexible.

Remark 10. After designing the FDI system, residual evaluation methods and appropriate level of threshold should be selected to take a decision about the occurrence of fault. According to (13) and (14), the residual signal for fault-free system $r^{0}(t)$ satisfies the following equation:

$$
\begin{align*}
\left\|r^{0}(t)\right\|_{2} & =\left\|r_{\bar{n}}(t)+r_{u}(t)\right\|_{2}  \tag{53}\\
& \leq\left\|r_{\bar{n}}(t)\right\|_{2}+\left\|r_{u}(t)\right\|_{2} \leq J_{\mathrm{th}, \bar{n}}+J_{\mathrm{th}, u},
\end{align*}
$$

where

$$
\begin{align*}
& J_{\mathrm{th}, \bar{n}}=\sup _{\left(\Delta A, \Delta B, \Delta A_{d}\right) \in \Omega_{i}}\left\|r_{\bar{n}}(t)\right\|_{2},  \tag{54}\\
& J_{\mathrm{th}, u}=\sup _{\left(\Delta A, \Delta B, \Delta A_{d}\right) \in \Omega_{i}}\left\|r_{u}(t)\right\|_{2} .
\end{align*}
$$

$J_{\text {th }, \bar{n}}$ can be computed offline, and under the assumption that $\bar{n} \in L_{2}$ we have $\sup _{\left(\Delta A, \Delta B, \Delta A_{d}\right) \in \Omega_{i}}\left\|r_{\bar{n}}(t)\right\|_{2}=M_{\bar{n}}$. Since the signal $u$ is supposed to be known online, the value of $J_{\text {th }, u}$ can be determined online by $J_{\text {th }, u}=\gamma_{u}\|u(t)\|_{2}$, where $\gamma_{u}=$ $\sup _{\left(\Delta A, \Delta B, \Delta A_{d}\right) \in \Omega_{i}}\left(\left\|r_{u}(t)\right\|_{2}\right) /\left(\|u(t)\|_{2}\right) \cdot \gamma_{u}$ can be computed by Theorem 7. Therefore, the threshold value can be evaluated by

$$
\begin{equation*}
J_{\text {th }}=M_{\bar{n}}+\gamma_{u}\|u(t)\|_{2} . \tag{55}
\end{equation*}
$$

Since values of (53) and (54) increase by passing the time and, consequently, need more memory in real application, one can use root-mean-square (RMS) norm of $r(t)$, defined in (56), to detect the fault signals:

$$
\begin{equation*}
\|r(t)\|_{2}^{T}=\int_{t_{1}}^{t_{2}} r^{T}(t) r(t) d t, \quad T=t_{2}-t_{1} \tag{56}
\end{equation*}
$$

where $T$ is designed parameter.

## 3. Simulation Results

The main objective of this section is to investigate the effectiveness of the designed UIO. To this end, a numerical example is used and simulation results are presented. Consider a system which is defined by (1) with the following matrices:

$$
\begin{array}{cc}
A=\left[\begin{array}{ccc}
-3.8 & 1.5 & -0.5 \\
0.5 & -3 & 1 \\
-0.3 & 0.7 & -2.4
\end{array}\right], \quad A_{d}=\left[\begin{array}{ccc}
0.4 & 0.1 & -0.2 \\
0.1 & -0.8 & 0.2 \\
0.7 & -0.1 & 0.5
\end{array}\right], \\
B=\left[\begin{array}{c}
0.1 \\
0.2 \\
-0.4
\end{array}\right], \quad F_{x}=\left[\begin{array}{c}
0.6 \\
-0.5 \\
0.4
\end{array}\right], \quad E=\left[\begin{array}{c}
-0.4 \\
0.1 \\
-0.3
\end{array}\right], \\
C=\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{array}\right], \quad F_{y}=\left[\begin{array}{c}
0.2 \\
0.8 \\
-1.2
\end{array}\right], \quad R=\left[\begin{array}{c}
0.1 \\
0.2 \\
-0.4
\end{array}\right], \\
D=\left[\begin{array}{c}
0.9 \\
0.2 \\
0.7
\end{array}\right] . \tag{57}
\end{array}
$$

Uncertainties are also defined by the following matrices in (2):

$$
\begin{align*}
& M_{1}=\left[\begin{array}{l}
0.1 \\
0.2 \\
0.1
\end{array}\right], \quad M_{2}=\left[\begin{array}{c}
0.1 \\
0 \\
-0.1
\end{array}\right], \quad M_{3}=\left[\begin{array}{c}
-0.1 \\
0.2 \\
0.1
\end{array}\right], \\
& N_{1}=\left[\begin{array}{lll}
0 & 0.1 & 0.3
\end{array}\right], \\
& N_{2} \\
&=\left[\begin{array}{lll}
0.1 & 0 & 0
\end{array}\right],  \tag{58}\\
& N_{3}=0.1 .
\end{align*}
$$

The dynamic characteristic of fault is considered as

$$
\begin{equation*}
A_{\theta}=0, \quad F_{\theta}=1 . \tag{59}
\end{equation*}
$$

The first step to design the fault detection system is to solve the LMIs (26) and (32) in Theorems 3 and 4. The Yalmip LMI toolbox is used to solve the LMIs. To start the iterative optimization method presented in Remark 6, the initial values $\alpha_{\text {int }}=3$ and $\beta_{\text {int }}=0$ are selected. Using this procedure, the following results are obtained:

$$
\begin{gather*}
V_{1}^{*}=\left[\begin{array}{ccc}
-0.0994 & -0.0532 & 0.2714 \\
-0.0532 & -5.4751 & 1.3859 \\
0.2714 & 1.3859 & -0.7651
\end{array}\right], \\
V_{2}^{*}=\left[\begin{array}{ccc}
-0.6315 & -0.8304 & 0.1871 \\
-0.8304 & 0.6711 & 0.5971 \\
0.1871 & 0.5971 & -0.5845
\end{array}\right], \\
\Phi_{1}^{*}=\left[\begin{array}{ccc}
662.5 & -250.8 & -2628.5 \\
-250.8 & 3855.7 & -281.2 \\
-2628.5 & -281.2 & 2055.8
\end{array}\right],  \tag{60}\\
\Phi_{2}^{*}=\left[\begin{array}{ccc}
984.8 & -1453.4 & -842.7 \\
-1453.4 & 2292 & 1253.8 \\
-842.7 & 1253.8 & 733.6
\end{array}\right], \\
\Phi_{3}^{*}=\left[\begin{array}{ccc}
-4.7395 & 13.0831 & 0.7977 \\
13.0831 & -8.5691 & -14.9116 \\
0.7977 & -14.9116 & -3.0645
\end{array}\right], \\
\alpha=2.4, \\
\beta=1 .
\end{gather*}
$$

Using these values, the LMI (51) is solved and the observer dynamic matrices are obtained as follows:

$$
\begin{aligned}
& F=\left[\begin{array}{lll}
-3.6557 & -3.3806 & 2.9608 \\
-0.3878 & -1.4059 & 0.3008 \\
-2.4404 & -3.2245 & 1.9455
\end{array}\right], \\
& G=\left[\begin{array}{lll}
-0.0559 & -0.7284 & -0.1686 \\
-0.0153 & -0.1629 & -0.0350 \\
-0.0445 & -0.557 & -0.1273
\end{array}\right], \\
& H=\left[\begin{array}{l}
0.6878 \\
0.1535 \\
0.5328
\end{array}\right],
\end{aligned}
$$



Figure 1: An abrupt fault occurs at $t=5$.
$L_{1}=\left[\begin{array}{ccc}0.0372 & -1.1695 & 0.8939 \\ -0.2145 & 0.7355 & 0.1978 \\ -0.7462 & -0.9036 & 1.6937\end{array}\right]$,
$K_{1 z}=\left[\begin{array}{lll}-0.7701 & -0.1957 & 0.9507 \\ -0.1730 & -0.0567 & 0.2159 \\ -0.5857 & -0.1278 & 0.7254\end{array}\right]$,
$K_{2 z}=\left[\begin{array}{lll}0.2122 & -0.3395 & -0.7167 \\ 0.0496 & -0.0777 & -0.1591 \\ 0.1603 & -0.2648 & -0.5534\end{array}\right]$.
(61)

To verify the sensitivity of designed UIO, an abrupt fault, shown in Figure 1, occurs in the 4 seconds elapsed from running of the system. The step disturbance signal exerted to the system between 3 to 7 seconds. The noise signal is assumed to be white Gaussian noise with power 0.0005 , and the uncertainty $\Sigma_{i}(t)$ is considered sinusoidal signal. The residual signals are shown in Figure 2. It can be seen that the residual signals change when the fault occurs; however, the residual signals do not show any sensitivity to the exerted disturbance. The value of threshold $J_{\text {th }}$ is presented in Figure 3. This figure indicates that the fault is detected rapidly and the difference between threshold and norm of faulty residual signal is high enough to detect the occurrence of fault in the system.

The RMS of residual signals (56) has been depicted in Figure 4. It can be seen that the RMS of faulty signals suddenly changes in contrast to RMS of fault-free signals. Therefore, the occurrence of fault can be effectively realized.

## 4. Conclusions

In this paper, a novel UIO-based residual generator is developed for robust fault detection purposes. The developed method is applicable to a variety of linear uncertain timedelay systems. The proposed approach is able to decouple


Figure 2: Residual signals of UIO fault detection filter.


Figure 3: Euclidean norm of residual fault-free and faulty signals and $J_{\text {th }}$.
thoroughly exogenous disturbances while minimizing uncertainties and noise effects. The fault effect is also maximized at the same time. To this end, first, the optimal fault detection filter is designed for system without considering uncertainties. Then, the fault detection filter is designed so that the $H_{\infty}$ norm between the fault detection filter and the optimal one is minimized. Superiority of the proposed approach has been verified through a numerical example. Simulation results show that the proposed approach is able to detect dynamic faults. As a future work, one can extend this approach to nonlinear systems and descriptor system. Moreover, developing


FIGURE 4: Detecting occurrence of fault using RMS norm of $\|r(t)\|_{2}^{T}$.
the data framework for designing unknown input observer for time-delay system is an interesting area.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

[1] L. Zhang, E.-K. Boukas, L. Baron, and H. R. Karimi, "Fault detection for discrete-time Markov jump linear systems with partially known transition probabilities," International Journal of Control, vol. 83, no. 8, pp. 1564-1572, 2010.
[2] D. Dongsheng, B. Jiang, P. Shi, and H. R. Karimi, "Fault detection for continues-time switched systems under asynchronous switching," International Journal of Robust and Nonlinear Control, 2013.
[3] J. Yu, G. Sun, and H. R. Karimi, "Fault reconstruction based cascaded sliding mode observers for descriptor linear systems," Mathematical Problems in Engineering, vol. 2012, Article ID 623426, 20 pages, 2012.
[4] S. Yin, X. Yang, and H. R. Karimi, "Data-driven adaptive observer for fault diagnosis," Mathematical Problems in Engineering, vol. 2012, Article ID 832836, 21 pages, 2012.
[5] J. Chen and R. J. Patton, Robust Model-Based Fault Diagnosis for Dynamic Systems, Kluwer Academic Publishers, 1999.
[6] J. Guo, X. Huang, and Y. Cui, "Design and analysis of robust fault detection filter using LMI tools," Computers and Mathematics with Applications, vol. 57, no. 11-12, pp. 1743-1747, 2009.
[7] M. Zhong, S. X. Ding, T. Bingyong, T. Jeinsch, and M. Sader, "An LMI approach to design robust fault detection observers," in Proceedings of the 4th World Congress on Intelligent Control and Automation, pp. 2705-2709, June 2002.
[8] M. Zhong, S. X. Ding, J. Lam, and H. Wang, "An LMI approach to design robust fault detection filter for uncertain LTI systems," Automatica, vol. 39, no. 3, pp. 543-550, 2003.
[9] T. Li, L. Wu, and X. Wei, "Robust fault detection filter design for uncertain LTI systems based on new bounded real lemma," International Journal of Control, Automation and Systems, vol. 7, no. 4, pp. 644-650, 2009.
[10] M. Zhong, S. X. Ding, J. Lam, and C. Zhang, "Fault detection filter design for LTI system with time delays," in Proceedings of the 42nd IEEE Conference on Decision and Control, pp. 14671472, December 2003.
[11] L. Bai, Z. Tian, and S. Shi, "Design of $\mathrm{H}_{\infty}$ robust fault detection filter for linear uncertain time-delay systems," ISA Transactions, vol. 45, no. 4, pp. 491-502, 2006.
[12] M. Zhong, H. Ye, T. Sun, and G. Wang, "An iterative LMI approach to RFDF for linear system with time-varying delays," Asian Journal of Control, vol. 8, no. 1, pp. 86-90, 2006.
[13] H. Wang, C. Wang, and H. Gao, "An LMI approach to robust fault detection filter design for uncertain state-delayed systems," in Advances in Machine Learning and Cybernetics, vol. 3930 of Lecture Notes in Computer Science, pp. 112-121, 2006.
[14] H. R. Karimi, M. Zapateiro, and N. Luo, "A linear matrix inequality approach to robust fault detection filter design of linear systems with mixed time-varying delays and nonlinear perturbations," Journal of the Franklin Institute, vol. 347, no. 6, pp. 957-973, 2010.
[15] M. Witczak, A. Obuchowicz, and J. Korbicz, "Genetic programming based approaches to identification and fault diagnosis of non-linear dynamic systems," International Journal of Control, vol. 75, no. 13, pp. 1012-1031, 2002.
[16] G.-R. Duan and R. J. Patton, "Robust fault detection using Luenberger-type unknown input observers-a parametric approach," International Journal of Systems Science, vol. 32, no. 4, pp. 533-540, 2001.
[17] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," Journal of Process Control, vol. 7, no. 6, pp. 403-424, 1997.
[18] S. Ahmadizadeh, J. Zarei, and H. R. Karimi, "Robust unknown input observer design for linear uncertain time delay systems with application to fault detection," Asian Journal of Control, vol. 16, pp. 1-14, 2014.
[19] W. Chen and M. Saif, "Fault detection and isolation based on novel unknown input observer design," in Proceedings of the American Control Conference, pp. 5129-5134, June 2006.
[20] S. Mondal, G. Chakraborty, and K. Bhattacharyya, "LMI approach to robust unknown input observer design for continuous systems with noise and uncertainties," International Journal of Control, Automation and Systems, vol. 8, no. 2, pp. 210-219, 2010.
[21] G.-Y. Tang and J. Li, "Optimal fault diagnosis for systems with delayed measurements," IET Control Theory and Applications, vol. 2, no. 11, pp. 990-998, 2008.
[22] Y.-M. Fu, G.-R. Duan, and S.-M. Song, "Design of unknown input observer for linear time-delay systems," International Journal of Control, Automation and Systems, vol. 2, no. 4, pp. 530-535, 2004.

# Adaptive Neural Network Dynamic Inversion with Prescribed Performance for Aircraft Flight Control 

Wendong Gai, ${ }^{1,2}$ Honglun Wang, ${ }^{2}$ Jing Zhang, ${ }^{1}$ and Yuxia Li $^{1}$<br>${ }^{1}$ College of Information and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266590, China<br>${ }^{2}$ Science and Technology on Aircraft Control Laboratory, Beihang University, Beijing 100191, China

Correspondence should be addressed to Honglun Wang; hl_wang_2002@126.com
Received 16 July 2013; Revised 23 September 2013; Accepted 4 October 2013
Academic Editor: Dewei Li
Copyright © 2013 Wendong Gai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

An adaptive neural network dynamic inversion with prescribed performance method is proposed for aircraft flight control. The aircraft nonlinear attitude angle model is analyzed. And we propose a new attitude angle controller design method based on prescribed performance which describes the convergence rate and overshoot of the tracking error. Then the model error is compensated by the adaptive neural network. Subsequently, the system stability is analyzed in detail. Finally, the proposed method is applied to the aircraft attitude tracking control system. The nonlinear simulation demonstrates that this method can guarantee the stability and tracking performance in the transient and steady behavior.


## 1. Introduction

Flight control design for aircraft continues to be one of the most important problems in the world of automatic control. The problem is driven by the nonlinear and uncertain nature of aircraft dynamics. Traditionally, the solution to this problem is to design the linear controller using linearized aircraft models at multiple trimmed conditions. And this procedure is time consuming and expensive.

Control of aircraft by dynamic model inversion is well known and has been applied to the control of high angle of attack fighter aircraft [1, 2]. The primary drawback of dynamic inversion for aircraft flight control is the need for high-fidelity nonlinear model which must be inverted in real time. However, it is difficult to obtain the exact aircraft dynamic model in practice. The neural network augmented model inversion in the attitude angular loop is implemented to compensate the model inversion error, and it uses proportional-derivative desired dynamics to design the attitude control system for the helicopter [3] and tilt-rotor aircraft [4].

The asymptotic tracking can be achieved using this method. However, the transient behavior of the output signals could be oscillatory when the tracking error magnitude is
decreased by increasing the adaption rate. Several solutions [5-8] have been proposed to overcome this problem. These methods guarantee the convergence of tracking error, but the required tracking error upper bounds can't be accurately computed. A new adaptive control method with prescribed performance is presented in [9], and this method guarantees the transient state tracking error in the prespecified performance bound. And this method is used to improve the performance of the planar two-link articulated manipulator [10, 11] and the 6-DOF PUMA 560 arm [12].

It is very important for aircraft to track the attitude command with a desired transient and steady performance, when the aircraft finishes the special flight tasks, such as automated aerial refueling $[13,14]$ and transition flight control $[15,16]$.

In this paper, we will investigate the aircraft attitude control problem of guaranteeing transient and steady performance in the adaptive compensation control system. By employing the prescribed performance bounds proposed in [9], we propose a new adaptive neural network dynamic inversion method. With certain transformation method, a new transformed error system is obtained through considering the prescribed performance bound into the original attitude control system. An adaptive dynamic inversion controller is designed for the transformed system. It is ensured
that the tracking error is guaranteed inside the prescribed error bound as long as the transformed error system is stable.

The paper is organized as follows: the problem and the control configuration are introduced in Section 2. Section 3 presents the adaptive neural network dynamic inversion with prescribed performance design, stability analysis, model error analysis, and neural network structure. And the simulations are described in Section 4. Finally, this paper concludes in Section 5.

## 2. Aircraft Nonlinear Attitude Angle Model

The aircraft nonlinear attitude dynamic model can be presented as

$$
\begin{align*}
& \dot{\phi}=p+(r \cos \phi+q \sin \phi) \tan \theta, \\
& \dot{\theta}=q \cos \phi-r \sin \phi,  \tag{1}\\
& \dot{\psi}=\frac{(r \cos \phi+q \sin \phi)}{\cos \theta}, \\
& \dot{p}=\left(c_{1} r+c_{2} p\right) q+c_{3} \bar{L}+c_{4} N, \\
& \dot{q}=c_{5} p r-c_{6}\left(p^{2}-r^{2}\right)+c_{7} M,  \tag{2}\\
& \dot{r}=\left(c_{8} p-c_{2} r\right) q+c_{4} \bar{L}+c_{9} N,
\end{align*}
$$

where $\phi, \theta$, and $\psi$ are the roll, pitch, and yaw attitude angles. $p, q$, and $r$ are the roll, pitch, and yaw angular rates. $c_{1}, \ldots, c_{9}$ can be found in [17]. $\bar{L}, M$, and $N$ are the roll, pitch, and yaw moments, which can be described as

$$
\begin{align*}
& \bar{L}=\frac{\rho_{a} V^{2} S b C_{l}}{2}, \\
& M=\frac{\rho_{a} V^{2} S \bar{c} C_{m}}{2},  \tag{3}\\
& N=\frac{\rho_{a} V^{2} S b C_{n}}{2},
\end{align*}
$$

where $\rho_{a}$ is the air density, $S$ is the wing reference area, $b$ is the wing span, $V$ is the flight velocity, and $\bar{c}$ is the wing mean geometric chord. $C_{l}, C_{m}$, and $C_{n}$ are the rolling, pitching, and yawing moment coefficients described as

$$
\begin{align*}
& C_{l}=C_{l \beta} \beta+C_{l \bar{p}} \bar{p}+C_{l \bar{r}} \bar{r}+C_{l \delta_{a}} \delta_{a}+C_{l \delta_{r}} \delta_{r}, \\
& C_{m}=C_{m, \alpha=0}+C_{m \alpha} \alpha+C_{m \bar{q}} \bar{q}+C_{m \bar{\alpha}} \overline{\dot{\alpha}}+C_{m \delta_{e}} \delta_{e},  \tag{4}\\
& C_{n}=C_{n \beta} \beta+C_{n \bar{p}} \bar{p}+C_{n \bar{r}} \bar{r}+C_{n \delta_{a}} \delta_{a}+C_{n \delta_{r}} \delta_{r},
\end{align*}
$$

where $C_{(*)}$ is the aerodynamic derivatives. $\alpha$ and $\beta$ are the angles of attack and sideslip. $\delta_{a}, \delta_{e}$, and $\delta_{r}$ are the aileron, elevator, and rudder deflections, which are the control actuators of the aircraft. $\bar{p}, \bar{q}, \bar{r}$, and $\bar{\alpha}$ are defined by

$$
\begin{array}{llrl}
\bar{p} & =p b /(2 V), & \bar{r} & =r b /(2 V) \\
\bar{q} & =q \bar{c} /(2 V), & \overline{\dot{\alpha}}=\dot{\alpha} \bar{c} /(2 V) \tag{5}
\end{array}
$$

and $\dot{\alpha}$ is the derivative of the angle of attack.

Substituting (3)-(4) into (2), and (2) can be rewritten in the affine nonlinear form as

$$
\left[\begin{array}{c}
\dot{p}  \tag{6}\\
\dot{q} \\
\dot{r}
\end{array}\right]=\left[\begin{array}{l}
f_{p} \\
f_{q} \\
f_{r}
\end{array}\right]+G_{u}\left[\begin{array}{l}
\delta_{a} \\
\delta_{e} \\
\delta_{r}
\end{array}\right],
$$

where $f_{q}, f_{q}, f_{r}$, and $G_{u}$ are

$$
\begin{gather*}
f_{p}=\left(c_{1} r+c_{2} p\right) q+c_{3} M_{x}^{0}+c_{4} M_{z}^{0}, \\
f_{q}=c_{5} p r-c_{6}\left(p^{2}-r^{2}\right)+c_{7} M_{y}^{0}  \tag{7}\\
f_{r}=\left(c_{8} p-c_{2} r\right) q+c_{4} M_{x}^{0}+c_{9} M_{z}^{0}, \\
G_{u}=\frac{\rho_{a} V^{2} S}{2}\left[\begin{array}{ccc}
c_{3} & 0 & c_{4} \\
0 & c_{7} & 0 \\
c_{4} & 0 & c_{9}
\end{array}\right]\left[\begin{array}{ccc}
b C_{l \delta_{a}} & 0 & b C_{l \delta_{r}} \\
0 & \bar{c} C_{m \delta_{e}} & 0 \\
b C_{n \delta_{a}} & 0 & b C_{n \delta_{r}}
\end{array}\right], \tag{8}
\end{gather*}
$$

and

$$
\begin{align*}
& M_{x}^{0}=\frac{\rho_{a} V^{2} S b\left(C_{l \beta} \beta+C_{l \bar{p}} \bar{p}+C_{l \bar{r}} \bar{r}\right)}{2}, \\
& M_{y}^{0}=\frac{\rho_{a} V^{2} S \bar{c}\left(C_{m, \alpha=0}+C_{m \alpha} \alpha+C_{m \bar{q}} \bar{q}+C_{m \bar{\alpha}} \bar{\alpha}\right)}{2},  \tag{9}\\
& M_{z}^{0}=\frac{\rho_{a} V^{2} S b\left(C_{n \beta} \beta+C_{n \bar{p}} \bar{p}+C_{n \bar{r}} \bar{r}\right)}{2} .
\end{align*}
$$

According to (1), we can derive the second derivatives of attitude angles as follows:

$$
\begin{equation*}
[\ddot{\phi}, \ddot{\theta}, \ddot{\psi}]^{T}=L(\phi, \theta)[\dot{p}, \dot{q}, \dot{r}]^{T}+g(\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}) \tag{10}
\end{equation*}
$$

where

$$
L(\phi, \theta)=\left[\begin{array}{ccc}
1 & \sin \phi \tan \theta & \cos \phi \tan \theta  \tag{11}\\
0 & \cos \phi & -\sin \phi \\
0 & \sin \phi \sec \theta & \cos \phi \sec \theta
\end{array}\right]
$$

$$
g(\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi})=\left[\begin{array}{c}
\dot{\theta} \dot{\psi} \sec \theta+\dot{\phi} \dot{\theta} \tan \theta  \tag{12}\\
-\dot{\phi} \dot{\psi} \cos \theta \\
\dot{\phi} \dot{\theta} \sec \theta+\dot{\theta} \dot{\psi} \tan \theta
\end{array}\right] .
$$

Substituting (6) into (10), we obtain

$$
\begin{align*}
{[\ddot{\phi}, \ddot{\theta}, \ddot{\psi}]^{T}=} & L(\phi, \theta)\left[f_{p}, f_{q}, f_{r}\right]^{T}+g(\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}) \\
& +L(\phi, \theta) G_{u}\left[\delta_{a}, \delta_{e}, \delta_{r}\right]^{T} \tag{13}
\end{align*}
$$

## 3. Prescribed Performance-Based Adaptive Neural Network Dynamic Inversion Design

The aircraft attitude model shown in (13) can be represented in the following shorthand notation:

$$
\begin{equation*}
\ddot{x}=f(x, \dot{x})+g(x) u \tag{14}
\end{equation*}
$$



Figure 1: Adaptive neural network dynamic inversion with prescribed performance architecture.
where the controlled state $x=[\phi, \theta, \psi]^{T}$ and the control vector $u=\left[\delta_{a}, \delta_{e}, \delta_{r}\right]^{T} . f(x, \dot{x})$ and $g(x)$ are nonlinear functions.

The state tracking error is defined as

$$
\begin{equation*}
e(t)=x(t)-x_{d}(t), \tag{15}
\end{equation*}
$$

where $x_{d}(t)$ is the desired state vector.
The proposed control architecture of the aircraft attitude control system is shown in Figure 1.
3.1. Dynamic Inversion. This section will show a brief introduction of dynamic inversion. And the readers could derive much more details from the reference [2].

We seek to linearize a nonlinear system through computing dynamic inversion to cancel the nonlinearities in the system. The aircraft dynamics are shown in (14). The number of control inputs and controlled states must be the same; that is to say, the nonlinear function $g(x)$ is invertible. Then, the control input can be calculated by

$$
\begin{equation*}
u_{c}=g^{-1}(x)\left(u_{m}-f(x, \dot{x})\right), \tag{16}
\end{equation*}
$$

where $u_{m}$ is the desired response of $\ddot{x}$. Replacing the $u$ in the right of (14) by the $u_{c}$ from (16), we derive

$$
\begin{equation*}
\ddot{x}=u_{m} \tag{17}
\end{equation*}
$$

and any nonlinearities in $f(x, \dot{x})$ and $g(x)$ are cancelled.
The achieved system dynamics will match the chosen desired dynamics when there are no errors between the design model and real object. However, the model error is inevitable. So a new method is proposed to compensate the model error and guarantee the system performances in the transient and steady behavior.

### 3.2. Performance Function and Error Transformation

Definition 1 (see [9]). A smooth function $\rho: \mathfrak{R}_{+} \rightarrow \mathfrak{R}_{+}$can be called a performance function if the following conditions are satisfied:

$$
\begin{gather*}
\rho(t)>0, \quad \dot{\rho}(t)<0, \\
\lim _{t \rightarrow \infty} \rho(t)=\rho_{\infty}>0 . \tag{18}
\end{gather*}
$$

For example, a performance function is

$$
\begin{equation*}
\rho(t)=\left(\rho_{0}-\rho_{\infty}\right) e^{-l t}+\rho_{\infty}, \tag{19}
\end{equation*}
$$

where $\rho_{0}, \rho_{\infty}$ and $l$ are positive constants, $\rho_{0}$ is the initial tracking error $e(t)$, and $\rho_{\infty}$ is the maximum allowable tracking error $e(t)$ at the steady state. The decrement of tracking error $e(t)$ will decrease when the parameter $l$ decreases. And we can derive the first and second derivatives of $\rho(t)$ as follows:

$$
\begin{align*}
& \dot{\rho}(t)=-l\left(\rho_{0}-\rho_{\infty}\right) e^{-l t} \\
& \ddot{\rho}(t)=l^{2}\left(\rho_{0}-\rho_{\infty}\right) e^{-l t} \tag{20}
\end{align*}
$$

Then by satisfying the following condition:

$$
\begin{equation*}
-\underline{\delta} \rho(t)<e(t)<\bar{\delta} \rho(t), \quad \forall t \geq 0 \tag{21}
\end{equation*}
$$

where $0 \leq \underline{\delta}$, and $\bar{\delta} \leq 1$ are prescribed scalars; the objective of guaranteeing transient and steady performance can be derived.

Remark 2. According to (21), $-\underline{\delta} \rho(0)$ and $\bar{\delta} \rho(0)$ are the lower bound of the negative overshoot and upper bound of the positive overshoot of $e(t)$, respectively. And a lower bound of the convergence speed of $e(t)$ is introduced by the decreasing rate of $\rho(t)$.

Remark 3. By changing the parameters of performance function $\rho(t)$ and the positive prescribed scalars $\underline{\delta}$, and $\bar{\delta}$, the maximum overshoot and convergence rate of $e(t)$ can be modified.

To transform the original system with the constrained tracking error performance (in (21)) into an equivalent constrained one, an error transformation is introduced. And the error transformation is defined as

$$
\begin{equation*}
e(t)=\rho(t) S(\varepsilon) \tag{22}
\end{equation*}
$$

where $\varepsilon$ is the transformed error and a smooth and strictly increasing function $S$ has the following properties:

$$
\begin{gather*}
-\underline{\delta}<S(\varepsilon)<\bar{\delta}  \tag{23}\\
\lim _{\varepsilon \rightarrow-\infty} S(\varepsilon)=-\underline{\delta}, \quad \lim _{\varepsilon \rightarrow+\infty} S(\varepsilon)=\bar{\delta},  \tag{24}\\
S(0)=0 . \tag{25}
\end{gather*}
$$

According to the first property in (23) and $\rho(t)>0$, we have

$$
\begin{equation*}
-\underline{\delta} \rho(t)<\rho(t) S(\varepsilon)<\bar{\delta} \rho(t) . \tag{26}
\end{equation*}
$$

According to (19), we obtain

$$
\begin{equation*}
-\underline{\delta} \rho(t)<e(t)<\bar{\delta} \rho(t) . \tag{27}
\end{equation*}
$$

In addition, from the third property in (25), $\lim _{t \rightarrow \infty^{\infty}} e(t)=0$ can be achieved if $\lim _{t \rightarrow \infty} \varepsilon(t)=0$ is satisfied.

Then (22) can be described as

$$
\begin{equation*}
\varepsilon(t)=S^{-1}\left(\frac{e(t)}{\rho(t)}\right) \tag{28}
\end{equation*}
$$

Lemma 4 (see [9]). Consider system in (14), the transient and steady state tracking error behavior bounds described by the performance function $\rho(t)$ and the error transformation equation (22). The following results hold.
(a) The system in (14) is invariant under the error transformation equation (22).
(b) Stabilization of the transformed system using (28) is sufficient to guarantee the prescribed performance.

In what follows, an adaptive neural network dynamic inversion method is proposed to stabilize the transformed system using (28).

### 3.3. Controller Design and Stability Analysis

Assumption 5. The desired states $x_{d}(t)$ are known bounded time functions, with known bounded derivatives.

Assumption 6. The states $x(t)$ of the nonlinear system in (14) are available for measurement.

We define the following error function $E_{i}(t)$, which describes the dynamics of the new error system using the error transformation equation (28).

$$
\begin{equation*}
E_{i}(t)=\dot{\varepsilon}_{i}(t)+\eta_{i} \varepsilon_{i}(t), \quad i=p, q, r, \tag{29}
\end{equation*}
$$

where $\eta_{i}, i=p, q$, and $r$ are positive constants to be chosen.
We define

$$
\begin{equation*}
\lambda(t)=\frac{e(t)}{\rho(t)} . \tag{30}
\end{equation*}
$$

The derivative of (28) is

$$
\begin{equation*}
\dot{\varepsilon}=\frac{\partial S^{-1}}{\partial \lambda} \dot{\lambda} \tag{31}
\end{equation*}
$$

where

$$
\begin{equation*}
\dot{\lambda}=\frac{\dot{e}(t)}{\rho(t)}-\frac{e(t) \dot{\rho}(t)}{\rho^{2}(t)} \tag{32}
\end{equation*}
$$

And the second derivative of (28) is

$$
\begin{align*}
\ddot{\varepsilon}= & \frac{\partial\left(\partial S^{-1} / \partial \lambda\right)}{\partial \lambda}\left(\frac{\dot{e}(t)}{\rho(t)}-\frac{e(t) \dot{\rho}(t)}{\rho^{2}(t)}\right)^{2} \\
& +\frac{\partial S^{-1}}{\partial \lambda}\left[\frac{\ddot{e}(t)}{\rho(t)}-\frac{2 \dot{e}(t) \dot{\rho}(t)}{\rho^{2}(t)}\right.  \tag{33}\\
& \left.\quad-\frac{e(t) \ddot{\rho}(t)}{\rho^{2}(t)}+\frac{2 \dot{e}(t) \dot{\rho}^{2}(t)}{\rho^{3}(t)}\right] .
\end{align*}
$$

Then we compute the time derivative of $E_{p}(t)$ for the roll error as

$$
\begin{equation*}
\dot{E}_{p}(t)=\ddot{\varepsilon}_{p}(t)+\eta_{p} \dot{\varepsilon}_{p}(t) \tag{34}
\end{equation*}
$$

And the pitch and yaw errors are derived by the similar method.

Substituting (30)-(33) into (34), we obtain

$$
\begin{equation*}
\dot{E}_{p}=\frac{\partial S_{p}^{-1}}{\partial \lambda_{p}} \frac{1}{\rho_{p}(t)} \ddot{e}_{p}(t)+E_{p}^{M} \tag{35}
\end{equation*}
$$

where

$$
\begin{align*}
E_{p}^{M}= & \frac{\partial\left(\partial S_{p}^{-1} / \partial \lambda_{p}\right)}{\partial \lambda_{p}}\left(\frac{\dot{e}_{p}(t)}{\rho_{p}(t)}-\frac{e_{p}(t) \dot{\rho}_{p}(t)}{\rho_{p}^{2}(t)}\right)^{2} \\
& +\eta_{p} \frac{\partial S_{p}^{-1}}{\partial \lambda_{p}} \dot{\lambda}_{p}-\frac{\partial S_{p}^{-1}}{\partial \lambda_{p}} \\
& \times\left[\frac{2 \dot{e}_{p}(t) \dot{\rho}_{p}(t)}{\rho_{p}^{2}(t)}+\frac{e_{p}(t) \ddot{\rho}_{p}(t)}{\rho_{p}^{2}(t)}-\frac{2 \dot{e}_{p}(t) \dot{\rho}_{p}^{2}(t)}{\rho_{p}^{3}(t)}\right] \tag{36}
\end{align*}
$$

Then we can derive

$$
\begin{equation*}
\dot{E}=E_{R} \ddot{e}(t)+E_{M}, \tag{37}
\end{equation*}
$$

where $\dot{E}=\left[\dot{E}_{p}, \dot{E}_{q}, \dot{E}_{r}\right]^{T}, \ddot{e}(t)=\left[\ddot{e}_{p}(t), \ddot{e}_{q}(t), \ddot{e}_{r}(t)\right]^{T}, E_{M}=$ [ $\left.E_{p}^{M}, E_{q}^{M}, E_{r}^{M}\right]^{T}$, and $E_{R}$ is

$$
E_{R}=\left[\begin{array}{lll}
E_{p}^{R} & &  \tag{38}\\
& E_{q}^{R} & \\
& & E_{r}^{R}
\end{array}\right]
$$

where $E_{p}^{R}=\left(\partial S_{p}^{-1} / \partial \lambda_{p}\right) / \rho_{p}(t), E_{q}^{R}=\left(\partial S_{q}{ }^{-1} / \partial \lambda_{q}\right) / \rho_{q}(t)$, $E_{r}^{R}=\left(\partial S_{r}^{-1} / \partial \lambda_{r}\right) / \rho_{r}(t)$, and $\ddot{e}_{p}(t), \ddot{e}_{q}(t), \ddot{e}_{r}(t)$, are

$$
\begin{align*}
& \ddot{e}_{p}(t)=\ddot{\phi}-\ddot{\phi}_{d} \\
& \ddot{e}_{q}(t)=\ddot{\theta}-\ddot{\theta}_{d}  \tag{39}\\
& \ddot{e}_{r}(t)=\ddot{\psi}-\ddot{\psi}_{d} .
\end{align*}
$$

To simplify the controller design progress, we linearize (2) in an equilibrium point which is the steady wings-level flight state.

$$
\begin{align*}
{[\dot{p}, \dot{q}, \dot{r}]^{T}=} & A_{\omega}\left[V_{0}+\Delta V, \alpha_{0}+\Delta \alpha, \beta_{0}+\Delta \beta, p_{0}\right. \\
& \left.+\Delta p, q_{0}+\Delta q, r_{0}+\Delta r\right]^{T}  \tag{40}\\
+ & B_{\omega}\left[\delta_{a 0}+\Delta \delta_{a}, \delta_{e 0}+\Delta \delta_{e}, \delta_{r 0}+\Delta \delta_{r}\right]^{T}
\end{align*}
$$

where the $A_{\omega}$ and $B_{\omega}$ are the appropriate dimension constant matrixes, $\beta_{0}=p_{0}=q_{0}=r_{0}=\delta_{a 0}=\delta_{r 0}=0$. And $V_{0}, \alpha_{0}$, and $\delta_{e 0}$ are the flight velocity, angle of attack and elevator deflection in some equilibrium point, respectively. The symbol $\Delta$ represents the small perturbation from the equilibrium value.

According to (2), (13), (14), and (40), we can obtain

$$
\begin{equation*}
\ddot{x}=F(x)+G(x) \Delta u+\chi, \tag{41}
\end{equation*}
$$

where $\Delta u=\left[\Delta \delta_{a}, \delta_{e 0}+\Delta \delta_{e}, \Delta \delta_{r}\right]^{T}$, and

$$
\begin{align*}
F(x)= & {\left[F_{p}, F_{q}, F_{r}\right]^{T}=g(\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}) } \\
& +L(\phi, \theta) A_{\omega}\left[V_{0}+\Delta V, \alpha_{0}+\Delta \alpha, \Delta \beta, \Delta p, \Delta q, \Delta r\right]^{T}, \\
G(x)= & {\left[G_{p}, G_{q}, G_{r}\right]^{T}=L(\phi, \theta) B_{\omega} } \\
\chi= & {\left[\chi_{p}, \chi_{q}, \chi_{r}\right]^{T}, } \tag{42}
\end{align*}
$$

where $\chi$ is the model error which will be analyzed in Section 3.4.

The formula $\ddot{x}=F(x)+G(x) \Delta u$ in (41) is named as the design model in some equilibrium point, which is different from the real nonlinear model in (14). And the difference between the design model and the nonlinear model is represented by the symbol $\chi$, which will be compensated by the adaptive neural network.

Because there are three channels in the attitude control and the form of each channel is the same, consider the following Theorem 7 for the roll channel. And the pitch and yaw channels are similar.

Theorem 7. Considering Assumption 5, Assumption 6, and the nonlinear system in (14), all the signals are bounded, and the tracking error $e(t)$ satisfies the performance described by the performance function $\rho(t)$, if the control input of system satisfies the following formula.

The control input of roll channel is

$$
\begin{equation*}
u_{p}=\delta_{a}=G_{p}^{-1}\left[-F_{p}-\left(E_{p}^{R}\right)^{-1}\left(E_{p}^{M}+k_{p} E_{p}\right)+\ddot{\phi}_{d}-u_{p}^{a d}\right] \tag{43}
\end{equation*}
$$

The adaptive signal of roll channel is

$$
\begin{equation*}
u_{p}^{a d}=w_{p}^{T} g_{p} \tag{44}
\end{equation*}
$$

The neural network weight update law of roll channel is

$$
\begin{align*}
& \dot{\tilde{w}}_{p}= \begin{cases}\gamma_{p}\left(g_{p}\left(E_{p}\right)^{T} E_{p}^{R}+\sigma_{p} w_{p}\right), & \left\|E_{p}\right\|>\zeta_{p} \\
0 & \left\|E_{p}\right\| \leq \zeta_{p}\end{cases} \\
& \zeta_{p}=\frac{\left\|E_{p}^{R}\right\| h_{p}+\sqrt{\left(\left\|E_{p}^{R}\right\| h_{p}\right)^{2}+k_{p} \sigma_{p}\left(w_{p}^{\max }\right)^{2}}}{2 k_{p}} \tag{45}
\end{align*}
$$

where the vector $g_{p}$ is a set of basis functions to approximate the uncertainty and the neural network weight vector $w_{p}$ is the set of coefficients of each basis function in the roll channel. The adaptation gain $\gamma_{p}$ determines the learning rate of neural network. The $\sigma_{p}$ is a modification term to limit the growth of the neural network weights. The constant $k_{p}$ is positive. And the positive constant $h_{p}$ is the neural network approximate error which is bounded. The neural network weight error is

$$
\begin{equation*}
\widetilde{w}_{p}=w_{p}-w_{p}^{*} \tag{46}
\end{equation*}
$$

where the $w_{p}^{*}$ is the true value of the neural network weight in the roll channel.

Proof. A suitable Lyapunov function of roll channel will be

$$
V_{p}= \begin{cases}\frac{1}{2}\left(E_{p}\right)^{T} E_{p}+\frac{1}{2 \gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \widetilde{w}_{p}, & \left\|E_{p}\right\|>\zeta_{p},  \tag{47}\\ \frac{1}{2}\left(E_{p 0}\right)^{T} E_{p 0}+\frac{1}{2 \gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \widetilde{w}_{p}, & \left\|E_{p}\right\| \leq \zeta_{p},\end{cases}
$$

where $\left\|E_{p 0}\right\|=\zeta_{p}$ and $\zeta_{p}$ is to be defined later.
Firstly, if $\left\|E_{p}\right\|>\zeta_{p}$ is satisfied, then the time derivative of (47) is given by

$$
\begin{equation*}
\dot{V}_{p}=\left(E_{p}\right)^{T} \dot{E}_{p}+\frac{1}{\gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \dot{\tilde{w}}_{p} \tag{48}
\end{equation*}
$$

Substituting (37) into (48), we derive

$$
\begin{equation*}
\dot{V}_{p}=\left(E_{p}\right)^{T}\left[E_{p}^{M}+E_{p}^{R}\left(\ddot{\phi}-\ddot{\phi}_{d}\right)\right]+\frac{1}{\gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \dot{\widetilde{w}}_{p} \tag{49}
\end{equation*}
$$

Considering (41)-(42) and (49), we have

$$
\begin{align*}
\dot{V}_{p}= & \left(E_{p}\right)^{T}\left[E_{p}^{M}+E_{p}^{R}\left(F_{p}+G_{p} u_{p}+\chi_{p}-\ddot{\phi}_{d}\right)\right] \\
& +\frac{1}{\gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \dot{\tilde{w}}_{p} \tag{50}
\end{align*}
$$

Let the control input $u_{p}$ satisfy (43), then (50) can be described as

$$
\begin{equation*}
\dot{V}_{p}=\left(E_{p}\right)^{T}\left[-k_{p} E_{p}+E_{p}^{R}\left(\chi_{p}-u_{p}^{a d}\right)\right]+\frac{1}{\gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \dot{\widetilde{w}}_{p} \tag{51}
\end{equation*}
$$

Substituting (44)-(46) into (51), we obtain

$$
\begin{align*}
\dot{V}_{p}= & -k_{p}\left(E_{p}\right)^{T} E_{p}+\left(E_{p}\right)^{T} E_{p}^{R} \\
& \times\left(\chi_{p}-\left(w_{p}^{*}\right)^{T} g_{p}\right)+\sigma_{p}\left(\widetilde{w}_{p}\right)^{T} w_{p} . \tag{52}
\end{align*}
$$

By using the norms of the terms on the right side of (52), we obtain the following inequality:

$$
\begin{align*}
\dot{V}_{p} \leq & -k_{p}\left\|E_{p}\right\|^{2}+\left\|E_{p}\right\|\left\|E_{p}^{R}\right\|  \tag{53}\\
& \times\left\|\chi_{p}-\left(w_{p}^{*}\right)^{T} g_{p}\right\|+\sigma_{p}\left\|\widetilde{w}_{p}\right\|\left\|w_{p}\right\| .
\end{align*}
$$

In addition, the approximate error of neural network is bounded, so the following equation is satisfied:

$$
\begin{equation*}
\left\|\chi_{p}-\left(w_{p}^{*}\right)^{T} g_{p}\right\| \leq h_{p} . \tag{54}
\end{equation*}
$$

The maximum weight of ideal neural network in the roll channel is $w_{p}^{\max }$, so we have

$$
\begin{equation*}
\left\|w_{p}^{*}\right\| \leq w_{p}^{\max } \tag{55}
\end{equation*}
$$

Substituting (46) and (54)-(55) into (53), we get

$$
\begin{equation*}
\dot{V}_{p} \leq-k_{p}\left\|E_{p}\right\|^{2}+\left\|E_{p}\right\|\left\|E_{p}^{R}\right\| h_{p}+\sigma_{p}\left(w_{p}^{\max }\left\|w_{p}\right\|-\left\|w_{p}\right\|^{2}\right) \tag{56}
\end{equation*}
$$

Considering (56), we obtain

$$
\begin{align*}
\dot{V}_{p} \leq & -k_{p}\left\|E_{p}\right\|^{2}+\left\|E_{p}\right\|\left\|E_{p}^{R}\right\| h_{p} \\
& -\sigma_{p}\left(\left\|w_{p}\right\|-\frac{w_{p}^{\max }}{2}\right)^{2}+\sigma_{p}\left(\frac{w_{p}^{\max }}{2}\right)^{2} . \tag{57}
\end{align*}
$$

If the system is stable, then $\dot{V}_{p}<0$. And (57) can be transformed to the following formula:

$$
\begin{equation*}
k_{p}\left\|E_{p}\right\|^{2}-\left\|E_{p}\right\|\left\|E_{p}^{R}\right\| h_{p}-\sigma_{p}\left(\frac{w_{p}^{\max }}{2}\right)^{2}>0 \tag{58}
\end{equation*}
$$

Then we can derive

$$
\begin{equation*}
\left\|E_{p}\right\|>\frac{\left\|E_{p}^{R}\right\| h_{p}+\sqrt{\left(\left\|E_{p}^{R}\right\| h_{p}\right)^{2}+k_{p} \sigma_{p}\left(w_{p}^{\max }\right)^{2}}}{2 k_{p}}=\zeta_{p} \tag{59}
\end{equation*}
$$

Next, if $\left\|E_{p}\right\| \leq \zeta_{p}$ is satisfied, then the time derivative of (47) is derived by

$$
\begin{equation*}
\dot{V}_{p}=\frac{1}{\gamma_{p}}\left(\widetilde{w}_{p}\right)^{T} \dot{\widetilde{w}}_{p} \tag{60}
\end{equation*}
$$

Here the weight update law is $\dot{w}_{p}=\dot{\tilde{w}}_{p}=0$, and $\dot{V}_{p}=0$. Therefore, the system is stable, and all the signals are bounded. Considering Lemma 4, the tracking error $e(t)$ satisfied the performance described by the performance function $\rho(t)$.

This completes the proof.
3.4. Analysis of the Model Error. According to (2)-(5), the moment model is nonlinear, complicated, and must be continuously varying with the flight condition. For simplicity, the linear model of (40) in an equilibrium point is used to replace the nonlinear model of (2).

We define the model error $\Lambda=\left[\Lambda_{p}, \Lambda_{q}, \Lambda_{r}\right]^{T}$, which is the error between the linear model Equation (40) and the nonlinear model equation (6). And the $\Lambda$ is

$$
\begin{align*}
\Lambda=[ & \left.f_{p}, f_{q}, f_{r}\right]^{T}-A_{\omega}\left[V_{0}+\Delta V, \alpha_{0}+\Delta \alpha, \Delta \beta, \Delta p, \Delta q, \Delta r\right]^{T} \\
& +G_{u}\left[\delta_{a}, \delta_{e}, \delta_{r}\right]^{T}-B_{\omega}\left[\Delta \delta_{a}, \delta_{e 0}+\Delta \delta_{e}, \Delta \delta_{r}\right]^{T} . \tag{61}
\end{align*}
$$

Then (6) can be rewritten as

$$
\begin{align*}
{[\dot{p}, \dot{q}, \dot{r}]^{T}=} & A_{\omega}\left[V_{0}+\Delta V, \alpha_{0}+\Delta \alpha, \Delta \beta, \Delta p, \Delta q, \Delta r\right]^{T} \\
& +B_{\omega}\left[\Delta \delta_{a}, \delta_{e 0}+\Delta \delta_{e}, \Delta \delta_{r}\right]^{T}-\Lambda \tag{62}
\end{align*}
$$

Substituting (62) into (10), we have

$$
\begin{equation*}
\ddot{x}=F(x)+G(x) \Delta u-L(\phi, \theta) \Lambda . \tag{63}
\end{equation*}
$$

Comparing (63) to (41), we obtain

$$
\begin{equation*}
\chi=-L(\phi, \theta) \Lambda . \tag{64}
\end{equation*}
$$

Therefore, the model error mainly depends on the different equilibrium points, attitude angles, actuator deflections, and so on.
3.5. Neural Network Structure. The first step in determining the appropriate network structure is identifying the network inputs. Based on the analysis of model error sources described in Section 3.3, there are three main categories of inputs: the attitude angles, attitude angle rates, and actuator deflections.

A Sigma-Pi neural network [18] is used to compensate the model error $\chi$, and the basis function of the pitch channel $g_{q}$ is

$$
\begin{equation*}
g_{q}=\operatorname{kron}\left(\operatorname{kron}\left(C_{1 q}, C_{2 q}\right), C_{3 q}\right) \tag{65}
\end{equation*}
$$

where kron $(\cdot, \cdot)$ represents the Kronecker products and is defined as follows:

$$
\begin{equation*}
C_{1 q}=\left[1, \bar{\theta}, \bar{\theta}^{2}\right]^{T}, \quad C_{2 q}=[1, \bar{q}]^{T}, \quad C_{3 q}=\left[1, \bar{\delta}_{e}, \bar{\delta}_{e}^{2}\right]^{T}, \tag{66}
\end{equation*}
$$



Figure 2: Neural network structure.
where $\bar{\theta}, \bar{q}, \bar{\delta}_{c}$ and $\bar{\delta}_{e}$ are normalized variables between -1 and 1. The normalization function is

$$
\begin{equation*}
y=f(x)=\frac{2}{1+e^{-0.1 x}}-1 \tag{67}
\end{equation*}
$$

where $x$ is the input parameter and $y$ is the output parameter.
And a general description of the neural network is shown in Figure 2.

And the basis function of roll channel $g_{p}$ and the basis function of yaw channel $g_{r}$ can be derived similarly as follows:

$$
\begin{equation*}
g_{k}=\operatorname{kron}\left(\operatorname{kron}\left(\operatorname{kron}\left(\operatorname{kron}\left(C_{1}, C_{2}\right), C_{3}\right), C_{4}\right), C_{k}\right), \tag{68}
\end{equation*}
$$

where $k=p, r$. Then $C_{i}, i=1,2,3,4, k$ is

$$
\begin{array}{ll}
C_{1}=\left[1, \bar{\phi}, \bar{\phi}^{2}\right]^{T}, & C_{2}=[1, \bar{p}]^{T}, \quad C_{3}=[1, \bar{r}]^{T}, \\
C_{4}=\left[1, \bar{\psi}, \bar{\psi}^{2}\right]^{T}, \quad C_{p}=\left[1, \bar{\delta}_{a}, \bar{\delta}_{a}^{2}\right]^{T}, \quad C_{r}=\left[1, \bar{\delta}_{r}, \bar{\delta}_{r}^{2}\right]^{T} . \tag{69}
\end{array}
$$

## 4. Simulation Results

In this section, we consider the attitude angles control problem for a fixed-wing aircraft, and the initial flight state is the wings-level flight. Then the attitude angles commands in three channels will be tracked, respectively.

In the following simulation, the initial flight height and velocity are 6000 m and $190 \mathrm{~m} / \mathrm{s}$, and the initial attitude angles and angular rates including $\phi, \theta, \psi, p, q$, and $r$ are zeros. In addition, all the initial actuator deflections are zeros.

The error transformation function [19] in the simulation is described as

$$
\begin{equation*}
S(\varepsilon)=\frac{\bar{\delta} e^{(\varepsilon+y)}-\underline{\delta} e^{-(\varepsilon+y)}}{e^{(\varepsilon+y)}+e^{-(\varepsilon+y)}} \tag{70}
\end{equation*}
$$

where $y=\ln (\bar{\delta} / \underline{\delta}) / 2$. It can be shown that $S(\varepsilon)$ satisfies the properties in (23)-(25).

The attitude angles commands of three channels are transformed into the desired attitude angles commands


Figure 3: Command filter.

Table 1: Performance parameters.

| $\rho_{0}^{\phi}$ | $-12^{\circ}$ | $\rho_{0}^{\psi}$ | $-8^{\circ}$ | $\rho_{0}^{\theta}$ | $-10^{\circ}$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| $\rho_{\infty}^{\phi}$ | $-0.3^{\circ}$ | $\rho_{\infty}^{\psi}$ | $-0.2^{\circ}$ | $\rho_{\infty}^{\theta}$ | $-0.2^{\circ}$ |
| $l_{\phi}$ | 0.7 | $l_{\psi}$ | 0.7 | $l_{\theta}$ | 0.7 |
| $\bar{\delta}_{\phi}$ | 0.6 | $\underline{\delta}_{\psi}$ | 0.5 | $\underline{\delta}_{\theta}$ | 0.6 |
| $\bar{\delta}_{\phi}$ | 1 | $\bar{\delta}_{\psi}$ | 1 | $\bar{\delta}_{\theta}$ | 1 |

Table 2: Controller parameters.

| $k_{p}$ | 10 | $k_{r}$ | 10 | $k_{q}$ | 10 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $\gamma_{p}$ | 200 | $\gamma_{r}$ | 200 | $\gamma_{q}$ | 50 |
| $\sigma_{p}$ | 0.1 | $\sigma_{r}$ | 0.1 | $\sigma_{q}$ | 0.3 |

through the command filters. And the structure of command filter for the roll channel is shown in Figure 3. In addition, the desired attitude angles commands for yaw and pitch channels can be obtained by the similar command filters.

The command filter parameters are set as $\xi_{i}=1, \omega_{i}=2.5$, and $i=\phi, \psi, \theta$.

Design the control inputs with prescribed performance for three channels through the procedures in Section 3.2. The performance and controller parameters are shown in Tables 1 and 2.

Remark 8. For the performance function $\rho(t), \rho_{0}$ is derived by subtracting the attitude command from the initial attitude angle. $\rho_{\infty}$ is the allowable attitude tracking error at the steady state. And the decrement of tracking error $e(t)$ will decrease when the parameter $l$ decreases.

Remark 9. For the controller parameters, the adaptation gain $\gamma$ will improve the attitude tracking performance, especially, when there are much larger model errors. The $\sigma_{p}$ is a modification term to limit the growth of the neural network weights; therefore, it is small. The transient performance of attitude tracking error can be improved by increasing the parameter $k$; however, the increase will increase the magnitude of the control input. Then a compromise must be reached.


Figure 4: Responses of the attitude angles.


Figure 5: Tracking errors of the attitude angles.

The design model I is derived at the trimmed flight condition of 6000 m and $190 \mathrm{~m} / \mathrm{s}$, and the model error is small.

The aircraft tracks the attitude angles commands from the initial flight state. And the attitude angles tracking responses and tracking errors are shown in Figures 4 and 5.

The two methods have achieved the attitude angles command tracking. Figure 4 shows the better attitude responses are achieved by the proposed method compared to the
method in [20]. And the coupling among different channels is smaller when the proposed method is used. For example, the roll angle response has a less change when the aircraft tracks the yaw command. In Figure 5, the attitude angles tracking errors satisfy the prescribed performance bound with the proposed method in the dynamic and steady state. The main reason is that the method in [20] does not consider the performance bound defined by the performance function $\rho(t)$ in the design process.


Figure 6: Responses of the Attitude angles with model error.

_- Roll tracking error
$\ldots \ldots \rho_{p}(t)$
$\ldots-0.6 \rho_{p}(t)$
(a)

(b)


$$
\begin{aligned}
& \text { _ Pitching tracking error } \\
& \ldots . \rho_{q}(t) \\
& =--0.6 \rho_{q}(t)
\end{aligned}
$$

(c)

Figure 7: Tracking errors of the attitude angles with model error.

(c)

Figure 8: Deflections of the control actuators in two design models.


Figure 9: The outputs of neural network in three channels.

In the real flight control system, there must be the model error. In order to verify that the similar tracking performance is also achieved when there is the large model error, we have conducted the following simulation study.

The flight condition is the same, and the initial flight height and velocity are 6000 m and $190 \mathrm{~m} / \mathrm{s}$. However, the design model II used to design the attitude angles controllers is derived at the trimmed flight condition of 4000 m and $150 \mathrm{~m} / \mathrm{s}$. Apparently, the model error is large.

And the attitude angles tracking responses and tracking errors are shown in Figures 6 and 7.

Figures 6 and 7 show the attitude angles tracking errors still satisfy the prescribed performance bound, although the model error is large in this situation. In addition, Figures 6 and 7 show the track performance is similar when the different design models are used.

The control actuators deflections for three channels are compared in Figure 8 when the two design models are used.

Figure 8 shows the deflections of the control actuators using the design model II increase to derive the desired attitude angles tracking performance. In addition, the outputs of neural network in three channels are shown in Figure 9.

Figure 9 shows the outputs of neural network using the design model II are larger than the one using the design model I. The main reason is that the model error is larger when the design model II is used, and the larger outputs of neural network are used to compensate the large model error.

## 5. Conclusion

In this paper, an adaptive neural network dynamic inversion with prescribed performance method is proposed for aircraft attitude control. By incorporating the adaptive neural network dynamic inversion with the prescribed performance concept, the proposed method guarantees the system tracking error satisfies the prescribed performance bound in the transient and steady behavior. The nonlinear simulation of the aircraft also verifies the effectiveness of the proposed approach.

Further investigation is needed for the situations in the presence of the external wind disturbance and unmodeled dynamics. And, these design parameters in this method should be decreased and optimized to achieve a real application.

## Acknowledgments

This work was supported by the Program for New Century Excellent Talents in University (Grant no. NCET-10-0032) and by the National Natural Science Foundation of China (Grant no. 61175084).

## References

[1] S. A. Snell, D. F. Enns, and W. L. Garrard Jr., "Nonlinear inversion flight control for a supermaneuverable aircraft," Journal of Guidance, Control, and Dynamics, vol. 15, no. 4, pp. 976-984, 1992.
[2] D. Enns, D. Bugajski, R. Hendrick, and G. Stein, "Dynamic inversion: an evolving methodology for flight control design," International Journal of Control, vol. 59, no. 1, pp. 71-91, 1994.
[3] J. Leitner, A. Calise, and J. V. R. Prasad, "Analysis of adaptive neural networks for helicopter flight control," Journal of Guidance, Control, and Dynamics, vol. 20, no. 5, pp. 972-979, 1997.
[4] R. Rysdyk and A. J. Calise, "Robust nonlinear adaptive flight control for consistent handling qualities," IEEE Transactions on Control Systems Technology, vol. 13, no. 6, pp. 896-910, 2005.
[5] C. Cao and N. Hovakimyan, "Design and analysis of a novel $L_{1}$ adaptive control architecture with guaranteed transient performance," IEEE Transactions on Automatic Control, vol. 53, no. 2, pp. 586-591, 2008.
[6] W. Gai, H. Wang, and D. Li, "Trajectory tracking for automated aerial refueling based on adaptive dynamic inversion," Journal of Beijing University of Aeronautics and Astronautics, vol. 38, no. 5, pp. 585-590, 2012 (Chinese).
[7] H. Xu and P. A. Ioannou, "Robust adaptive control for a class of MIMO nonlinear systems with guaranteed error bounds," IEEE Transactions on Automatic Control, vol. 48, no. 5, pp. 728-742, 2003.
[8] V. Stepanyan and K. Krishnakumar, "Adaptive control with reference model modification," Journal of Guidance, Control, and Dynamics, vol. 35, no. 4, pp. 1370-1374, 2012.
[9] C. P. Bechlioulis and G. A. Rovithakis, "Adaptive control with guaranteed transient and steady state tracking error bounds for strict feedback systems," Automatica, vol. 45, no. 2, pp. 532-538, 2009.
[10] C. P. Bechlioulis and G. A. Rovithakis, "Prescribed performance adaptive control for multi-input multi-output affine in the control nonlinear systems," IEEE Transactions on Automatic Control, vol. 55, no. 5, pp. 1220-1226, 2010.
[11] A. Kostarigka and G. Rovithakis, "Adaptive dynamic output feedback neural network control of uncertain MIMO nonlinear systems with prescribed performance," IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 1, pp. 138149, 2012.
[12] C. P. Bechlioulis, Z. Doulgeri, and G. A. Rovithakis, "Guaranteeing prescribed performance and contact maintenance via an approximation free robot force/position controller," Automatica, vol. 48, no. 2, pp. 360-365, 2012.
[13] M. L. Fravolini, A. Ficola, G. Campa, M. R. Napolitano, and B. Seanor, "Modeling and control issues for autonomous aerial refueling for UAVs using a probe-drogue refueling system," Aerospace Science and Technology, vol. 8, no. 7, pp. 611-618, 2004.
[14] J. P. Nalepka and J. L. Hinchman, "Automated aerial refueling: extending the effectiveness of unmanned air vehicles," in Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, pp. 240-247, August 2005, AIAA Paper 2005-6005.
[15] Y. Xili, F. Yong, and Z. Jihong, "Transition flight control of two vertical/short takeoff and landing aircraft," Journal of Guidance, Control, and Dynamics, vol. 31, no. 2, pp. 371-385, 2008.
[16] W. Gai and H. Wang, "Closed-loop dynamic control allocation for aircraft with multiple actuators," Chinese Journal of Aeronautics, vol. 26, no. 3, pp. 676-686, 2013.
[17] B. Stevens and L. Lewis, Aircraft Control and Simulation, Wiley, New York, NY, USA, 2nd edition, 2003.
[18] B. S. Kim and A. J. Calise, "Nonlinear flight control using neural networks," Journal of Guidance, Control, and Dynamics, vol. 20, no. 1, pp. 26-33, 1997.
[19] W. Wang and C. Wen, "Adaptive actuator failure compensation control of uncertain nonlinear systems with guaranteed transient performance," Automatica, vol. 46, no. 12, pp. 2082-2091, 2010.
[20] R. T. Rysdyk and A. J. Calise, "Adaptive model inversion flight control for tilt-rotor aircraft," Journal of Guidance, Control, and Dynamics, vol. 22, no. 3, pp. 402-407, 1999.

## Research Article

# Control Strategy Based on Wavelet Transform and Neural Network for Hybrid Power System 

Y. D. Song, ${ }^{1,2}$ Qian Cao, ${ }^{2}$ Xiaoqiang Du, ${ }^{2}$ and Hamid Reza Karimi ${ }^{3}$<br>${ }^{1}$ School of Automation, Chongqing University, Chongqing 400044, China<br>${ }^{2}$ School of Energy Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China<br>${ }^{3}$ Department of Engineering, Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Qian Cao; 403226080@qq.com
Received 31 July 2013; Accepted 26 September 2013
Academic Editor: Tao Zou
Copyright © 2013 Y. D. Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper deals with an energy management of a hybrid power generation system. The proposed control strategy for the energy management is based on the combination of wavelet transform and neural network arithmetic. The hybrid system in this paper consists of an emulated wind turbine generator, PV panels, DC and AC loads, lithium ion battery, and super capacitor, which are all connected on a DC bus with unified DC voltage. The control strategy is responsible for compensating the difference between the generated power from the wind and solar generators and the demanded power by the loads. Wavelet transform decomposes the power difference into smoothed component and fast fluctuated component. In consideration of battery protection, the neural network is introduced to calculate the reference power of battery. Super capacitor (SC) is controlled to regulate the DC bus voltage. The model of the hybrid system is developed in detail under Matlab/Simulink software environment.


## 1. Introduction

Faced with shortage of oil, rising of the petroleum price and the increasing pollution of the environment, people all over the world are searching for solutions to energy crises. Since the end of the twentieth century, great attention has been paid to renewable energy sources [1]. Wind energy and solar energy are both sustainable and nonpolluting sources and they are potentially to be alternative sources over traditional energy [2]. Utilizing renewable energy sources like the sun and wind, hybrid renewable energy system is becoming popular in economic and environmental ways.

However, the power of wind generator and solar panels is fluctuant and discontinuous, so it is highly necessary to add storage system to a single energy or hybrid energy supply to smooth the electrical power. Researches are being studied all over the world to solve the energy management problem of a DC microgrid with hybrid power generators, especially with renewable energy power generators. Kalantar and Mousavi [3] studied the dynamic behavior of stand-alone
hybrid power generation system with battery storage, using a supervisory controller based on programming to balance the energy within the system. Chen et al. [4] designed and implemented an energy management system (EMS) with fuzzy control for a DC microgrid system. Ko et al. [5] proposed a fuzzy controller and utilized sliding mode nonlinear control to keep a hybrid power system robust. Zhang et al. [6-8] investigated the problem of robust static output feedback (SOF) control and step tracking control problem for discrete-time nonlinear systems. Yin et al. [9-11] proposed a method of fault diagnosis scheme with parameters directly identified from the process data and compared the results of this data-driven method with process monitoring method.

Lithium ion batteries possess high energy density, relative high power density, long life span, and environmental friendliness and thus have been used in a wide range of areas [12]. Other types of batteries do not have the above advantages at the same time. For example, lead-acid batteries have much lower energy density than lithium ion batteries and often need to be in float charging state. The favorable characteristics
of lithium ion battery are very beneficial for hybrid energy system to improve its energy capacity. Yet the transient power frequency and fluctuating output voltage of the generating system may pose great pressure on the battery, which may reduce its lifetime span and worsen its performance. Besides, the relatively low power density makes it difficult for battery alone to meet the rapid changing of power generated by the nonstable renewable energy. Therefore, another auxiliary storage unit, along with battery storage system, may improve the system performance in prolonging the life span of lithium ion battery and providing more smoothed power flow for the loads.

Recently, super capacitors are being researched for many good qualities, such as considerably higher power densities than those of batteries and tremendous higher energy density than that of regular capacitors [13]. The high power density of SC is suitable for smoothing and the difference between the rapid changing generated power and the load demands caused by the power fluctuation. Besides, since their operation does not employ a chemical reaction, SCs are much more robust than batteries, which provide a long cycle life that is at least 500 times more than that of standard batteries [14]. For the listed reasons above, we can come to the conclusion that battery system with super capacity as auxiliary storage unit is beneficial for a hybrid power system to provide high quality power and meet the loads demands. In Kamel's research, ultracapacitor based energy storage system is developed to smooth the output power of wind turbine and enhance MG's performance in islanding mode [15]. Erdinc added ultracapacitor to hybrid vehicular power system to improve the efficiency and dynamic response of a vehicular system [16].

Due to the different characteristics of the renewable energy and the storage system, an energy management strategy is proposed in this paper to reduce the difference between the generated power and the power demanded by loads. Wavelet transform is applied to decompose the different power into smoothed component and high fluctuation component. The two components are suitable for battery and SC to compensate, respectively, according to their different features. What is more, to prevent the battery from overcharge and deep discharge, neural network algorithm is employed after the wavelet transform. The battery is mainly in charge of compensating most of the difference between the generated power and the demanded power, while the SC is mainly responsible for stabilizing the voltage on the DC bus of the whole system, which indirectly compensates the rest of the power difference.

This paper describes a standalone hybrid renewable energy power generation system with hybrid storage system consisting of lithium ion battery and SC, using a wavelet neural network based control strategy. The contents are organized as follows. Section 2 illustrates the modeling of the system components, respectively, including wind power generator, PV generator, lithium ion battery, and SC. Section 3 elaborates the wavelet neural network control strategy for the energy management system. The test results and discussion are given in Section 4.

## 2. System Description and Components Modeling

The proposed hybrid system consists of a wind turbine, a PV panel and a lithium ion battery, and SC based energy storage system. All the components are connected to a voltage uniformed DC bus with converters as is shown in Figure 1.
2.1. Wind Turbine Modeling. We assume a wind turbine driven by a permanent synchronous generator (PMSG) in this study. The rated output power and rated wind speed of the wind turbine are 600 W and $13 \mathrm{~m} / \mathrm{s}$. The starting wind speed is $3 \mathrm{~m} / \mathrm{s}$, and the maximum wind speed is $18 \mathrm{~m} / \mathrm{s}$. The maximum output power is 800 W . The model of wind turbine is built in Matlab/Simulink software. According to the aerodynamic theory, the output power of the wind turbine can be expressed as

$$
\begin{equation*}
P_{m}=0.5 \rho \pi R^{2} V_{w}^{2} C_{p}(\lambda, \beta) \tag{1}
\end{equation*}
$$

where $P_{m}$ is the output power extracted from wind turbine generator. $\rho, R$, and $V_{w}$ represent the air density, the blade radium, and the wind speed, respectively. $C_{p}$ is the power coefficient, which can be expressed as a function of tip speed ratio $\lambda$ and the blade pitch angle $\beta$ as follows:

$$
\begin{equation*}
C_{p}(\lambda, \beta)=0.73\left(\frac{151}{\lambda}-0.58 \beta-0.002 \beta^{2.14}-13.2\right) e^{-18.4 / \lambda} \tag{2}
\end{equation*}
$$

in which

$$
\begin{equation*}
\lambda=\frac{1}{\left(1 /(\lambda-0.02 \beta)-0.003 /\left(\beta^{3}+1\right)\right)} \tag{3}
\end{equation*}
$$

A $13 \mathrm{~m} / \mathrm{s}$ turbulent wind is generated from Bladed/FAST software and plotted in Matlab, shown in Figure 2. The output power of the wind turbine under MPPT control is shown in Figure 3.
2.2. PV Array Modeling. PV array is built by several PV cells connecting in series and parallel. The short-circuit current of each PV cell is calculated by the following equation:

$$
\begin{equation*}
I_{\mathrm{sc}}=I_{\mathrm{sc} 0}\left(\frac{G}{G_{0}}\right)^{\alpha} \tag{4}
\end{equation*}
$$

where $I_{\mathrm{sc}}$ and $I_{\mathrm{sc} 0}$ represent the short-circuit currents under standard and normal solar radiation $G$ and $G_{0}$ and $\alpha$ reflects all the nonlinear effects.

The open-circuit voltage of the PV cell is

$$
\begin{equation*}
V_{\mathrm{oc}}=\frac{V_{\mathrm{oc} 0}}{1+\beta \ln \left(G / G_{0}\right)} \cdot\left(\frac{T_{0}}{T}\right)^{\gamma} \tag{5}
\end{equation*}
$$

where $V_{\text {oc }}$ and $V_{\text {oco }}$ represent the open-circuit voltage under standard and normal solar radiation $G$ and $G_{0}$ and $T$ is the temperature of PV cell. $\beta$ is a specific technology-related coefficient of PV cell and $\gamma$ reflects the nonlinear effects of temperature.


Figure 1: Hybrid power system structure.


Figure 2: Turbulent wind of $13 \mathrm{~m} / \mathrm{s}$ from Bladed/FAST software.


Figure 3: The output power of the 600 W wind turbine with turbulent wind of $13 \mathrm{~m} / \mathrm{s}$.

The maximum output power from the PV array with $N_{s}$ PV cells connected in series and $N_{p}$ series groups in parallel can be written as

$$
\begin{align*}
P_{\max }= & N_{p} \cdot N_{s} \cdot \frac{V_{\mathrm{oc} 0} / n k T q-\ln \left(V_{\mathrm{oc}} / n k T q+0.72\right)}{1+V_{\mathrm{oc} 0} / n k T q}  \tag{6}\\
& \cdot\left(1-\frac{R_{s}}{V_{\mathrm{oc}} / I_{\mathrm{sc}}}\right) \cdot V_{\mathrm{oc}} \cdot I_{\mathrm{sc}} .
\end{align*}
$$



Figure 4: The maximum output power of PV array.

The maximum output power from PV array is shown in Figure 4.

Since the power of wind turbine and PV array are both obtained, the total generated power can be calculated by adding them together. In addition, we assume that the load demand is constant 1 kW . So the power difference between the generated power and the demanded power by the loads $\Delta P$ is

$$
\begin{equation*}
\Delta P=P_{\mathrm{WP}}+P_{\mathrm{PV}}-P_{L} \tag{7}
\end{equation*}
$$

in which $P_{\mathrm{WP}}, P_{\mathrm{PV}}$, and $P_{L}$ are wind turbine output power, PV array output power, and load demanded power, respectively. Variation of $\Delta P$ over time in this paper is shown in Figure 5.
2.3. Energy Storage System. As is mentioned before, the power generated from the renewable energy includes many sharp and transient variations, which makes it hard to meet the relatively smoothed load power demand. Therefore, it is necessary to compensate the gap between the generated power and the demanded power. In this scheme, the power that generated by the renewable energies and power that is demanded by the load are coordinated by the energy


Figure 5: Variation of $\Delta P$.
storage system which is composed of lithium ion battery bank and super capacitor (SC). Battery has high energy density, whereas it has relatively slow charging and discharging speed. On the other hand, super capacitor has the advantage of quick charge, large power density, and long cycle life [17]. SC in a hybrid energy storage system can quickly respond to the power smoothing instructions and better complete power smoothing tasks [18]. Based on the above characteristics, a modified coordinated control strategy, by which the total generated power can be smoothed and the loads demand can be met as well, is proposed and elaborated in the following chapters.
(1) Lithium Ion Battery Bank Modeling. The battery module from the Matlab/Simulation software is adopted in this paper. This model allows users to apply parameters based on battery type and nominal values. The battery bank in this paper consists of four batteries connected in parallel, each of which is of 24 V nominal voltage and 5 Ah rated capacity. That makes the total capacity of the battery bank 20 Ah . The parameters of the lithium ion battery are listed in Table 1.
(2) Super Capacitor Modeling. A classical equivalent model for SC is shown in Figure 6, which consists of a capacitance $(C)$, an equivalent series resistance $\left(R_{S}\right)$ representing the charging and discharging resistance, and an equivalent parallel resistance $\left(R_{P}\right)$ representing the self-discharging losses [19]. Instead of using the common RC equivalent circuit, a modified electrical equivalent circuit for super capacity, as is shown in Figure 7, is applied in this paper. The RC branch in the modified equivalent circuit composed of $R_{S}$ and $C_{1}$ is called the "fast branch" and is used to represent the immediate behavior of the SC in the time range of seconds. The RC branch comprising $R_{2}$ and $C_{2}$ is called the "slow branch" and this RC branch presents the internal energy distribution at the end of charge or discharge [20].

## 3. System Control Strategy

In this section, a wavelet transform and neural network based control strategy is introduced to manage the system energy. The advantage of wavelet analysis, as opposed to conventional techniques, is that wavelet transform decomposes a signal into a series of short duration waves or local basis functions (wavelets) on the time axis which allows the analysis of


Figure 6: Classical equivalent model for SC.


Figure 7: Modified equivalent model for SC.

Table 1: Lithium ion battery bank parameters.

| Parameter | Value |
| :--- | :---: |
| Nominal voltage | 24 V |
| Rated capacity | $20 \mathrm{Ah}(4 * 5 \mathrm{Ah})$ |
| Maximum capacity | $20 \mathrm{Ah}(4 * 5 \mathrm{Ah})$ |
| Fully charged voltage | 27.9357 V |
| Nominal discharge current | 8.6957 A |
| Internal resistance | $12 \mathrm{~m} \Omega(48 / 4 \mathrm{~m} \Omega)$ |

local phenomena in signals consisting of many transients [21]. In this case, we apply a three-level Haar wavelet. The original signal for wavelet transform is $\Delta P$, the difference between the total generated power $P_{\mathrm{WP}}+P_{\mathrm{PV}}$ and the load demand $P_{L}$. Then $\Delta P$ is decomposed into approximation component $\Delta P_{a}$ and detailed component $\Delta P_{d}$ by wavelet transform. According to the different respond speed and power density characteristics of the two types of storage devices mentioned above, the majority of $\Delta P_{a}$, which is the smoothed component of the total difference, is convenient to be met by battery, while the detailed part $\Delta P_{d}$, which contains a lot of high frequency components, is suitable for super capacity to compensate. The decomposition is illustrated in Figure 8.

However, if we assign $\Delta P_{a}$ to be the exact reference power of battery directly, it is highly possible that the batteries would reach out their acceptable SOC limit, which would cause some certain damages to the batteries and reduce its lifetime. Therefore, neural network is introduced to maintain the SOC of batteries within a reasonable range, so that they can function in good condition. In this paper, adaptive linear (ADALINE) neural network is applied to obtain the reference power of battery storage system. Figure 9 shows the general neural network model. $p_{1}, p_{2}$ are the inputs of network controller; $W_{1}$ and $W_{2}$ are the corresponding


Figure 8: Decomposition of power difference $\Delta P$ using Haar wavelet transform.


Figure 9: Model of a simple ADALINE neural network.
weight of the two inputs parameters; $b$ represents bias and $n$ represents the net input; $a$ is the output of the network controller. In this case, the inputs of the neural network are $\Delta P_{a}$-the approximation component of the total different $\Delta P$ after wavelet transformation, and the SOC of the battery. By utilizing the input parameters, the neural network controller determines the reference power for battery. The first half data of $\Delta P_{a}$ is used to train the perception. The target value of battery SOC is set as 0.7 . If the SOC of battery is around the desired value, then the reference power for battery is $\Delta P_{a}$. However, if the SOC is more/less than the desired value, the reference power is more/less than $\Delta P_{a}$, in order
to decrease/increase the SOC to the target. This way, the battery would be protected from being overcharged or deep discharged.

The whole configuration of the wavelet neural network controller is illustrated in Figure 10. This scheme is composed of a battery bank and a super capacity using two bidirectional DC/DC converters, respectively, for power tracking and voltage regulation. The reference power for the battery is obtained by the wavelet neural network elaborated above. A PI controller is then followed to track the battery's actual power and the system response can be achieved by generating PWM switching signals to DC/DC converters. Besides, the DC bus voltage is regulated by controlling the bidirectional DC/DC converter coupled with SC.

## 4. Simulation Results

We set the learning step of the neural network 0.008 and the mean square error 0.005 . The network is trained using the first half of operation data according to the principles mentioned in Section 3. The output power of battery is demonstrated in Figure 11. It is obvious that the variation of battery output power is similar to $\Delta P_{a}$ the approximation component of power difference after wavelet transform. As


Figure 10: Wavelet and neural network controller based energy management system.


Figure 11: The output power of the lithium ion battery.


Figure 12: The variation of SOC of the lithium ion battery.
we can see, the changes of the battery power possess less sharp transition part than the original power difference $\Delta P$ shown in Figure 5. This protects the battery from being damaged by extremely fast charging and discharging operation.

The variation of battery SOC is demonstrated in Figure 12. The SOC is sustained near 0.7 as we designed in the neural network controller.


Figure 13: The output power of the super capacitor.


Figure 14: The variation of DC bus voltage.

Figure 13 shows the output power of super capacitor, which resembles the detailed part of the power difference after wavelet transform.

The variation of the DC bus voltage is shown in Figure 14. The voltage is sustained within a reasonable limit by generating proper PWM signals to super capacitor, which is important for the loads connected on the DC bus. As is illustrated in the figure, the voltage of the DC bus maintained at the point of 48 V after a short fluctuation at the beginning.

The simulation results show that the proposed method effectively keeps the microgrid operating under stable state
with power fluctuation injected in it by making full use of two different storage units. The salient features of the proposed method include: (1) effective in dealing with modeling uncertainties; (2) structurally simple and computationally inexpensive; and (3) the design parameters can be readily determined, which makes it much easier than tuning conventional PID controller.

## 5. Conclusion

This paper proposes a wavelet transform and neural network based energy management system for hybrid power system. The hybrid power system consists of wind power subsystem, solar power subsystem, and an energy storage system. The wind turbine and PV array are all controlled under MPPT to obtain maximum electrical power. The energy storage system includes lithium ion battery bank and super capacitor which are controlled under the proposed energy management system. In the proposed control strategy, wavelet is in charge of decomposing and then reconfiguring the power difference between generated power and consumed power by loads. The approximation part is compensated by battery bank. In consideration of sustaining its SOC within an acceptable limit, neural network controller is introduced. Then the voltage of system DC bus is maintained by rapidly charging and discharging the super capacitor.

The simulation results demonstrate that the proposed strategy is capable of compensating the variation power difference caused by the renewable energy and the loads, as well as maintaining the DC voltage stable. Furthermore, the SOC of batteries is within the recommended range, thus protecting them from being damaged by overcharge and deep discharge. Since the modeling of the system is complex and difficult, future work will be more focused on data-driven controller design for microgrid.

## Acknowledgments

This work is supported in part by the National Basic Research Program of China (973 Program no. 2012CB215200) and in part by the National Natural Science Foundation of China (no. 51205046).

## References

[1] R. Pecen and A. Nayir, "Design and implementation of a 12 kW wind-solar distributed power and instrumentation system as an educational testbed for Electrical Engineering Technology students," in Proceedings of the International Symposium on Modern Electric Power Systems (MEPS '10), pp. 1-6, September 2010.
[2] Y. Xie, J. Sun, C. Mi, and J. S. Freudenberg, "Analysis and modeling of a DC hybrid power system testbed for power management strategy development," in Proceedings of the 5th IEEE Vehicle Power and Propulsion Conference (VPPC '09), pp. 926-933, September 2009.
[3] M. Kalantar and S. M. Mousavi, "Dynamic behavior of a stand-alone hybrid power generation system of wind turbine,
microturbine, solar array and battery storage," Applied Energy, vol. 87, no. 10, pp. 3051-3064, 2010.
[4] Y. K. Chen, Y. C. Wu, C. C. Song, and Y. S. Chen, "Design and implementation of energy management system with fuzzy control for DC microgrid systems," IEEE Transactions on Power Electronics, vol. 28, no. 4, pp. 1563-1570, 2013.
[5] H.-S. Ko, G.-G. Yoon, W.-P. Hong, and J. Jatskevich, "Control of hybrid wind power generation system with dump load using advanced fuzzy-robust controller,' in Proceedings of the International Conference on Electrical Machines and Systems (ICEMS '07), pp. 219-226, October 2007.
[6] H. Zhang, Y. Shi, and A. Saadat Mehr, "Robust static output feedback control and remote PID design for networked motor systems," IEEE Transactions on Industrial Electronics, vol. 58, no. 12, pp. 5396-5405, 2011.
[7] H. Zhang, Y. Shi, and A. Saadat Mehr, "Robust non-fragile dynamic vibration absorbers with uncertain factors," Journal of Sound and Vibration, vol. 330, no. 4, pp. 559-566, 2011.
[8] H. Zhang, Y. Shi, and M. Liu, " $H_{\infty}$ step tracking control for networked discrete-time nonlinear systems with integral and predictive actions," IEEE Transactions on Industrial Informatics, vol. 9, no. 1, pp. 337-345, 2013.
[9] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," Journal of Process Control, vol. 22, no. 9, pp. 1567-1581, 2012.
[10] S. Yin, X. Yang, and H. R. Karimi, "Data-driven adaptive observer for fault diagnosis," Mathematical Problems in Engineering, vol. 2012, Article ID 832836, 21 pages, 2012.
[11] S. Yin, S. X. Ding, A. H. Abandan Sari, and H. Hao, "Data-driven monitoring for stochastic systems and its application on batch process," International Journal of Systems Science, vol. 44, no. 7, pp. 1366-1376, 2013.
[12] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," Journal of Power Sources, vol. 226, pp. 272-288, 2013.
[13] M. Akbari, S. M. M. Tafreshi, and M. A. Golkar, "Voltage control of a hybrid ac/dc microgrid in stand-alone operation mode," in Proceedings of the IEEE PES International Conference on Innovative Smart Grid Technologies-India (ISGT India '11), pp. 363-367, December 2011.
[14] Y. Ates, O. Erdinc, M. Uzunoglu, and B. Vural, "Energy management of an FC/UC hybrid vehicular power system using a combined neural network-wavelet transform based strategy," International Journal of Hydrogen Energy, vol. 35, no. 2, pp. 774783, 2010.
[15] R. M. Kamel, A. Chaouachi, and K. Nagasaka, "Wind power smoothing using fuzzy logic pitch controller and energy capacitor system for improvement Micro-Grid performance in islanding mode," Energy, vol. 35, no. 5, pp. 2119-2129, 2010.
[16] O. Erdinc, B. Vural, M. Uzunoglu, and Y. Ates, "Modeling and analysis of an FC/UC hybrid vehicular power system using a wavelet-fuzzy logic based load sharing and control algorithm," International Journal of Hydrogen Energy, vol. 34, no. 12, pp. 5223-5233, 2009.
[17] R. Lu, C. Zhu, L. Tian, and Q. Wang, "Super-capacitor stacks management system with dynamic equalization techniques," IEEE Transactions on Magnetics, vol. 43, no. 1, pp. 254-258, 2007.
[18] X. Han, F. Chen, X. Cui, Y. Li, and X. Li, "A power smoothing control strategy and optimized allocation of battery capacity
based on hybrid storage energy technology," Energies, vol. 5, no. 5, pp. 1593-1612, 2012.
[19] M. Uzunoglu, O. C. Onar, and M. S. Alam, "Modeling, control and simulation of a PV/FC/UC based hybrid power generation system for stand-alone applications," Renewable Energy, vol. 34, no. 3, pp. 509-520, 2009.
[20] O. Erdinc, B. Vural, M. Uzunoglu, and Y. Ates, "Modeling and analysis of an FC/UC hybrid vehicular power system using a wavelet-fuzzy logic based load sharing and control algorithm," International Journal of Hydrogen Energy, vol. 34, no. 12, pp. 5223-5233, 2009.
[21] S. P. Valsan and K. S. Swarup, "Wavelet based transformer protection using high frequency power directional signals," Electric Power Systems Research, vol. 78, no. 4, pp. 547-558, 2008.

# A Hybrid Multiobjective Differential Evolution Algorithm and Its Application to the Optimization of Grinding and Classification 

Yalin Wang, ${ }^{1}$ Xiaofang Chen, ${ }^{1}$ Weihua Gui, ${ }^{1}$ Chunhua Yang, ${ }^{1}$ Lou Caccetta, ${ }^{2}$ and Honglei $\mathbf{X u}^{2}$<br>${ }^{1}$ School of Information Science and Engineering, Central South University, Changsha 410083, China<br>${ }^{2}$ Department of Mathematics \& Statistics, Curtin University, Perth, WA 6845, Australia<br>Correspondence should be addressed to Xiaofang Chen; xiaofangchen@csu.edu.cn

Received 19 July 2013; Accepted 5 September 2013
Academic Editor: Dewei Li
Copyright © 2013 Yalin Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The grinding-classification is the prerequisite process for full recovery of the nonrenewable minerals with both production quality and quantity objectives concerned. Its natural formulation is a constrained multiobjective optimization problem of complex expression since the process is composed of one grinding machine and two classification machines. In this paper, a hybrid differential evolution (DE) algorithm with multi-population is proposed. Some infeasible solutions with better performance are allowed to be saved, and they participate randomly in the evolution. In order to exploit the meaningful infeasible solutions, a functionally partitioned multi-population mechanism is designed to find an optimal solution from all possible directions. Meanwhile, a simplex method for local search is inserted into the evolution process to enhance the searching strategy in the optimization process. Simulation results from the test of some benchmark problems indicate that the proposed algorithm tends to converge quickly and effectively to the Pareto frontier with better distribution. Finally, the proposed algorithm is applied to solve a multiobjective optimization model of a grinding and classification process. Based on the technique for order performance by similarity to ideal solution (TOPSIS), the satisfactory solution is obtained by using a decision-making method for multiple attributes.


## 1. Introduction

Grinding-classification is an important prerequisite process for most mineral processing plants. The grinding process reduces the particle size of raw ores and is usually followed by classification to separate them into different sizes. Grindingclassification operation is required to produce pulp with suitable concentration and fineness for flotation. The pulp quality will directly influence the subsequent flotation efficiency and recovery of valuable metals from tailings. In order to improve economic efficiency and energy consumption, the process optimization objectives include product quality and output yields. Under certain mineral source conditions, the objectives are decided by a series of operation variables such as the solid flow of feed ore to ball mill, the steel ball filling rate, and the flow rates of water added to the first and the
second classifier recycles. To solve the optimization model of products' output and quality in the grinding-classification process is of great significance to improve the technical and economic specifications, and it has been a continuous endeavor of the scientists and engineers [1-3].

Grinding-classification is an energy-intensive process influenced by many interacting factors with mutual restraints. The goals of grinding-classification optimization problem are decided by multiple constrained input control variables of nonlinear relationships. So, the optimization model of grinding-classification operation is a complex constrained multiobjective optimization problem (CMOP). Generally, constrained multiobjective problems are so difficult to be solved that the constraint handling techniques and multiobjective optimization methods need to be combined for optimization.

Multiobjective optimization problems (MOPs), in the case of traditional optimization methods, are often handled by aggregating multiple objectives into a single scalar objective through weighting factors. MOPs have a set of equally good (nondominating) solutions instead of a single one, called a Pareto optimum which was introduced by Edgeworth in 1881 [4] and later generalized by Pareto in 1896 [5]. The practical MOPs are often implicated in series of equations, functions, or procedures with complicated constraints. Therefore, the evolutionary algorithms are attractive approaches for low requirements on mathematical expression [6]. Since the mid 1980s, there has been a growing interest in solving MOPs using evolutionary approaches [7-10]. One of the most successful evolutionary algorithms for the optimization of continuous space functions is the differential evolution (DE) [11]. DE is simple and efficiently converges to the global optimum in most cases [12, 13]. Its efficiency has been proven [14] in many application fields such as pattern recognition [15] and mechanical engineering [16].

There have been many improvements for DE to solve MOPs. Abbass [17] firstly provided a Pareto DE (PDE) algorithm for MOPs in which DE was employed to create new solutions, and only the nondominated solutions were kept as the basis for the next generation. Madavan [18] developed a Pareto differential evolution approach (PDEA) in which new solutions were created by DE and kept in an auxiliary population. Xue et al. [19] introduced multiobjective differential evolution (MODE) and used Pareto-based ranking assignment and crowding distance metric, but in a different manner from PDEA. Robic and Filipi [20], also adopting Paretobased ranking assignment and crowding distance metric, developed a DE for multiobjective optimization (DEMO) with a different population update strategy and achieved good results. Huang et al. [21] extended the self-adaptive DE (SADE) to solve MOPs by a so called multiobjective selfadaptive DE (MOSADE). They further extended MOSADE by using objectivewise learning strategies [22]. Adeyemo and Otieno [23] provided multiobjective differential evolution algorithm (MDEA). In MDEA, a new solution was generated by DE variant and compared with target solution. If it dominates the target solution, then it was added to the new population; otherwise, a target solution was added.

On the other hand, single-objective constrained optimization problems have been studied intensively in the past years [24-28]. Different constraint handling techniques have been proposed to solve constrained optimization problems. Michalewicz and Schoenauer [29] divided constraints handling methods used in evolutionary algorithms into four categories: preserving feasibility of solutions, penalty functions, separating the feasible and infeasible solutions, and hybrid methods. The differences among these methods are how to deal with the infeasible individuals throughout the search phases. Currently, the penalty function method is most widely used, and this algorithm strongly depends on the choice of the penalty parameter.

Although the multiobjective optimization and the constraint handling problem have received lots of contribution, respectively, the CMOPs are still difficult to be solved in practice. Coello and Christiansen [30] proposed a simple
approach to solve CMOPs by ignoring any solution that violates any of the assigned constraints. Deb et al. [8] proposed a constrained multiobjective algorithm based on the concept of constrained domination, which is also known as superiority of the feasible solution. Woldesenbet et al. [31] introduced a constraint handling technique based on adaptive penalty functions and distance measures by extending the corresponding version for the single-objective constrained optimization.

In the MOP of grinding and classification process, the definitions of Pareto solutions, Pareto frontier, and Pareto dominance are in consistency with the classic definitions. Clearly, the Pareto frontier is a mapping of the Pareto-optimal solutions to the objective space. In the minimization sense, general constrained MOPs can be formulated as follows

$$
\begin{array}{ll}
\min & F(X)=\min \left[f_{1}(X), f_{2}(X), \ldots, f_{r}(X)\right], \\
\text { s.t. } & g_{i}(X) \leq 0 \quad(i=1,2, \ldots, p)  \tag{1}\\
& h_{j}(X)=0 \quad(j=p+1, \ldots, q) \\
& x_{k} \in\left[x_{k \min }, x_{k \max }\right] \quad(k=1,2, \ldots, n)
\end{array}
$$

where $F(X)$ is the objective vector, $X=\left(x_{1}, \ldots, x_{n}\right) \in \mathbb{R}^{n}$ is a parameter vector, $g_{i}(X)$ is the $i$ th inequality constraint, and $h_{j}(X)$ is the $j$ th equality constraint. $x_{k \text { min }}$ and $x_{k \text { max }}$ are, respectively, the lower and upper bounds of the decision variable $x_{k}$.

In this paper, based on the specific industrial background of continuous bauxite grinding-classification operation, a new hybrid DE algorithm is proposed to solve complex constrained multiobjective optimization problems. Firstly, a hybrid DE algorithm for MOPs with simplex method (SMDEMO) is designed to overcome the problems of global performance degradation and being trapped in local optimum. Then, for the MOPs with complicated constraints, the proposed algorithm is formed by combining SM-DEMO and functional partitioned multi-population. In this method, the construction of penalty functions is not required, and the meaningful infeasible solutions are fully utilized.

The remainder of the paper is structured as follows. Section 2 describes the SM-DEMO algorithm for unconstrained cases. The proposed algorithm of multipopulation for constrained MOPs is given in Section 3 with verification of performance by benchmark testing results. Section 4 describes the model of products' output and quality in the grinding-classification process in detail and the application of the proposed algorithm in the optimization model. Finally, the conclusions based on the present study are drawn in Section 5.

## 2. SM-DEMO Algorithm for Unconstrained MOPs

In order to efficiently solve multiobjective optimization problem and find the approximately complete and nearoptimal Pareto frontier, we proposed a hybrid DE algorithm for unconstrained multiobjective optimization with simplex method.

The differential evolution, with initialization, crossover, and selection as in usual genetic algorithms, uses a perturbation of two members as the mutation operator to produce
a new individual. The mutation operator of the DE algorithm is described as follows.

Considering each target individual $x_{i}^{G}$, in the $G$ th generation of size $N p$, a mutant individual $\widehat{x}_{i}^{G+1}$ is defined by

$$
\begin{equation*}
\widehat{x}_{i}^{G+1}=x_{r 3}^{G}+F\left(x_{r 1}^{G}-x_{r 2}^{G}\right), \tag{2}
\end{equation*}
$$

where indexes $r_{1}, r_{2}$, and $r_{3}$ represent mutually different integers that are different from $i$ and that are randomly generated over $[1, N p]$, and $F$ is the scaling factor.

The simplex method, proposed by Spendley, Hext, and Himsworth and later refined by Nelder and Mead (NM) [32], is a derivative-free line-search method that is particularly designed for traditional unconstrained minimization scenarios. Clearly, NM method can be deemed as a direct linesearch method of the steepest descent kind. The ingredients of the replacement process consist of four basic operations: reflection, expansion, contraction, and shrinkage. Through these operations, the simplex can improve itself and approximate to a local optimum point sequentially. Furthermore, the simplex can vary its shape, size, and orientation to adapt itself to the local contour of the objective function.
2.1. Main Strategy of SM-DEMO. The SM-DEMO algorithm is improved by the following three points compared with DE .
2.1.1. Modified Selection Operation. After traditional DE evolution, the individual $u_{i j}^{G+1}$ may violate the boundary constraints $x_{i j}^{\text {max }}$ and $x_{i j}^{\min } \cdot u_{i j}^{G+1}$ is replaced by new individual $w_{i j}^{G+1}$ being adjusted as follows:
$w_{i j}^{G+1}$

$$
= \begin{cases}x_{i j}^{\max }+\operatorname{rand}() *\left(x_{i j}^{\max }-u_{i j}^{G+1}\right), & \text { if }\left(u_{i j}^{G+1}>x_{i j}^{\max }\right),  \tag{3}\\ x_{i j}^{\min }+\operatorname{rand}() *\left(x_{i j}^{\min }-u_{i j}^{G+1}\right), & \text { if }\left(u_{i j}^{G+1}<x_{i j}^{\min }\right), \\ u_{i j}^{G+1}, & \text { otherwise. }\end{cases}
$$

The new population is combined with the existing parent population to form a new set $M g$ of bigger size than $N p$. A nondominated ranking of $M g$ is performed, and the $N p$ best individuals are selected. This approach allows a global nondomination checking between both the parent and the new generation rather than only in the new generation as is done in other approaches, whereas it requires additional computational cost in sorting the combined.
2.1.2. Nondominated Ranking Based on Euclidean Distance. The solutions within each nondominated frontier that reside in the less crowded region in the frontier are assigned a higher rank, as the NSGA-II algorithm [8] developed by Deb et al. indicated. The crowding distance of the $i$ th solution in its frontier (marked with solid circles) is the average side length of the cuboids (shown with a dashed box in Figure 1(a)). The crowding-distance computation requires sorting the population according to each objective function value in ascending order of magnitude. As shown in Figure 1, $A$ and $C$ are two solutions near $B$ in the same rank, and $\sigma_{0}(B)$ is the
crowding distance of the $B$ th solution, traditionally calculated as follows:

$$
\begin{equation*}
\sigma_{0}(B)=\sum_{j=1}^{n}\left|f_{j}(A)-f_{j}(C)\right|, \tag{4}
\end{equation*}
$$

where $f_{j}(A), f_{j}(C)$ are the objective vectors. For each objective function, the boundary solutions (solutions with the smallest and the largest function values) are assigned an infinite distance value.

A crowding-distance metric is used to estimate the density of solutions surrounding a particular solution in the population and is obtained from the average distance of the two solutions on either side of the solution along each of the objectives. As shown in Figure 1, $A, B, C$ are the individuals of the generation on the same frontier, and we easily know that the density in Figure 1(a) is better than that in Figure 1(b). If we use (4) to calculate the crowding distance of $C$, we only know that in Figure 1(a) it is better than in Figure 1(b); the crowding distance of $C$ in Figures 1(a) and 1(c) is equal, which is against the knowledge.

To distinguish the mentioned situations, we propose an improved crowding-distance metric based on Euclidean distance. $M$ is the center point of the line $A B, f_{j}$ is the $j$ th objective vector, and the crowding distance $\sigma(B)$ is defined as follows:

$$
\begin{align*}
\sigma(B)= & |A C|-|B M| \\
= & \sqrt{\sum_{j=1}^{n}\left[f_{j}(A)-f_{j}(C)\right]^{2}}  \tag{5}\\
& -\sqrt{\sum_{j=1}^{n}\left\{f_{j}(B)-\frac{\left[f_{j}(A)+f_{j}(C)\right]}{2}\right\}} .
\end{align*}
$$

The crowded-comparison operator guides the selection process at the various stages of the algorithm toward a uniformly distributed Pareto-optimal frontier. To carry out the comparison, we assume that every individual in the population has two attributes: (1) nondomination rank $i_{\text {rank }}$ and (2) crowding distance $i_{\text {distance. }}$. Then, a partial order is defined as $i<j$. If $i<j$, that is, between two solutions with different nondomination ranks, we prefer the solution with the lower (better) rank, namely, $i_{\text {rank }} \prec j_{\text {rank }}$. Otherwise, if both solutions belong to the same frontier, that is, $i_{\text {rank }}=$ $j_{\text {rank }}$, then we prefer the solution that is located in a lesser crowded region, that is, $i_{\text {distance }} \succ j_{\text {distance }}$.
2.1.3. Simplex Method for Local Search. The simplex method for local search is mixed in the evolution process to enhance the searching strategy in the optimization process. The goal of integrating NM simplex method [32] and DE is to enrich the population diversity and avoid being trapped in local minimum. We apply NM simplex method operator to the present population when the number of iterations is greater than a particular value (like $G_{\max } / 2$ ). The individuals that achieved the single extreme value in each objective function are marked as the initial vertex points of simplex method. The


Figure 1: Crowding-distance diagram.
present population is updated according to simplex method until the terminal conditions are satisfied.

The computation steps of the algorithm are included in Section 3.2.
2.2. Evaluation Criteria. Unlike the single-objective optimization, it is more complicated for solution quality evaluation in the case of multiobjective optimization. Many of the suggested methods can be summarized in two types. One is to evaluate the convergence degree by computing the proximity between the solution frontier and the actual Pareto frontier. The other is to evaluate the distribution degree of the solutions in objective space by computing the distances among the individuals. Here, we choose both methods to evaluate the performance of the SM-DEMO algorithm.
(1) Convergence Evaluation. Deb et al. [8] proposed this method in 2002. It is described as follows:

$$
\begin{equation*}
\gamma=\frac{1}{\mathrm{Q}}\left(\sum_{i=1}^{\mathrm{Q}} \min \left\|P^{*}-P_{\mathrm{FT}}\right\|\right) \tag{6}
\end{equation*}
$$

where $\gamma$ is the extent of convergence to a known of Paretooptimal set, $P^{*}$ is the obtained nondomination Pareto frontier, $P_{\mathrm{FT}}$ is the real nondomination Pareto frontier, $\left\|P^{*}-P_{\mathrm{FT}}\right\|$ is the Euclidean distance of $P^{*}$ with $P_{\mathrm{FT}}$, and $Q$ is the number of obtained solutions.
(2) Distribution Degree Evaluation. The nonuniformity in the distribution is measured by SP as follows:

$$
\begin{equation*}
\mathrm{SP}=\sqrt{\frac{1}{(\mathrm{Q}-1)} \sum_{i=1}^{\mathrm{Q}}\left(\bar{d}-d_{i}\right)^{2}} \tag{7}
\end{equation*}
$$

where $d_{i}$ is the Euclidean distance among consecutive solutions in the obtained nondominated set of solutions and parameter $\bar{d}$ is the average distance.
2.3. Experimental Studies. Four well-known benchmark test functions [33] are used here to compare the performance of SM-DEMO with NSGA-II, DEMO/Parent. These four problems are called ZDT2, ZDT3, ZDT4, and ZDT6; each has two objective functions. We describe them in Table 1.

Table 1: Test problems.

| Test problems | $\begin{gathered} \text { Objective functions } \\ \min F(X)=\min \left[f_{1}(X), f_{2}(X)\right] \end{gathered}$ | Range of variable |
| :---: | :---: | :---: |
| ZDT2 | $\begin{gathered} f_{1}(X)=x_{1}, f_{2}(X)=g(X)\left(1-\left(\frac{f_{1}}{g(X)}\right)^{2}\right), \\ g(X)=1+9 \sum_{i=2}^{n} \frac{x_{i}}{n-1} \end{gathered}$ | $\begin{gathered} n=30 \\ 0 \leq x_{i} \leq 1 \end{gathered}$ |
| ZDT3 | $\begin{gathered} f_{1}(X)=x_{1}, f_{2}(X)=g\left(1-\sqrt{\left(\frac{f_{1}}{g}\right)}-\left(\frac{f_{1}}{g}\right) \sin \left(10 \pi f_{1}\right)\right), \\ g(X)=1+9 \sum_{i=2}^{n} \frac{x_{i}}{n-1} \end{gathered}$ | $\begin{gathered} n=30 \\ 0 \leq x_{i} \leq 1 \end{gathered}$ |
| ZDT4 | $\begin{gathered} f_{1}(X)=x_{1}, f_{2}(X)=g\left(1-\sqrt{\left(\frac{f_{1}}{g}\right)}\right) \\ g(X)=1+10(n-1)+\sum_{i=2}^{n}\left(x_{i}^{2}-10 \cos \left(4 \pi x_{i}\right)\right) \end{gathered}$ | $\begin{gathered} n=10 \\ 0 \leq x_{1} \leq 1 \\ -5 \leq x_{i} \leq 5, i=2, \ldots, n \end{gathered}$ |
| ZDT6 | $\begin{gathered} f_{1}(X)=1-\exp \left(-4 x_{1}\right) \sin ^{6}\left(6 \pi x_{1}\right), f_{2}(X)=g\left(1-\left(\frac{f_{1}}{g}\right)^{2}\right), \\ g(X)=1+9 \sum_{i=2}^{n} \frac{x_{i}}{(n-1)^{0.25}} \end{gathered}$ | $\begin{gathered} n=10 \\ 0 \leq x_{i} \leq 1 \end{gathered}$ |

Table 2: The performance results of the each algorithm on the test function.

| Test function | Algorithm | $\gamma$ | SP |
| :--- | :---: | :---: | :---: |
|  | DEMO/Parent | $0.005120 \pm 0.000312$ | $0.000630 \pm 0.000010$ |
| ZDT2 | NSGA-2 | $0.007120 \pm 0.000413$ | $0.000540 \pm 0.000940$ |
|  | SM-DEMO | $0.004013 \pm 0.000230$ | $0.000423 \pm 0.000011$ |
| ZDT3 | DEMO/Parent | $0.009704 \pm 0.000027$ | $0.007512 \pm 0.000165$ |
|  | NSGA-2 | $0.014067 \pm 0.000059$ | $0.006540 \pm 0.000124$ |
|  | SM-DEMO | $0.004704 \pm 0.000003$ | $0.004450 \pm 0.000153$ |
|  | DEMO/Parent | $2.009704 \pm 0.901164$ | $0.011031 \pm 0.001104$ |
| ZDT4 | NSGA-2 | $3.144067 \pm 2.100740$ | $0.010122 \pm 0.000072$ |
|  | SM-DEMO | $0.874001 \pm 0.014323$ | $0.008721 \pm 0.000159$ |
|  | DEMO/Parent | $0.649704 \pm 0.004912$ | $0.104520 \pm 0.015486$ |
| ZDT6 | NSGA-2 | $1.014067 \pm 0.010421$ | $0.007942 \pm 0.000105$ |
|  | SM-DEMO | $0.007750 \pm 0.000083$ | $0.002014 \pm 0.000117$ |

The simulation is carried out under the environment of Intel Pentium 4, CPU 3.06 GHz , 512 MB memory, Windows XP Professional, Matlab7.1. Initialization parameters are set as follows: population size $N p=100$, scaling factor $F=0.8$, cross rate $C_{R}=0.6$, maximum evolution generation $G_{\max }=$ 250 , and number of SM evolution iterations $G_{S M}=100$.

All of the three algorithms are real coded, with equal population size and equal maximum evolution generation. Each algorithm independently runs 20 times for each test function. Because we cannot get the real Pareto-optimal set, we will take 60 Pareto-optimal solutions obtained by the three algorithms as a true Pareto-optimal solution set.

We evaluated the algorithms based on the two performance indexes $\gamma$ and SP. Table 2 shows the mean and variance of $\gamma$ and SP using three algorithms: SM-DEMO, NSGAII, and DEMO/Parent. We can learn from Table 2, for the ZDT2 function, that all of the three algorithms have a good
performance, while the SM-DEMO is slightly better than the other two algorithms. In terms of convergences, for ZDT3, ZDT4 and ZDT6, which are more complex than ZDT2, SMDEMO is significantly better than DEMO/Parent and NSGAII.

Figure 2 shows a random running of SM-DEMO algorithm. It is clear that SM-DEMO algorithm can produce a good approximation and a uniform distribution.

## 3. Proposed Hybrid Algorithm for CMOP

The space of constrained multiobjective optimization problem can be divided into the feasible solution space and the infeasible solution space, as shown in Figure 3, where $S$ is the search space, $\Omega$ is the feasible solution space, and $Z$ is the infeasible solution space. $x_{i}(i=1,2,3,4)$ is the feasible solution, and $y_{i}(i=1,2,3,4)$ is the infeasible solution.


Figure 2: SM-DEMO simulation curve.

Assume that $x^{*}$ is the global optimal solution and $y_{1}$ is the closest one to $x^{*}$. If the infeasible population $y_{1}$ is not excluded by the evolution algorithm, it is permitted to explore boundary regions from new directions, where the optimum is likely to be found.

### 3.1. General Idea of the Proposed Algorithm. Researchers

 have gradually realized the merit of infeasible solutions in searching for the global optimum in the feasible region. Some infeasible solutions with better performance are allowed to be saved. Farmani et al. [34] formulated a method to ensure that infeasible solutions with a slight violation become feasible in evolution. Based on the constraints processing approach of multiobjective optimization problems, the proposed hybrid DE algorithm avoids constructing penalty function and deleting meaningful infeasible solutions directly.Here, the proposed algorithm will produce multiple groups of functional partitions, which include an evolutionary population Pg of size $N p$, an intermediate population $M g$ to save feasible individuals, an intermediate population $S g$ to save infeasible individuals, a population $P f$ to save the optimal feasible solution found in the search process and a population Pc to save the optimal infeasible solution. The relationship of multi-population is shown in Figure 4.

With the description of (1), equality constraints are always transformed into inequality constraints as $\left|h_{j}(X)\right|-\delta \leq 0$, where $j=p+1, \ldots, q$ and $\delta$ is a positive tolerance value. To evaluate the infeasible solution, the degree of constraint violation of individual $X$ on the $j$ th constraint is calculated as follows:

$$
V_{i}(X)= \begin{cases}\max \left\{0, g_{i}(X)\right\} & (i=1,2, \ldots, p),  \tag{8}\\ \max \left\{0,\left|h_{j}(X)\right|-\delta\right\} & (j=p+1, \ldots, q)\end{cases}
$$



Figure 3: Distribution diagram of search space.


Figure 4: The relationship diagram of multipopulation demonstrating.

The final constraint violation of each individual in the population can be obtained by calculating the mean of the normalized constraint violations.

In order to take advantage of the infeasible solutions with better performance, we proposed the following adaptive differential mutation operator to generate individual variation learning from the mutation operators in $\mathrm{DE} /$ rand-tobest/1/bin, according to rules defined by Price et al. [11]. Considering each individual vector $x_{i}^{G}$, a mutant individual $\widehat{x}_{i}^{G+1}$ is defined by
$\widehat{x}_{i}^{G+1}$
$= \begin{cases}x_{i}^{G}+F_{1} \cdot\left(x_{f 1}^{G}-x_{r 1}^{G}\right)+F_{2} \cdot\left(x_{f 2}^{G}-x_{r 2}^{G}\right), & R_{C} \geq \operatorname{rand}(), \\ x_{i}^{G}+F_{1} \cdot\left(y_{i}^{G}-x_{r 1}^{G}\right)+F_{2} \cdot\left(x_{f 1}^{G}-x_{r 2}^{G}\right), & R_{C}<\operatorname{rand}(),\end{cases}$
where $r_{1}$ and $r_{2}$ represent different integers and also different from $i$, randomly generated over $[1, N p] ; F$ is the scaling factor; $x_{f i}^{G}(i=1,2, \ldots, n)$ is randomly generated from $P f$,
$y_{i}^{G}(i=1,2, \ldots, n)$ is randomly generated from $P c$; and $R_{C}$ is the mutation factor as follows:

$$
\begin{equation*}
R_{C}=R c_{0} \cdot \frac{\gamma^{G}+\text { const }}{\gamma^{G-1}+\text { const }}, \tag{10}
\end{equation*}
$$

where $R c_{0}$ is the initial value of the variability factor, const is a small constant, to ensure that the fractional is meaningful, and $\gamma^{G}$ is defined as follows:

$$
\begin{equation*}
\gamma^{G}=\frac{\text { the number of infeasible solutions in } P g}{N p} . \tag{11}
\end{equation*}
$$

3.2. Framework of the Proposed Algorithm. The proposed algorithm is described as follows.

Step 1 (initialization). Generate the population $P g, P f$, and $P c$ of size $N p, N P_{1}$, and $N P_{2}$. Set the value of $C_{R}$ (crossover probability), $G_{\max }$ (the number of function evaluations), $G_{S M}$ (the iterative number of evolution by NM simplex method), $g=1$ (the current generation number), and positive control parameter for scaling the difference vectors $F_{1}, F_{2}$. Randomly generate the parent population Pg from the decision space. Set the $P f$, and $P c$, and let the intermediate populations $S g$ and $M g$ be empty.

Step 2 (DE reproduction). By (3) and (9) for mutation, crossover, and selection, an offspring $S g$ is created. Judge the constraints of all individuals in Pg . In accordance with (8), we first calculate constraint violation degree $V_{i}(X)$ of all of the individuals. If $V_{i}(X)=0$, the solution is feasible and preserved to the intermediate set $M g$; if $V_{i}(X)>0$, the solution is infeasible and preserved to the intermediate set $S g$.
Step 3 (simplex method). Apply NM simplex method operator to the present population if $g \geq G_{\max } / 2$. Update the present population $M g$ when the number of iterations exceeds maximum iterations.

Step 4 (Pf construction). Rank chromosomes in $M g$ based on (5), and generate the elitist population $P f$ (the size is $N p$ ) from the ranked population $M g$.
Step 5 (Pc construction). Add the chromosomes in Sg with slight constraint violations to the Pc.

Step 6 (mixing the population). Combine Sg with the existing parent population to form a new set Mg . Remove the duplicate individuals in Mg and the existing parent population.
Step 7 (evolution). Randomly choose chromosomes from $P c, P g$, and $P f$. Use the adaptive differential mutation and uniform discrete crossover to obtain the offspring population $\mathrm{Pg}+1$.

Step 8 (termination). If the stopping criterion is met, stop and output the best solution; else, go to Step 2.
3.3. Experimental Study. In this section, we choose three problems CTP, TNK, and BNH, as shown in Table 3, to test the proposed method, and compare the method with the current CNSGA-II [35].

Table 3: Test functions.

| Test function | Objective function $\min F(X)=\min \left[f_{1}(X), f_{2}(X)\right]$ | Constraints | Range of variable |
| :---: | :---: | :---: | :---: |
| CTP | $\begin{aligned} & f_{1}(X)=x_{1} \\ & f_{2}(X)= \exp \left(-\frac{f_{1}(X)}{c(X)}\right) \\ & \times\left\{41+\sum_{i=2}^{5}\left[x_{i}^{2}-10 \cos \left(2 \pi x_{i}\right)\right]\right\} \end{aligned}$ | $\begin{gathered} g_{1}(X)=\cos (\theta)\left(f_{2}(X)-e\right)-\sin (\theta) f_{1}(X), \\ g_{2}(X)=a \mid \sin \left\{b \pi \left[\sin (\theta)\left(f_{2}(X)-e\right)\right.\right. \\ \left.\left.+\cos (\theta) f_{1}(X)\right]^{c}\right\}\left.\right\|^{d}, \\ g_{1}(X) \geq g_{2}(X) \end{gathered}$ | $\begin{gathered} 0 \leq x_{1} \leq 1 \\ -5 \leq x_{i} \leq 5 \\ i=2,3,4,5 \end{gathered}$ |
| BNH | $\begin{gathered} f_{1}(X)=4 x_{1}^{2}+4 x_{2}^{2}, \\ f_{2}(X)=\left(x_{1}-5\right)^{2}+\left(x_{2}-5\right)^{2} \end{gathered}$ | $\begin{gathered} g_{1}(X)=\left(x_{1}-5\right)^{2}+x_{2}{ }^{2}-25, \\ g_{2}(X)=-\left(x_{1}-8\right)^{2}+\left(x_{2}+3\right)+7.7, \\ g_{1}(X) \leq 0, g_{2}(X) \leq 0 \\ \hline \end{gathered}$ | $\begin{aligned} & 0 \leq x_{1} \leq 5 \\ & 0 \leq x_{2} \leq 3 \end{aligned}$ |
| TNK | $\begin{aligned} f_{1}(X) & =x_{1}, \\ f_{2}(X) & =x_{2} \end{aligned}$ | $\begin{aligned} & g_{1}(X)=-x_{1}^{2}-x_{2}^{2}+1 \\ &+0.1 \cos \left(16 \arctan \left(\frac{x_{1}}{x_{2}}\right)\right), \\ & g_{2}(X)=\left(x_{1}-0.5\right)^{2}+\left(x_{2}-0.5\right), \\ & g_{1}(X) \leq 0, g_{2}(X)-0.5 \leq 0 \end{aligned}$ | $\begin{gathered} 0 \leq x_{i} \leq \pi \\ i=1,2 \end{gathered}$ |



Figure 5: CTP solution space.

For CTP problem, there are the six parameters $\theta, a, b, c$, $d$, and $e$ that must be chosen in a way so that a portion of the unconstrained Pareto-optimal region is infeasible. Each constraint is an implicit non-linear function of decision variables. Thus, it may be difficult to find a number of solutions on a non-linear constraint boundary. We take two sets of values for six parameters in CTP problem, which are determined as (1) CTP1: $\theta=0.1 \pi, a=40, b=0.5, c=1$, $d=2$, and $e=-2$; (2) CTP2: $\theta=-0.2 \pi, a=0.2, b=10$, $c=1, d=6$, and $e=1$. The Pareto frontiers, the feasible solution spaces, and the infeasible solution spaces are shown in Figure 5.

The parameters are initialized as follows. The size of population Pg is $N p=200$, size of $P f$ is $N P_{1}=150$, size of $P c$ is
$N P_{2}=10$, scaling factors $F_{1}$ and $F_{2}$ are randomly generated within $[0.5,1]$, cross rate is $C_{R}=0.6$, maximum evolution generation is $G_{\max }=300$, and number of SM evolution iterations is $G_{S M}=100$. All of the proposed algorithms and CNSGA-II are real coded with equal population size and equal maximum evolution generation. Each algorithm runs 20 times independently for each test function.

Figure 6 shows the result of a random running of the proposed algorithm and NSGA-II, the smooth curve "-" represents the Pareto frontier, and " " stands for the solution achieved by the proposed algorithm or NSGA-II.

It is obvious that the proposed algorithm returns a better approximation to the true Pareto-optimal frontie and a distribution of higher uniformity. We also evaluated the


Figure 6: Continued.


Figure 6: The comparison chart of Pareto frontier.

Table 4: The comparison of performance.

| Test function | Algorithm | $\gamma$ | SP |
| :--- | :---: | :---: | :---: |
| CTP1 | CNSGA- $\Pi$ | $0.021317 \pm 0.000323$ | $0.873321 \pm 0.08725$ |
|  | The proposed | $0.009836 \pm 0.000410$ | $0.567933 \pm 0.01845$ |
| CTP2 | CNSGA- $\Pi$ | $0.011120 \pm 0.000753$ | $0.78314 \pm 0.02843$ |
|  | The proposed | $0.007013 \pm 0.000554$ | $0.296543 \pm 0.00453$ |
| BNH | CNSGA- $\Pi$ | $0.014947 \pm 0.000632$ | $0.336941 \pm 0.00917$ |
|  | The proposed | $0.013766 \pm 0.000043$ | $0.209321 \pm 0.00561$ |
| TNK | CNSGA- $\Pi$ | $0.013235 \pm 0.000740$ | $0.464542 \pm 0.00730$ |
|  | The proposed | $0.006435 \pm 0.000017$ | $0.224560 \pm 0.00159$ |



Figure 7: Flow diagram of the grinding and classification process.
algorithms based on the two criterions $\gamma$ and SP , as shown in Table 4. It can be observed from the data in Table 4 that the proposed algorithm performs significantly better than the classical CNSGA-II algorithm in convergence and distribution uniformity. The simulation results show that this algorithm can accurately converge to global Pareto solutions and can maintain diversity of population.

## 4. Optimization of Grinding and Classification Process

4.1. Bauxite Grinding and Classification Process. The grinding and classification process is the key preparation for the bauxite mineral processing. Here, we consider a bauxite grinding process in a certain mineral company with single grinding and two-stage classification, as shown in Figure 7.

The process consists in a grinding ball mill and two spiral classifiers. First classifier recycle will be put back to the ball mill for regrinding, and the first-stage overflow will be put into second spiral classifier after being mixed with water; the second classifier recycle will be prepared for Bayer production as the rough concentrate, and the secondstage overflow will be sent to the next flotation process. The production objectives are composed of the production yields, technically represented by the solid flow of feed ore since the process is nonstorable, and the mineral processing quality, represented by percentage of the small-size fractions of mineral particles in the second-stage overflows.

Table 5: Notations for the model of the grinding and classification process.

| Notation | Description |
| :---: | :---: |
| $p_{i}$ | Particle percentage of the $i$ th size fraction in the ball mill overflow |
| $f_{j}$ | Particle percentage of the $i$ th size fraction in the Feed ore |
| C | Rate of the first classifier recycle |
| $E a_{i}$ | The efficiency of the first spiral classifier |
| $\tau$ | The mean residence time |
| $P_{1}$ | The internal concentration in ball mill |
| $M_{\text {MF }}$ | The solid flow of feed ore |
| $W_{1}$ | The water addition of the first classifier recycle |
| $W_{2}$ | The classifier water addition |
| $b_{i j}$ | The breakage distribution function |
| $S_{i}$ | The breakage rate function |
| $d_{i}$ | The particle with the $i$ th size |
| $\alpha c_{i}$ | The particle percentage of the $i$ th size in the first classifier overflow |
| $d_{0}$ | The unit size of the particle |
| $\mathrm{P}_{2}$ | First-stage overflow |
| $\phi_{B}$ | Ball filling rate |
| $A_{c}, B_{c}$ | Parameters of the first-stage classifier overflow size fraction distribution |
| $E a_{i}$ | The efficiency of the first spiral classifier |
| $d_{50}$ | The particle size fraction after correction separation |
| $\underline{m}$ | The separation accuracy |
| $k$ | The intermix index |
| $E a_{\text {min }}^{\prime}, d_{50 c}^{\prime}, m^{\prime}, k^{\prime}$ | The corresponding key parameters to the efficiency of the second spiral classifier |
| a, $\alpha, \mu, \Lambda$ | Four key parameters to control the breakage rate function |
| $d_{\text {min }}, d_{\text {max }}$ | The minimum and maximum particle sizes |
| $A_{\text {MF }}, B_{\text {MF }}$ | Parameters of feed ore size fraction distribution |
| $F(i)$ | The cumulative particle percentage less than the $i$ th size fraction in feed ore |
| $\alpha c_{i}^{\prime}$ | The particle percentage of the $i$ th size fraction in the second classifier overflow |
| $E a_{i}^{\prime}$ | The efficiency of the second spiral classifier |

4.2. Predictive Model of the Grinding and Classification Process. Here, we establish the mathematical predictive model of each unit process in the bauxite grinding and classification process. The notations of the indexes, decision variables, and parameters are listed in Table 5. These notations will be used for the model of the grinding and classification process.
4.2.1. Ball Mill Circuit Model. Here, $p_{i}$ is the particle percentage of $i$ th size fraction in the ball mill overflow, $f_{j}$ is the particle percentage of $i$ th size fraction in feed ore, rate of the first classifier recycle $C$ is known, and $E a_{i}$ is the efficiency of
the first spiral classifier. According to a technical report of field investigation and study, we have that

$$
\begin{align*}
& p_{i}(1+C)=\frac{d_{i i} f_{i}+\sum_{j=1, i>1}^{i-1} d_{i j}\left[E a_{i} p_{j}(1+C)+f_{j}\right]}{1-d_{i i} E a_{i}}, \\
& d_{i j}= \begin{cases}e_{j}, & i=j, \\
\sum_{k=j}^{i-1} c_{i k} c_{j k}\left(e_{k}-e_{i}\right), & i>j,\end{cases} \\
& c_{i j}= \begin{cases}-\sum_{k=i}^{j-1} c_{i k} c_{j k}, & i<j, \\
1, & i=j, \\
\frac{1}{S_{i}-S_{j}} \sum_{k=j}^{i-1} S_{k} b_{i k} c_{k j}, & i>j,\end{cases} \\
& e_{j}=\frac{1}{\left(1+0.5 \cdot \tau S_{j}\right)\left(1+0.25 \cdot \tau S_{j}\right)^{2}}, \quad \tau=\frac{8 P_{1}}{M_{\mathrm{MF}}}, \tag{12}
\end{align*}
$$

where $\tau$ is the mean residence time of minerals, $P_{1}$ is the internal concentration in ball mill, and

$$
\begin{equation*}
P_{1}=\frac{M_{\mathrm{MF}}(1+C)}{M_{\mathrm{MF}}(1+C)+W_{1}+0.3\left(W_{1}+W_{2}\right)} \tag{13}
\end{equation*}
$$

where $M_{\mathrm{MF}}(t / h)$ is the solid flow of feed ore, $W_{1}$ is the water addition of the first classifier recycle, and $W_{2}$ is the classifier water addition. $b_{i j}$ is the breakage distribution function; $S_{i}$ is the breakage rate function, and it satisfied the following equation:

$$
\begin{equation*}
S_{i}=\frac{\left(a\left(d_{i} / d_{o}\right)^{\alpha}\right)}{\left(1+\left(d_{i} / \mu\right)^{\Lambda}\right)}, \tag{14}
\end{equation*}
$$

where $d_{i}$ is the particle with the $i$ th size, $d_{o}$ it is a unit, when per millimeter is a unit, $d_{o}=1, d_{i}=i(\mathrm{~mm})$, and $a, \alpha, \mu$, and $\Lambda$ are four key parameters to control the breakage rate function.

In a concrete grinding and classification process, the ball mill size is fixed, and the speed of ball mill is constant. Through data acquisition and testing of grinding and classification steady-state loop, the regression model between $a$, $\alpha, \mu, \Lambda$ and condition variables, size fraction distribution can be established. The input variables are ball filling rate $\phi_{B}$, solid flow of feed ore $M_{\mathrm{MF}}$, water addition of the first classifier recycle $W_{1}$, and parameters of feed ore size fraction distribution $A_{\mathrm{MF}}, B_{\mathrm{MF}}$. The regression model is

$$
\begin{align*}
{\left[\begin{array}{llll}
a & \alpha & \mu & \Lambda
\end{array}\right]^{T}=} & {\left[\begin{array}{ccc}
x_{11} & \cdots & x_{1 j} \\
\vdots & \ddots & \vdots \\
x_{i 1} & \cdots & x_{i j}
\end{array}\right] }  \tag{15}\\
& \cdot\left[\begin{array}{lllll}
W_{1} & \phi_{B} & A_{\mathrm{MF}} & B_{\mathrm{MF}} & M_{\mathrm{MF}}
\end{array}\right]^{T}
\end{align*}
$$

The value of $x_{i j}(i=1,2,3,4 ; j=1,2, \ldots, 5)$ can be obtained by the experimental data regression, $A_{\mathrm{MF}}, B_{\mathrm{MF}}$ can
be obtained from feed ore size fraction distribution, and $F(i)$ is the cumulative particle percentage less than the $i$ th size fraction in feed ore, and it is represented as follows:

$$
\begin{equation*}
F(i)=1-\exp \left(-A_{\mathrm{MF}} d_{i}^{B_{\mathrm{MF}}}\right) . \tag{16}
\end{equation*}
$$

4.2.2. Spiral Classifier Model. $\alpha c_{i}$ is the particle percentage of the $i$ th size fraction in the first classifier overflow, and $p_{i}$ is the particle percentage of the $i$ th size fraction in ball mill overflow. The spiral classifier model is as follows:

$$
\begin{equation*}
\alpha c_{i}=\frac{p_{i} \times\left(1-E a_{i}\right)}{\sum_{i=1}\left(p_{i} \times\left(1-E a_{i}\right)\right)} \times 100 \% \tag{17}
\end{equation*}
$$

where $E a_{i}$ is the efficiency of the first spiral classifier and the mechanism formula of $E a_{i}$ is shown as follows:

$$
\begin{align*}
E a_{i}= & 1-\exp \left[-0.693\left(\frac{d_{i}-d_{\min }}{d_{50 c}-d_{\min }}\right)^{m}\right] \\
& +E a_{\min } \cdot\left[1-\left(\frac{d_{i}-d_{\min }}{d_{\max }}\right)\right]^{k} \tag{18}
\end{align*}
$$

where $d_{i}$ is the particle with the $i$ th size, $d_{\text {min }}$ and $d_{\text {max }}$ represent maximum and minimum particle sizes, $d_{50 c}$ is the particle size fraction after correction separation, $m$ is separation accuracy, and $k$ is intermix index.

Through data acquisition and testing of grinding and classification steady-state loop, the regression model between classification parameters and condition variables, size fraction distribution can be established. The input variables include the solid flow of feed ore $M_{\mathrm{MF}}$, the classifier water addition $W_{2}$, and the parameters of ball mill overflow size fraction distribution $A_{\mathrm{MF}}, B_{\mathrm{MF}}$. The regression model is shown as follows:

$$
\begin{align*}
{\left[\begin{array}{llll}
E a_{\min } & d_{50 c} & m & k
\end{array}\right]^{T}=} & {\left[\begin{array}{ccc}
y_{11} & \cdots & y_{1 j} \\
\vdots & \ddots & \vdots \\
y_{i 1} & \cdots & y_{i j}
\end{array}\right] }  \tag{19}\\
& \cdot\left[\begin{array}{llll}
M_{\mathrm{MF}} & W_{2} & A_{\mathrm{MP}} & B_{\mathrm{MP}}
\end{array}\right]^{T},
\end{align*}
$$

where the value of $y_{i j}(i=1,2,3,4 ; j=1,2,3,4)$ can be obtained by data regression.

The first-stage overflow $P_{2}$ calculation formula is as follows:

$$
\begin{equation*}
P_{2}=\frac{M_{\mathrm{MF}}}{\left(M_{\mathrm{MF}}+W_{1}+W_{2}\right)} \tag{20}
\end{equation*}
$$

Similarly, we can get the second spiral classifier model as follows:

$$
\begin{equation*}
\alpha c_{i}^{\prime}=\frac{\alpha c_{i} \times\left(1-E a_{i}^{\prime}\right)}{\sum_{i}\left(\alpha c_{i} \times\left(1-E a^{\prime}\right)\right)} \times 100 \%, \tag{21}
\end{equation*}
$$

where $\alpha c_{i}^{\prime}$ is the particle percentage of the $i$ th size fraction in the second classifier overflow, $\alpha c_{i}$ is the particle percentage of the $i$ th size fraction in the first classifier overflow, and
$E a_{i}^{\prime}$ is the efficiency of the second spiral classifier. The spiral classifier model is as follows:

$$
\begin{align*}
E a_{i}^{\prime}= & 1-\exp \left[-0.693\left(\frac{d_{i}^{\prime}-d_{\min }^{\prime}}{d_{50 c}^{\prime}-d_{\min }^{\prime}}\right)^{m^{\prime}}\right]  \tag{22}\\
& +E a_{\min }^{\prime} \cdot\left[1-\left(\frac{d_{i}^{\prime}-d_{\min }^{\prime}}{d_{\max }^{\prime}}\right)\right]^{k^{\prime}},
\end{align*}
$$

where, $E a_{\min }^{\prime}, d_{50 c}^{\prime}, m^{\prime}$, and $k^{\prime}$ are key parameters to the efficiency of the second spiral classifier. Through data acquisition and testing of grinding and classification steady-state loop, the regression model between classification parameters and condition variables, size fraction distribution can be established. The input variables include solid flow of feed ore $M_{\mathrm{MF}}$ and parameters of the first-stage classifier overflow size fraction distribution $A_{c}, B_{c}$, which are solved by similar equation to (20). The regression model is shown as follows:

$$
\begin{align*}
{\left[\begin{array}{llll}
E a_{\min }^{\prime} & d_{50 c}^{\prime} & m^{\prime} & k^{\prime}
\end{array}\right]^{T}=} & {\left[\begin{array}{ccc}
y_{11}^{\prime} & \cdots & y_{1 j}^{\prime} \\
\vdots & \ddots & \vdots \\
y_{i 1}^{\prime} & \cdots & y_{i j}^{\prime}
\end{array}\right] }  \tag{23}\\
& \cdot\left[\begin{array}{llll}
M_{\mathrm{MF}} & A_{c} & B_{c}
\end{array}\right]^{T},
\end{align*}
$$

where the value of $y_{i j}^{\prime}(i=1,2,3,4 ; j=1,2,3)$ can be obtained by experimental data regression.
4.3. Optimization Model of Grinding and Classification Process. Two objective functions in the process are identified: one is to maximize output $f_{1}(X)$, and the other is to maximize the small-size fractions (less than 0.075 mm fractions) in the second-stage overflow $f_{2}(X)$. It is also necessary to ensure that the grinding product meets all of other technical requirements and the least disturbance in the following flotation circuit. As the constraints, the feed load of the grinding circuit $M_{\mathrm{MF}}$, the steel ball filling rate $\phi_{B}$, the first and the second overflows $P_{1}$ and $P_{2}$, and the particle percentage of fine size fraction in the first and the second classifier overflows $\alpha c_{-0.075}$ and $\alpha c_{-0.075}^{\prime}$ should be within the user specified bounds.

The operation variables are the solid flow of feed ore $M_{\mathrm{MF}}$, water addition of the first classifier recycle $W_{1}$, ball filling rate $\phi_{B}$, and water addition of the second classifier $W_{2}$. Based on all of the above, grinding and classification process multiobjective optimization model is as follows:

$$
\begin{array}{ll}
\max & F=\max \left[f_{1}(X), f_{2}(X)\right], \\
& f_{1}(X)=M_{\mathrm{MF}}, \\
& f_{2}(X)=\alpha c_{-0.075}^{\prime} \\
& \\
& =f\left(M_{\mathrm{MF}}, W_{1}, W_{2}, \phi_{B}\right),  \tag{24}\\
\text { s.t. } & M_{\mathrm{MFmin}} \leq M_{\mathrm{MF}} \leq M_{\mathrm{MFmax}}, \\
& \phi_{B \min } \leq \phi_{B} \leq \phi_{B \max }, \\
& P_{1 \min } \leq P_{1} \leq P_{1 \max }, \\
& P_{2 \min } \leq P_{2} \leq P_{2 \max }, \\
& \alpha c_{-0.075} \geq \alpha c_{\min }, \\
& \alpha c_{-0.075}^{\prime} \geq \alpha c_{\min }^{\prime} .
\end{array}
$$

Table 6: The optimization results calculated by the proposed algorithm.

| Number | $f_{1}(X)(t / h)$ | $f_{2}(X)(\%)$ |
| :--- | :---: | :---: |
| 1 | 92.972 | 90.814 |
| 2 | 91.810 | 91.900 |
| 3 | 92.360 | 90.923 |
| 4 | 90.620 | 93.460 |
| 5 | 91.310 | 92.494 |
| 6 | 90.170 | 93.800 |
| 7 | 89.390 | 95.200 |
| 8 | 89.480 | 94.932 |
| 9 | 91.530 | 92.400 |
| 10 | 89.298 | 95.763 |
| 11 | 89.824 | 94.549 |
| 12 | 89.800 | 94.395 |
| 13 | 89.710 | 94.618 |
| 14 | 91.170 | 92.800 |
| 15 | 91.880 | 91.840 |
| 16 | 92.190 | 91.281 |
| 17 | 89.870 | 94.090 |
| 18 | 91.000 | 93.285 |
| 19 | 91.960 | 91.520 |
| 20 | 91.124 | 93.221 |



Figure 8: The comparison chart between optimization results and industrial data.

In (24), $f\left(M_{\text {MF }}, W_{1}, W_{2}, \phi_{B}\right)$ implicates the model of grinding and classification represented by (12)-(23). The proposed algorithm is applied to solve the problem, and the optimization results are shown in Table 6.

With the practical process data from a grinding circuit of a mineral plant, the simulation of this hybrid intelligent method adopted the same parameters on the variation in fresh slurry feed velocity, density, particle size distribution, and cyclone feed operating configurations.

The comparison of production data and optimization results in Table 6 is shown in Figure 8, where " $\downarrow$ " represents the proposed algorithm optimization results and " o " represents the original data collected from the field without optimization of raw data. According to the objectives, the data point closer to the upper right edge is more beneficial. Obviously, the proposed optimization result is far better than the original data, indicating the effectiveness of the optimization approach.
4.4. TOPSIS Method for Solution Selection. The resolution of a multiobjective optimization problem does not end when the Pareto-optimal set is found. In practical operational problems, a single solution must be selected. TOPSIS [36] is a useful technique in dealing with multiattribute or multicriteria decision-making (MADM/MCDM) problems in the real world. The standard TOPSIS method attempts to choose alternatives that simultaneously have the shortest distance from the positive-ideal solution and the farthest distance from the negative-ideal solution. According to the TOPSIS method, the relative closeness coefficient is calculated, and the best solution in Table 6 is the solution number 10 as $f_{1}(X)=89.298(t / h)$ and $f_{2}(X)=95.763(\%)$. The corresponding decision variables are $M_{\mathrm{MF}}=89.298(t / h)$, $\phi_{B}=32 \%, W_{1}=75.127(t / h)$, and $W_{2}=16.296(t / h)$.

## 5. Conclusions

Promoted by the requirements of engineering optimization in complex practical processes of grinding and classification, we proposed a hybrid multiobjective differential evolution algorithm with a few beneficial features integrated. Firstly, an archiving mechanism for infeasible solutions is established with functional partitioned multi-population, which aims to direct the population to approach or land in the feasible region from different directions during evolution. Secondly, we propose an infeasible constraint violations function to select infeasible population with better performance, so that they are allowed to be saved and to participate in the subsequent evolution. Thirdly, a nondominating ranking strategy is designed to improve the crowding-distance sorting and return uniform distribution of Pareto solutions. Finally, the simplex method is inserted in the differential evolution process to purposefully enrich the diversity without excessive computation cost. The advantage of the proposed algorithm is the exemption from constructing penalty function and the preservation of meaningful infeasible solutions directly. Simulation results on benchmarks indicate that the proposed algorithm can converge quickly and effectively to the true Pareto frontier with better distribution.

Based on the investigated information about grinding circuit process, we established a multiobjective optimal model with equations from mechanism knowledge, parameters recognized by data regression, and constraints of technical requirements. The nonlinear multiobjective optimization model is too complicated to be solved by traditional gradientbased algorithms. The proposed hybrid differential evolution algorithm is applied and tested to achieve a Pareto solution
set. It is proven to be valuable for operation decision making in the industrial process and showed superiority to the operation carried out in the production. In fact, many operating parameters in complex processes are highly coupled and conflicting with each other. The optimal operation of the entire production process is very difficult to obtain by manual calculation; let alone the fluctuation situation of process conditions. The application case indicates that the proposed method has good performance and is helpful to inspire further research on evolutionary methods for engineering optimization.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (61374156, 61273187, and 61134006), the National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2012BAK09B04), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (61321003), the Fund for Doctor Station of the Ministry of Education (20110162130011 and 20100162120019) and the Hunan Province Science and Technology Plan Project (2012CK4018).

## References

[1] K. Mitra and R. Gopinath, "Multiobjective optimization of an industrial grinding operation using elitist nondominated sorting genetic algorithm," Chemical Engineering Science, vol. 59, no. 2, pp. 385-396, 2004.
[2] G. Yu, T. Y. Chai, and X. C. Luo, "Multiobjective production planning optimization using hybrid evolutionary algorithms for mineral processing," IEEE Transactions on Evolutionary Computation, vol. 15, no. 4, pp. 487-514, 2011.
[3] T. Y. Ma, W. H. Gui, Y. L. Wang, and C. H. Yang, "Dynamic optimization control for grinding and classification process," Control and Decision, vol. 27, no. 2, pp. 286-290, 2012 (Chinese).
[4] F. Y. Edgeworth, Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences, C. Kegan Paul \& Company, London, UK, 1881.
[5] V. Pareto, Cours d'Economie Politique, F. Rouge, Lausanne, Switzerland, 1896.
[6] X. F. Chen, W. H. Gui, Y. L. Wang, and L. Cen, "Multi-step optimal control of complex process: a genetic programming strategy and its application," Engineering Applications of Artificial Intelligence, vol. 17, no. 5, pp. 491-500, 2004.
[7] C. A. C. Coello, "Evolutionary multi-objective optimization: a historical view of the field," IEEE Computational Intelligence Magazine, vol. 1, no. 1, pp. 28-36, 2006.
[8] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II", IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182197, 2002.
[9] C. M. Fonesca and P. J. Fleming, "Genetic algorithms for multi objective optimization: formulation, discussion and generalization," in Proceedings of the 5th International Conference on Genetic Algorithms, pp. 415-423, Morgan Kaufmann, San Mateo, Calif, USA, 1993.
[10] S. Sengupta, S. Das, M. Nasir, A. V. Vasilakos, and W. Pedrycz, "Energy-efficient differentiated coverage of dynamic objects using an improved evolutionary multi-objective optimization algorithm with fuzzy-dominance," in Proceedings of IEEE Congress on Evolutionary Computation, pp. 1-8, IEEE Press, Brisbane, Australia, 2012.
[11] K. V. Price, R. M. Storn, and J. A. Lampinen, Differential Evolution a Practical Approach to Global Optimization, Springer, Berlin, Germany, 2005.
[12] W. Y. Gong and Z. H. Cai, "An improved multiobjective differential evolution based on Pareto-adaptive $\varepsilon$-dominance and orthogonal design," European Journal of Operational Research, vol. 198, no. 2, pp. 576-601, 2009.
[13] X. H. Zeng, W. K. Wong, and S. Y. Leung, "An operator allocation optimization model for balancing control of the hybrid assembly lines using Pareto utility discrete differential evolution algorithm," Computers and Operations Research, vol. 39, no. 5, pp. 1145-1159, 2012.
[14] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," IEEE Transactions on Evolutionary Computation, vol. 3, no. 4, pp. 257-271, 1999.
[15] A. Al-Ani, A. Alsukker, and R. N. Khushaba, "Feature subset selection using differential evolution and a wheel based search strategy," Swarm and Evolutionary Computation, vol. 9, pp. 1526, 2013.
[16] G. Y. Li and M. G. Liu, "The summary of differential evolution algorithm and its improvements," in Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE '10), pp. V3153-V3156, Chengdu, China, August 2010.
[17] H. A. Abbass, "The self-adaptive Pareto differential evolution algorithm," in Proceedings of the IEEE Congress on Evolutionary Computation, pp. 831-836, IEEE Press, Honolulu, Hawaii, USA, May 2002.
[18] N. K. Madavan, "Multiobjective optimization using a Pareto differential evolution approach," in Proceedings of the IEEE Congress on Evolutionary Computation, pp. 1145-1150, IEEE Press, Honolulu, Hawaii, USA, May 2002.
[19] F. Xue, A. C. Sanderson, and R. J. Graves, "Pareto-based multiobjective differential evolution," in Proceedings of the IEEE Congress on Evolutionary Computation,, pp. 862-869, IEEE Press, Piscataway, NJ, USA, December 2003.
[20] T. Robic and B. Filipi, "DEMO: differential evolution for multiobjective optimization," in Evolutionary Multi-Criterion Optimization, pp. 520-533, Springer, Berlin, Germany, 2005.
[21] V. L. Huang, A. K. Qin, P. N. Suganthan, and M. F. Tasgetiren, "Multi-objective optimization based on self-adaptive differential evolution algorithm," in Proceedings of the IEEE Congress on Evolutionary Computation (CEC '07), pp. 3601-3608, IEEE Press, Singapore, September 2007.
[22] V. L. Huang, S. Z. Zhao, R. Mallipeddi, and P. N. Suganthan, "Multi-objective optimization using self-adaptive differential evolution algorithm," in Proceedings of the IEEE Congress on Evolutionary Computation (CEC '09), pp. 190-194, IEEE Press, Trondheim, Norway, May 2009.
[23] J. A. Adeyemo and F. A. O. Otieno, "Multi-objective differential evolution algorithm for solving engineering problems," Journal of Applied Sciences, vol. 9, no. 20, pp. 3652-3661, 2009.
[24] Y. Wang and Z. Cai, "Combining multiobjective optimization with differential evolution to solve constrained optimization problems," IEEE Transactions on Evolutionary Computation, vol. 16, no. 1, pp. 117-134, 2012.
[25] A. Homaifar, C. X. Qi, and S. H. Lai, "Constrained optimization via genetic algorithms," Simulation, vol. 62, no. 4, pp. 242-253, 1994.
[26] J. A. Joines and C. R. Houck, "On the use of non-stationary penalty functions to solve nonlinear constrained optimization problems with GA's," in Proceedings of the 1st IEEE Conference on Evolutionary Computation, pp. 579-584, Orlando, Fla, USA, June 1994.
[27] F. Jiménez and J. L. Verdegay, "Evolutionary techniques for constrained optimization problems," in Proceedings of the 7th European Congress on Intelligent Techniques and Soft Computing (EUFIT '99), Aachen, Germany, 1999.
[28] C. Sun, J. Zeng, and J. Pan, "An improved vector particle swarm optimization for constrained optimization problems," Information Sciences, vol. 181, no. 6, pp. 1153-1163, 2011.
[29] Z. Michalewicz and M. Schoenauer, "Evolutionary algorithms for constrained parameter optimization problems," Evolutionary Computation, vol. 4, no. 1, pp. 1-32, 1996.
[30] C. A. C. Coello and A. D. Christiansen, "Moses: a multiobjective optimization tool for engineering design," Engineering Optimization, vol. 31, no. 1-3, pp. 337-368, 1999.
[31] Y. G. Woldesenbet, G. G. Yen, and B. G. Tessema, "Constraint handling in multiobjective evolutionary optimization," IEEE Transactions on Evolutionary Computation, vol. 13, no. 3, pp. 514-525, 2009.
[32] J. A. Nelder and R. Mead, "A simplex method for function minimization," The Computer Journal, vol. 7, no. 4, pp. 308-313, 1965.
[33] D. A. van Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithms: analyzing the state-of-the-art", Evolutionary Computation, vol. 8, no. 2, pp. 125-147, 2000.
[34] R. Farmani, D. A. Savic, and G. A. Walters, "Evolutionary multiobjective optimization in water distribution network design," Engineering Optimization, vol. 37, no. 2, pp. 167-183, 2005.
[35] K. Deb and T. Goel, Controlled Elitist Non-Dominated Sorting Genetic Algorithms for Better Convergence, Springer, Berlin, Germany, 2001.
[36] Y. J. Lai, T. Y. Liu, and C. L. Hwang, "Topsis for MODM," European Journal of Operational Research, vol. 76, no. 3, pp. 486-500, 1994.

## Research Article

# Discrete Current Control Strategy of Permanent Magnet Synchronous Motors 

Yan Dong, ${ }^{1}$ Kai Jing, ${ }^{1}$ Hexu Sun, ${ }^{1,2}$ and Yi Zheng ${ }^{1}$<br>${ }^{1}$ School of Control Science and Engineering of Hebei University of Technology, Tianjin 300130, China<br>${ }^{2}$ Hebei University of Science and Technology, Shijiazhuang 050018, China<br>Correspondence should be addressed to Hexu Sun; hxsun@hebust.edu.cn

Received 1 July 2013; Revised 22 September 2013; Accepted 28 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 Yan Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

A control strategy of permanent magnet synchronous motors (PMSMs), which is different from the traditional vector control (VC) and direct torque control (DTC), is proposed. Firstly, the circular rotating magnetic field is analyzed on the simplified model and discredited into stepping magnetic field. The stepping magnetomotive force will drive the rotor to run as the stepping motor. Secondly, the stator current orientation is used to build the control model instead of rotor flux orientation. Then, the discrete current control strategy is set and adopted in positioning control. Three methods of the strategy are simulated in computer and tested on the experiment platform of PMSM. The control precision is also verified through the experiment.


## 1. Introduction

The permanent magnet synchronous motors (PMSMs) have become the popular AC motors and are used in various situations for their advantages of high efficiency, power factor, small size, and avoidance of exciting current. As servo motors, PMSMs are usually controlled with two methods, that is, vector control (VC) by flux orientation and direct torque control (DTC).

VC was put forward in 1971 for asynchronous motor by German engineer Blaschke [1], which was used in PMSM soon afterwards. Generally, the theory is to keep the d component of stator current being 0 in rotor flux reference frame and the torque will be proportionate to the $q$ component of stator current which leads the constant rotor flux by $90^{\circ}$. It is good at torque responding and speed accuracy, but the decoupling of flux and torque needs more focus to design regulator for both. The robustness will be vulnerable [2].

DTC is proposed by Professor Depenbrock in 1985 [3], which is used to directly control the flux and torque by selecting proper voltage vector. This method avoids the decoupling and is simpler than VC, but the torque ripple cannot be avoided which will weaken the dynamic characteristic [4,5].

Both methods are based on rotor flux which needs to be tested by an observer or to be controlled with other
variables $[6,7]$. This paper proposes a strategy based on stator current frame and uses the discrete stator current to control the motor. By using this strategy, the motor will run step by step, and it not only reflects the simply structure and large capacity of PMSM but also provides the advantages of stepping motor such as digital control, discrete operation, and nonaccumulating error. The proposed strategy is a novel control method on PMSMs with simple control structure as compared with the above two classical methods. The wide application prospects and the deep research of it will promote the development of drive technology.

## 2. Discretization of Circular Rotating Magnetic Field

2.1. Stator Model of PMSM. In PMSM, distributed winding, which is used in normal AC motor, is often coiled as shown in Figure 1. Figure 1 shows two structures of 2-pole, 24-slot single-layer 3-phase motor stator winding.

Despite the differences of poles number, slots number, and the coiling form of the 3-phase AC motor, the physical model of stator can be described as in Figure 1 for the symmetry of the magnetic circuit and the magnetomotive force (MMF) generated by powered winding. Figure 2 shows


Figure 1: Distributed winding form.


Figure 2: Simplified stator model of synchronic motor.
the distances of $2 \tau$ about a pair of magnetic poles equivalent to $360^{\circ}$ of electrical angle. Every stator of 3-phase AC motor can be analyzed with this model.
2.2. Circular Rotating Magnetic Field. When powering the stator model with the 3-phase current as (1), setting the positive direction from $a$ to $x, b$ to $y$, and $c$ to $z$, the 3-phase MMF is generated which can be considered as sinusoidal distribution in the stator when excluding space harmonics. Then the MMF can be expressed as (2):

$$
\begin{gathered}
i_{a}=I_{m} \cos \omega t, \\
i_{b}=I_{m} \cos \left(\omega t-120^{\circ}\right),
\end{gathered}
$$

$$
\begin{equation*}
i_{c}=I_{m} \cos \left(\omega t+120^{\circ}\right), \tag{1}
\end{equation*}
$$

$$
\mathbf{F}_{a}(t)=0.5 \mathbf{F}_{a}\left(e^{j \omega t}+e^{-j \omega t}\right)
$$

$$
\begin{equation*}
\mathbf{F}_{b}(t)=0.5 \mathbf{F}_{b}\left(e^{j \omega t} e^{-j 120^{\circ}}+e^{-j \omega t} e^{j 120^{\circ}}\right) \tag{2}
\end{equation*}
$$

$$
\mathbf{F}_{c}(t)=0.5 \mathbf{F}_{c}\left(e^{j \omega t} e^{j 120^{\circ}}+e^{-j \omega t} e^{-j 120^{\circ}}\right)
$$

$\mathrm{F}_{a}$ is an MMF vector generated by the maximum current of A phase, the direction of which is assumed as the horizontal axis of static frame. $\mathbf{F}_{a}(t)$ is determined by $i_{a}$ varied with time $t . \mathbf{F}_{b}$ and $\mathbf{F}_{c}$ are similar to $\mathbf{F}_{a}$, which lead $\mathbf{F}_{a}$ by $120^{\circ}$ and $240^{\circ}$, respectively; $\mathbf{F}_{b}(t)$ and $\mathbf{F}_{c}(t)$ are with the same meaning of $\mathrm{F}_{a}(t)$.

The composite MMF in the air gap will be expressed as

$$
\begin{equation*}
\Sigma \mathbf{F}(t)=\mathbf{F}_{a}(t)+\mathbf{F}_{b}(t)+\mathbf{F}_{c}(t)=1.5 \mathbf{F}_{a} e^{j \omega t} \tag{3}
\end{equation*}
$$

It is a rotating MMF vector, of which the amplitude is 1.5 times of each phase. The electric angle of the MMF rotating in the space corresponds to that of the current changing in the winding, which is

$$
\begin{equation*}
\theta=\omega t \tag{4}
\end{equation*}
$$

When the current changes by a cycle, the rotating MMF goes $2 \tau$ distances in the air gap. The revolution per second is

$$
\begin{equation*}
n_{1}=\frac{f}{p_{\tau}} \tag{5}
\end{equation*}
$$

Where $f$ is the frequency of the stator current and $p_{\tau}$ is the number of pole pairs of the motor.


Figure 3: Stepping MMF of three-phase winding as $b_{H}=6$.
2.3. Discrete Magnetic Field and Positioning Torque. The MMF $\mathbf{F}_{s}$ generated by stator is to drive the rotor MMF $\mathbf{F}_{r}$ to rotate synchronously. The electromagnetic toque $T_{e}$ can be described in terms of $\mathbf{F}_{s}$ and $\mathbf{F}_{r}$ :

$$
\begin{equation*}
T_{e} \propto\left|\mathbf{F}_{r} \times \mathbf{F}_{s}\right|=\mathbf{F}_{r} \mathbf{F}_{s} \sin \theta \tag{6}
\end{equation*}
$$

The $\theta$ is the angle form $\mathbf{F}_{r}$ to $\mathbf{F}_{s}$. If $\mathbf{F}_{s}$ stops rotating at some position and $\mathbf{F}_{r}$ coincides with it, $\theta=0$, the electromagnetic toque will be equal to zero, which will be a positioning point.

If the motor is powered with the currents described in

$$
\begin{gather*}
i_{a}(k)=I_{m} \cos \frac{2 \pi}{b_{H}} k \\
i_{b}(k)=I_{m} \cos \left(\frac{2 \pi}{b_{H}} k-120^{\circ}\right),  \tag{7}\\
i_{c}(k)=I_{m} \cos \left(\frac{2 \pi}{b_{H}} k+120^{\circ}\right),
\end{gather*}
$$

where $b_{H}$ is the number of pulse distributor's beats per cycle, the composite MMF will stop at some point as the pulse number $k$ which is a positive integer not to change. When the next pulse emits, $k=k+1$, the composite MMF will go forward with a little angle just like a step. Then, the rotating MMF in the last section is discretized into stepping MMF [8] expressed in

$$
\begin{equation*}
\Sigma \mathbf{F}(k)=1.5 \mathbf{F}_{a} e^{j\left(2 \pi / b_{H}\right) k} \tag{8}
\end{equation*}
$$

An example as $b_{H}=6$ will illustrate the stepping MMF graphically.

Each MMF will generate a positioning point, and the torque driving the rotor MMF to approach this point is defined as positioning torque. Here, the angle is calculated by electric angle; the actual step number $b$ per revolution and the stepping angle $\alpha$ are expressed as the following formula with the number of pole pairs $p_{\tau}$ :

$$
\begin{gather*}
b=p_{\tau} b_{H}, \\
\alpha=\frac{360^{\circ}}{b}=\frac{360^{\circ}}{p_{\tau} b_{H}} . \tag{9}
\end{gather*}
$$



Figure 4: Vector diagram of stator current orientation.

The stepping angle is determined by $p_{\tau}$ and $b_{H}$. If one wants to increase the stepping number per revolution, it is better to increase $b_{H}$, since the number of pole pairs is constrained by motor structure.

## 3. PMSM Model for Step Motion

3.1. Motor Model by Stator Current Orientation. Make the angular speed of the rotating frame equal to that of stator current vector in general frame of PMSM which is shown in Figure 4 based on the $\alpha-\beta$ static frame. The rotating frame is built by $\mathbf{i}_{s}$, the horizontal axis coinciding with $\mathbf{i}_{s}$ is named $d$ axis, and the vertical axis orthogonal to $d$-axis is $q$-axis. Then, general frame becomes the $d-q$ frame orientated by stator current [9]. In the figure, the angle from $\psi_{r}$ to $\mathbf{i}_{s}$ is assumed as $\varepsilon$, and $\theta_{s}$ and $\theta_{r}$ represent the angle form $\alpha$-axis to $\mathbf{i}_{s}$ and $\psi_{r}$, respectively. $\omega$ is the angular speed of the rotating frame.

The two components of $\mathbf{i}_{s}$ in the frame, named $i_{s d}$ and $i_{s q}$, are expressed as

$$
\begin{gather*}
i_{s d}=i_{s}=\left|\mathbf{i}_{s}\right|, \\
i_{s q}=0 . \tag{10}
\end{gather*}
$$

According to the mathematical expression of PMSM on rotating frame, the flux function can be rewritten as the following equation:

$$
\begin{gather*}
\psi_{s d}=L_{d} i_{s d}+\psi_{r d}=L_{d} i_{s}+\psi_{r d}  \tag{11}\\
\psi_{s q}=L_{q} i_{s q}+\psi_{r q}=\psi_{r q}
\end{gather*}
$$

where $\psi_{s d}$ and $\psi_{s q}$ are $d-q$ components of stator flux in rotating frame, $\psi_{r d}$ and $\psi_{r q}$ are $d-q$ components of rotor flux, and $L_{d}$ and $L_{q}$ are $d-q$ components of stator selfinductance. The torque function can be expressed as the following formula with (10) and (11):

$$
\begin{align*}
T_{e} & =p_{\tau}\left|\psi_{s} \times \mathbf{i}_{s}\right| \\
& =p_{\tau} \psi_{s d} i_{s q}-p_{\tau} \psi_{s q} i_{s d}  \tag{12}\\
& =-p_{\tau} \psi_{r q} i_{s} .
\end{align*}
$$

Substituting $\sin (-\varepsilon)=\psi_{r q} / \psi_{r}$ into (12), the electromagnetic torque function can be rewritten as

$$
\begin{equation*}
T_{e}=p_{\tau} i_{s} \psi_{r} \sin \varepsilon \tag{13}
\end{equation*}
$$

$\varepsilon$ is also defined as torque angle; when it is greater than zero, with $\psi_{r}$ being drawn by $\mathbf{i}_{s}$, the electromagnetic torque is positive.
3.2. Structure of the Control System. Unlike the VC and DTC, in this control method, magnitude and phase of stator current are regulated dynamically for best torque responding, instead of keeping the amplitude of stator current and rotor flux or maintaining the angle $\varepsilon$ between the current and the flux equal to $90^{\circ}$. Because the rotor flux is unchanged, the regulable variables of the control system are no other than the magnitude of stator current $\left|i_{s}\right|$ and the angle $\varepsilon$.

The structure of motor control system can be simplified as shown in Figure 5 which includes an inner loop and an outer loop.

The outer loop is the only one closed loop to control the speed or position. In the loop, the input is the rotor angle frequency difference or angle difference of preset and feedback, and the output is preset current vector including the magnitude and the rotation angle. To regulate the two variables, we give the motor the maximum current for maximum torque to start or brake and supply the rated current and adjust the $\varepsilon$ to change the electromagnetic torque when the motor operates steadily.

The inner loop is current loop, in which the three-phase stator current is transformed into current vector on $d$ - $p$ frame and the vector is compared with the preset current vector from the previous regulator. The difference of the current vector is to select the voltage vector for inverter control. It can use the method of direct current control (DCC) in [10], which follows the synchronized on-off principle. The current vector at every time interval is predicted for two possible cases as the following formula:

$$
\begin{align*}
\mathbf{i}_{\alpha, \beta}(k+1)= & \underbrace{\mathbf{i}_{\alpha, \beta}(k)\left(1-\left(T / T_{s}\right)\right)}_{\mathbf{i}_{0(\alpha, \beta)}(k+1)} \\
& +\underbrace{\left(T / L_{s}\right) \mathbf{u}_{\alpha, \beta}(k)}_{\mathbf{i}_{u(\alpha, \beta)}(k+1)}, \tag{14}
\end{align*}
$$

where $\mathbf{i}_{0}$ is the radial naturally decreased current vector, $\mathbf{i}_{u}$ is the applied current vector generated by constant voltage during the sampling interval, and the subscripts $\alpha$ and $\beta$ represent the vector components of static frame. The voltage vector $\mathbf{u}$ at instant $k$ can take the following value by decomposing on static $\alpha-\beta$ frame:

$$
\begin{align*}
u_{\alpha, \beta}(k) & =U_{\mathrm{DC}}\left[\begin{array}{l}
K_{U \alpha}(k) \\
K_{U \beta}(k)
\end{array}\right] \\
& =U_{\mathrm{DC}}\left[\begin{array}{ccc}
\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\
0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}}
\end{array}\right]\left[\begin{array}{l}
s_{T 1} \\
s_{T 2} \\
s_{T 3}
\end{array}\right] . \tag{15}
\end{align*}
$$

$U_{\mathrm{DC}}$ is the DC-link voltage. $s_{T 1}, s_{T 2}$, and $s_{T 3}$ denote the states ( 0 or 1 ) of upper transistors in the inverter, which include six effective vectors $(100,110,010,011,001,101)$ and two zero vectors $(000,111)$. After calculating $i_{0}$, the six voltage vector closest to the direction of the error between $i_{s}^{*}$ and $i_{s}$ is
chosen. Figure 6 shows the particular case of selecting upper transistors 010.
3.3. Discrete Current Control. When the stator is powered with the discrete current as (7), the stator current vector $\mathbf{i}_{s}$ has $b_{H}$ positioning points at the stator circle shaping a regular polygon MMF shown in Figure 3, for example,

$$
\begin{equation*}
\mathbf{i}_{s}=\frac{3}{2} I_{m} e^{j\left(2 \pi / b_{H}\right) k} \tag{16}
\end{equation*}
$$

The angle between the two adjacent current vectors is defined as stepping angle just like the step motor, which is

$$
\begin{equation*}
\theta_{b}=\frac{2 \pi}{b_{H}} \tag{17}
\end{equation*}
$$

Therefore, the torque of PMSM can be written as

$$
\begin{equation*}
T_{e}=T_{\max } \sin \left(k \theta_{b}-p_{n} \alpha\right) \tag{18}
\end{equation*}
$$

where $\alpha$ is the mechanical angle of rotating and $p_{n}$ is the number of pole pairs.

This torque is also called reposition torque, impelling the rotor to run forward to catch up with the stator. Therefore, the stopping point of the stator current vector is the very positing point achieving incremental movement of a motor. Take $b_{H}=$ 12 and $T_{Z}=0$ (motor idling), for example, so the discrete current vector and the position are shown in Figure 7.

The proposed strategy of PMSM is called discrete current control, in which the main control variable is the torque angle between stator current vector and rotor flux vector, and the amplitude of stator current is the rating (except for starting and braking which is the maximum). It is different from VC and DTC, and the latter is to control the angle of flux of stator and rotor keeping the stator flux constant. The proposed strategy is more suitable for positioning because of the characteristic of positioning torque generated by discrete current and stepping motion, and the control process is also easier than the two classical methods.

## 4. Discrete Current Control of PMSM

To describe the proposed control strategy, two errors generated in the operation must be declared.
(1) Static angle error: generated by load torque. It needs an electromagnetic torque to balance, so the torque angle cannot be decreased to zero which become an error for the control.
(2) Dynamic angle error: the following process of rotor is not synchronous with stator current vector. The rotor will lag behind the vector when driving or go beyond the positioning point when braking. But the dynamic error will be disappeared when the rotor stops.
4.1. Pointing Control. Pointing control is a typical discrete control method, controlling the motor to move a step forward every time. Only when the transient process of the first step is completely terminated, the second step begins.


Figure 5: Block diagram of stator current oriented control system PMSM.


FIGURE 6: Current vector regulation based on voltage space vector.

The one-step torque $T_{b}$ should be greater than static load torque, so that the static angle error can be less than a stepping angle. The dynamic angle error, for example, should be less than $150^{\circ}$ to keep the operation not losing its step when $b_{H}=$ 12. The angle of one step is $k \theta_{b}$; the minimum one is $\theta_{b}$ and the maximum one must be less than the dynamic angle error.

The greater the stepping angle, the more serious the oscillation phenomenon near the positing point, which needs to be avoided if possible. The simulation result is shown in Figure 8. The motor is triggered by the step pulse every 0.4 seconds with the rise time of 0.025 s and the overshoot of about $32 \%$. The rotor stopped at the given point after the second oscillation.

The oscillation of pointing control is produced by $\Delta \theta=$ $k \theta_{b}-p_{n} \alpha$, and $\omega$ is not equal to zero at the same time, and the torque near the positing point will be so small. These problems can be solved with "bang-bang control" of optimal time and maximum torque.
(1) The time-optimal method is to brake at a proper time to remove the overshoot. As shown in Figure 9, the preset current vector angle is $\theta_{s}=-\theta_{b}=30^{\circ}$; then the rotor accelerated for $\varepsilon$ which is equal to $\theta_{b}$ when $t=0$. When $t=0.018 \mathrm{~s}$, the vector was back to $0^{\circ}$ and $\varepsilon<0$, and the motor began to decelerate. When $t=0.026 \mathrm{~s}, \theta_{s}=k \theta_{b}=\theta_{r}=p_{n} \alpha$, and $\omega=0$, the vector was set at $\theta_{s}=30^{\circ}$ again and the rotor stopped at the


Figure 7: Positioning star diagram.


Figure 8: Position and speed curve under point control.
positioning point. In the process, the transient time is 0.031 s , which decreases to its $1 / 6$.
(2) Maximum torque control is to give the maximum torque at the accelerating stage and brake with the maximum negative torque when the position is vicinity to the stator current vector. The maximum torque is generated as $\varepsilon=90^{\circ}$. In the simulation shown as Figure 10, transforming time of the vector is at $t_{1}=0.010 \mathrm{~s}$ and $t_{2}=0.017 \mathrm{~s}$. Before $t_{1}$, let $k=3$ and after it $k=-2$, and at $t_{2}$, make $k=1$ to keep the rotor stable. In this control, the transient time is only 0.027 s , which decreases to $1 / 8$ of the original time.


Figure 9: Position and speed curve under time optimization.


Figure 10: Position and speed curve under maximum torque.
4.2. Constant Frequency Control. Some motors need a constant frequency control method, which is only to change the step number in a constant frequency and to keep it not losing its steps. The angle frequency of motor $\omega_{r}$ will follow the given frequency $\omega_{s}$ by $\varepsilon$ which must be less than $180^{\circ}$. After a bit oscillations, the rotor will reach the state of $\omega_{r}=\omega_{s}$, while the given frequency has a maximum critical value named jumping frequency, which is defined as the highest frequency so that the motor does not lose its step. If $\omega_{s}$ is more than the jumping frequency, $\omega_{r}$ cannot catch up with $\omega_{s}$ and the position of rotor will lag behind the stator current vector, which will lead to a serious fault.

In the positioning control of this method, the motor responses will oscillate in starting and braking time. These oscillations can be eliminated by optimal controls as which is used in pointing control. The response curves generated by this method will be shown in the experiment in Section 5.
4.3. Up-Down Frequency Control. It needs more time to accelerate or decelerate for the large-capacity motor, because the rotor could store more kinetic energy. If only give the motor a step change in constant frequency, the dynamic angle error may be over the maximum and lead to steps losing.


Figure 11: Experiment platform.


Figure 12: Digital driving controller.

It is necessary to preset an increment or decrement frequency of the motor to accelerate or decelerate.

The highest frequency is limited by the electromagnetic torque which is a function of angle frequency. A frequency of stator current vector, which is less than the jumping frequency, is given to accelerate at $t=0(+)$. Then the frequency increases gradually and the time interval of every step decreases. The $\varepsilon$ had better to be control in the range of $90^{\circ} \pm \theta_{b}$ to maintain the maximum torque and not to lose its step.

Generally, to obtain a better result of control, this control is designed with closed loop to get an optimal up frequency curve. Moreover, the curve of frequency will be designed as two, three, or five segments according to the travel length. The experiment of three-segment curve is shown in Section 5.

## 5. Experiments

The experiments are based on a device of PMSM, which includes motor and transmission platform and digital driving controller. The platform is shown in Figure 11. The PMSM is of the type of M205B produced by KOLLMONGEN in US with rated power of 1.6 kW , rated voltage of 230 V , continuous rated current of 5.3 A , continuous torque of 4.47 Nm , and maximum revolution of 3600 rpm . The load is a DC generator with 1.1 kW rated power and the transmission ratio is $1: 1$ of the gear box. The connecting mechanism between the two motors is with toque sensor, harmonic reducer,


Figure 13: The structure diagram of control system.


Figure 14: Experiment curve of pointing control.
and inertia wheel. The application of PMSM can be well approximated by these devices.

The digital driving controller is composed of control unit and power amplifier shown in Figure 12. The kernel of control part is a TMS320F240 chip of DSP produced by TI and around it are the peripheral circuit and A/D circuit. The main part of power amplifier is PM15RSH120, which is
a intelligent power module (IPM) produced by Mitsubishi. Beside the IPM, the accessory circuit includes trigger signal driver circuit, special power supply module of JS158, position detecting circuit, current sampling circuit, and protection circuit.

The structure diagram of the control system is shown in Figure 13.


Figure 15: Experiment curve of constant frequency control.


Figure 16: Experiment curve of up-down frequency control.
5.1. Control Curve. In the experiment, the motor is with 2 pairs of pole and the electric angle is $720^{\circ}$ per revolution. We divided the cycle of stator current into 12 parts and the electric angle will be $30^{\circ}$ per step. The number of positioning point will be $12 \times 2$ per revolution and every step is corresponded to $15^{\circ}$.
5.1.1. Pointing Control. The stator current vector is given as formula (7). When $t=0, k=0$ and the motor stays at the initial position. When $t=0.6 \mathrm{~s}$, let $k=1$; the vector will lead the rotor flux by a stepping angle that is equal to $30^{\circ}$ and the rotor will follow the vector by the reposition torque.

The current change of A phase is shown in Figure 14(a) and the responded curve of position and speed is in Figures 14(b) and 14(c).
5.1.2. Constant Frequency Control. In order to watch the control process, this experiment uses a frequency of 0.5 Hz . From Figure 15, the rotor position is following the stator current vector closely and the positioning performance is obvious in the discrete control.
5.1.3. Up-Down Frequency Control. Three-segment-speed curve of motor is used in rapid positioning, which only

Table 1: Experiment data of position precision incensement motion.

| Given step | 1 | 12 | 24 | 100 | 500 | 1000 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Pulse number | 181 | 2052 | 4224 | 17002 | 85452 | 170702 |
| Rotating angle | $15.9^{\circ}$ | $180.35^{\circ}$ | $371.25^{\circ}$ | $1494^{\circ}$ | $7510.4^{\circ}$ | $62464.02^{\circ}$ |
| Actual step | 1.1 | 12.02 | 12.7 | 99.6 | 500.7 | 1000.2 |

TABLE 2: Experiment data of operating 160 revolutions.

| No. | 1 | 2 | 3 | 4 | 5 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Distance (pulse <br> number) | 655202 | 655365 | 655369 | 655406 | 655485 |
| Time (s) | 5.78 | 5.80 | 5.60 | 5.69 | 5.72 |
| Error (pulses <br> number) | 158 | 5 | 9 | 46 | 125 |

includes accelerating, constant speed, and decelerating. The experiment curve is shown in Figure 16(a) and the positioning accuracy is limited below a stepping angle. The currentfollowing curve is shown in Figure 16(b), in which the actual current curve is moved down a division of oscilloscope for watching clearly.
5.2. Analyses. Analyzing the error of stepping control of PMSM, we can gain the precision of it used in positioning. The steady error is less than one stepping angle which is $15^{\circ}$ here. If we use the pulses of rotary encoder, of which $360^{\circ}$ is corresponded to 4096 pulses, to stand for the absolute position, we can get a table of precision.

When driving the motor to run 160 revolutions, the emitting pulses and the operation time are shown in Tables 1 and 2.

If use open loop control method and let the speed follow the three-segment curve, when the rotor moves 160 revolutions, then the number of pulses is 655360 , and we get the result recorded in Table 2.

It is proved that the discrete current vector method of PMSM has more advantages than existing methods. Firstly, the structure is simply just using single loop. Secondly, the control method with discrete MMF can generate the larger torque to start or drive the high inertia loads. Thirdly, positioning precision is determined by the stepping angle that can get higher accuracy. Moreover, the reliability and robustness of this method are better than those of the original driver which needs to often change its parameter especially for high inertia loads.

## 6. Conclusion

In this paper, a stepping control method of PMSM is presented. In the method, the circle of rotating MMF is discretized to regular polygon, and in this case, the positioning on stator current orientation has been discussed with the mechanism model of PMSM. The three methods of control are simulated and tested in experiment, which is available with a general DSP controller.

Although good performance is achieved, the method needs deeper studies in theory and applications, such as current responding, harmony wave analysis of discrete current, and influence of the method to grid. Our further works in this area will be oriented to implementation of this method in transmission technology of valve and artillery in order to improve the performance and efficiency and simplify structure.

## Acknowledgments

This work is supported by the Natural Science Funds of Hebei Province (E2013202108) and by the National High Technology Research and Development Program of China (863 Program) (2006AA040306).

## References

[1] T. F. Blaschke, "Das Prinzip der Feldorientierung, die Grundlage fur die TRANSVEKTOR-Regelung von Asynchronmaschinen," Siemens Zeitschrift, vol. 45, no. 10, pp. 757-760, 1971.
[2] A. B. Nikolic and B. I. Jeftenic, "Precise vector control of CSI fed induction motor drive," European Transactions on Electrical Power, vol. 16, no. 2, pp. 175-188, 2006.
[3] M. Depenbrock, "Direkte selbstregelung, (DSR) für hochdynamische mit umrichtersptisun,"ETZ-Archiv, vol. 7, no. 7, pp. 211-216, 1985.
[4] S. M. Gadoue, D. Giaouris, and J. W. Finch, "Artificial intelli-gence-based speed control of DTC induction motor drives-a comparative study," Electric Power Systems Research, vol. 79, no. 1, pp. 210-219, 2009.
[5] Z. Sorchini and P. T. Krein, "Formal derivation of direct torque control for induction machines," IEEE Transactions on Power Electronics, vol. 21, no. 5, pp. 1428-1436, 2006.
[6] L. Yituo, L. Haifeng, Q. Wenlong et al., "A novel initial rotor position estimation method for permanent magnet synchronous motors," Proceedings of the Chinese Society for Electrical Engineering, vol. 33, no. 5, pp. 75-82, 2013.
[7] M. Oettmeier, M. Spichartz, V. Staudt, and A. Steimel, "Stator-flux-oriented control of PMSM in traction-Experimental results," in Electrical Systems for Aircraft, Railway and Ship Propulsion (ESARS '12), pp. 1-6, Bologna, Italy, 2012.
[8] Y. Yan, D. Yan, and C. Mingjun, "Research on application of PMSM step motion control technology in weapon location system," Journal of Gun Launch \& Control, no. 3, pp. 19-23, 2008.
[9] D. Yan, S. Hexu, B. Zhiyuan, and W. Wei, "Permanent magnet synchronous motor position control system based on torqueangle control," Transactions of China Electrotechnical Society, vol. 21, no. 1, pp. 86-91, 2006.
[10] V. Ambrožič, R. Fišer, and D. Nedeljković, "Direct current control-a new current regulation principle," IEEE Transactions on Power Electronics, vol. 18, no. 1, pp. 495-503, 2003.

## Research Article

# Cylinder Position Servo Control Based on Fuzzy PID 

Shibo Cai, Shunlei Wu, and Guanjun Bao<br>Key Laboratory of E\&M (Zhejiang University of Technology), Ministry of Education \& Zhejiang Province, Hangzhou 310032, China<br>Correspondence should be addressed to Guanjun Bao; gjbao@zjut.edu.cn

Received 25 July 2013; Revised 4 September 2013; Accepted 5 September 2013
Academic Editor: Tao Zou
Copyright © 2013 Shibo Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The arbitrary position control of cylinder has always been the hard challenge in pneumatic system. We try to develop a cylinder position servo control method by combining fuzzy PID with the theoretical model of the proportional valve-controlled cylinder system. The pressure differential equation of cylinder, pressure-flow equation of proportional valve, and moment equilibrium equation of cylinder are established. And the mathematical models of the cylinder driving system are linearized. Then fuzzy PID control algorithm is designed for the cylinder position control, including the detail analysis of fuzzy variables and domain, fuzzy logic rules, and defuzzification. The stability of the proposed fuzzy PID controller is theoretically proved according to the small gain theorem. Experiments for targets position of $250 \mathrm{~mm}, 300 \mathrm{~mm}$, and 350 mm were done and the results showed that the absolute error of the position control is less than 0.25 mm . And comparative experiment between fuzzy PID and classical PID verified the advantage of the proposed algorithm.


## 1. Introduction

In 1956, Shearer [1] first developed the pneumatic servo control system, using the high temperature and high pressure gas $\left(500^{\circ} \mathrm{C}, 20 \sim 30 \mathrm{MPa}\right.$ ) from the aerospace craft and missile propulsion as the working media. This pneumatic servo control system was successfully applied in the position, orientation, and stable flying control for aerospace crafts and missiles. In the subsequent period of time, efforts are contributed to investigate the pneumatic servo technology parallel with the hydraulic servo technique. But the early study made slow progress and there were few achievements that could be used, because of the difficulty in mathematic system models and lack of powerful analysis and calculating tools.

With the development of computer technology and modern control technique, the pneumatic servo control problem was revisited by scholars. Scavanda et al. [2] and Liu and Bobrow [3] broadened the linear model to several working points adopting the state-space method. But the influence of nonlinear factors such as mechanical friction is neglected. Baoren [4], Yunbo [5], and Guoliang et al. [6, 7] identified the system model based on experimental data, which can reflect the pneumatic system characteristics more accurately than the former methods. But it is not suitable for cases such as
long cylinder journey, large parametric variation, or heavy friction. Lee et al. [8] established a nonlinear model for pneumatic system and verified the model with experiments. Still, the model is complicated and requires rigid application conditions.

In this paper, we investigated a proportional valvecontrolled cylinder system and developed a position control method. Firstly, nonlinear mathematic model of the cylinder is established in Section 1. Then Section 2 gives the mathematic model of the whole pneumatic cylinder system. In Section 3, we designed a fuzzy PID controller for the proposed pneumatic position system, including all the detailed information. Experiments for different positions and comparison with classical PID were carried out, which are deeply discussed in Section 4. Finally, Section 5 summaries the main contribution and meaning of our work.

## 2. NonLinear Mathematic Model of Cylinder

The dynamic characteristics of cylinder are mainly described by three equations: the pressure differential equation of cylinder, pressure-flow equation of proportional valve and moment equilibrium equation of cylinder.

The flowing state of air inside the pneumatic system is extremely complicated. To simplify the system mathematical model, we use the following hypothesis.
(1) The working media (here refers to air) in the system is taken as ideal gas.
(2) The flowing state while the air runs through the valve port or other chokes is taken as the isentropic and adiabatic process.
(3) The lumped parameter model is adopted, ignoring the influences on the system from the distributed resistance in the air tube and flexibility of the pipeline.
(4) The air pressure and temperature inside the same chamber are equal everywhere.
(5) There is no leakage of the cylinder, both inside and outside.
(6) The pressures of air source and atmosphere are constant.
2.1. Pressure Differential Equation of the Cylinder. We suppose that the flowing air inside the thermodynamic system has no energy exchange with the outside and the pressure changes slightly, during the fast inflating process from air source to cylinder chamber. And then, this flowing process can be taken as the isentropic and adiabatic process. According to the energy equation of adiabatic inflating process from constant pressure air source to limited volume, there are four kinds of energy changing processes inside the volume during the movement [9].
(1) The air will bring in or take out the energy $q_{m} e$ itself during flowing in or out of the volume. Defining the internal energy of unit mass gas as $u$, kinetic energy as $v^{2} / 2$ and static energy as $g z$, we get

$$
\begin{equation*}
q_{m} e=q_{m}\left(\frac{u+v^{2}}{2+g z}\right) \tag{1}
\end{equation*}
$$

(2) The flowing work between the volume and the outside during the air runs in and out of the chamber is $\Delta W_{f}=q_{m} p v$, where $p$ is the air pressure and $v$ denotes the air specific volume.
(3) The thermoexchange between the chamber and the outside is $\Delta Q$.
(4) The work from the chamber to the outside during the piston movement is $\Delta W=p \Delta V$.

If we ignore the leakage of cylinder and valve, according to the energy conservation principle, the total internal energy $E$ of the chamber is

$$
\begin{align*}
& \frac{\mathrm{d} E_{1}}{\mathrm{~d} t}=q_{m 1} e_{1}+q_{m 1} p_{1} v_{1}+\frac{\mathrm{d} Q_{1}}{\mathrm{~d} t}-\frac{\mathrm{d} W_{s 1}}{\mathrm{~d} t} \\
& \frac{\mathrm{~d} E_{2}}{\mathrm{~d} t}=q_{m 2} e_{2}+q_{m 2} p_{2} v_{2}+\frac{\mathrm{d} Q_{2}}{\mathrm{~d} t}-\frac{\mathrm{d} W_{s 2}}{\mathrm{~d} t} \tag{2}
\end{align*}
$$

Supposing that the gas is ideal air and disregarding the kinetic energy and static energy of the air, we can get

$$
\begin{equation*}
q_{m} e+q_{m} p v=q_{m}(u+p v)=q_{m} h \tag{3}
\end{equation*}
$$

where $h$ is the specific enthalpy of air, $h=C_{p} T_{s}, C_{p}$ is the constant-pressure specific heat, and $T_{s}$ is the air temperature at the valve port.

As is well known, the internal energy of air is $E=m C_{v} T$, where $C_{v}$ is the constant-volume specific heat. According to the ideal air state equations, we have $m T=p v / R$, where $R$ is the gas constant, with the value of $287.1 \mathrm{j} /(\mathrm{kg} * \mathrm{k})$ and $R=C_{p}-$ C ${ }^{\text {. }}$

Substituting the above equations by formula (2), we can get

$$
\begin{align*}
& \frac{\mathrm{d} p_{1}}{\mathrm{~d} t}=R \frac{C_{p}}{C_{v}} T_{s} \frac{q_{m 1}}{V_{1}}-\frac{C_{p}}{C_{v}} \frac{p_{1}}{V_{1}} \frac{\mathrm{~d} V_{1}}{\mathrm{~d} t}+\frac{R}{C_{v} V_{1}} \frac{\mathrm{~d} Q_{1}}{\mathrm{~d} t} \\
& \frac{\mathrm{~d} p_{2}}{\mathrm{~d} t}=R \frac{C_{p}}{C_{v}} T_{s} \frac{q_{m 2}}{V_{2}}-\frac{C_{p}}{C_{v}} \frac{p_{2}}{V_{2}} \frac{\mathrm{~d} V_{2}}{\mathrm{~d} t}+\frac{R}{C_{v} V_{2}} \frac{\mathrm{~d} Q_{2}}{\mathrm{~d} t} \tag{4}
\end{align*}
$$

Generally, the rate of heat exchange $\mathrm{d} Q / \mathrm{d} t$ is determined by the temperature difference between the inside and outside of the cylinder and the coefficient of heat conduction of the cylinder block.
2.2. Pressure-Flow Equation of Proportional Valve. In the proportional valve-controlled cylinder system, the air mass flow running into and out of the cylinder chamber is controlled by the port area of the proportional valve. And the air mass flow $Q_{m}$ running through the valve port is determined by the effective port area of the valve $A_{e}$ and the upstream and downstream air pressure $P_{u}$ and $P_{d}$, that is,

$$
\begin{align*}
& Q_{m} \\
& = \begin{cases}A_{e} P_{u} \sqrt{\frac{2}{R T_{u}} \frac{k}{k-1}} \sqrt{\left(\frac{P_{d}}{P_{u}}\right)^{2 / k}-\left(\frac{P_{d}}{P_{u}}\right)^{(k+1) / k}} & 0.528<\frac{p_{d}}{P_{u}} \leq 1 \\
A_{e} P_{u} \sqrt{\frac{k}{R T_{u}}} \sqrt{\left(\frac{2}{k+1}\right)^{(k+1) / 2(k-1)}} & 0 \leq \frac{p_{d}}{P_{u}} \leq 0.528,\end{cases} \tag{5}
\end{align*}
$$

where $A_{e}$ is the effective port area of the valve, $\mathrm{m}^{2} ; T_{u}$ represents the stagnation temperature of the orifice upstream, $K ; Q_{m}$ denotes the air mass flow running through the valve port, $\mathrm{Kg} / \mathrm{s}$.
2.3. Force Equilibrium Equation of Cylinder. We can obtain the kinetic equilibrium between the cylinder and load by the force analysis for the system

$$
\begin{equation*}
A_{1} p_{1}-A_{2} p_{2}=m \frac{\mathrm{~d}^{2} y}{\mathrm{~d} t^{2}}+b \frac{\mathrm{~d} y}{\mathrm{~d} t}+F_{L}+F_{e} \operatorname{sign}(e) \tag{6}
\end{equation*}
$$

where $A_{1}$ and $A_{2}$ are the pressure working areas inside the two chambers of the cylinder, respectively; $m$ means the mass load; $b$ is the viscous damping coefficient between the piston and load; $F_{L}$ denotes the external load; $F_{e}$ represents the Coulomb friction and $e$ is the displacement deviation.


Figure 1: Cylinder position servo control diagram.

Combining the Coulomb friction and external load as $F$ and linearizing the force equilibrium equation, we can get

$$
\begin{equation*}
A_{1} p_{1}-A_{2} p_{2}=m s^{2} y+b s e+F \tag{7}
\end{equation*}
$$

## 3. Mathematic Model of the Pneumatic Position Servo System

From the above dynamic characteristics basical equations, it is clear that the system is nonlinear. So we linearize the system near the cylinder equilibrium point based on the linear system theory.

Generally, the spool opening area of proportional servo valve can be taken as the linear function of the controlling voltage; that is, the spool displacement is directly proportional to the controlling signal:

$$
\begin{equation*}
A_{e}=k_{a} u \tag{8}
\end{equation*}
$$

where $k_{a}$ is the voltage proportional coefficient.
Linearizing the flow equation of the proportional valve and applying the Laplace transform, we can get

$$
\begin{align*}
& Q_{m 1}=K_{q 1} u-K_{p 1} p_{1}  \tag{9}\\
& Q_{m 2}=K_{q 2} u-K_{p 2} p_{2}
\end{align*}
$$

where $K_{q 1}$ and $K_{q 2}$ are the flow gains at the working point of the controlling valves for the cylinder chambers, $K_{q}=$ $\partial q_{m} / \partial U ; K_{p 1}$ and $K_{p 2}$ are the flow pressure coefficients of the controlling valves for the cylinder chambers, $K_{P}=\partial q_{m} / \partial p$.

Linearizing the pressure differential equations of the cylinder chambers (2) and applying the Laplace transform, we can get

$$
\begin{align*}
& p_{1}=\frac{k R T q_{m 1}}{V_{k 1}} \frac{1}{s}-\frac{k p_{k 1} A_{1}}{V_{k 1}} e \\
& p_{2}=\frac{k R T q_{m 2}}{V_{k 2}} \frac{1}{s}-\frac{k p_{k 2} A_{2}}{V_{k 2}} e \tag{10}
\end{align*}
$$

The force equilibrium equation (6) can be transformed as

$$
\begin{equation*}
A_{1} p_{1}-A_{2} p_{2}=m s^{2} y+b s y+f e \tag{11}
\end{equation*}
$$

From the above analysis, the cylinder position servo control diagram can be drawn as Figure 1.

If $0 \leq u<5$, then

$$
\begin{gather*}
p_{1}=\frac{k R T K_{q 1} u}{V_{k 1} s+k R T K_{p 1}}-\frac{k p_{k 1} A_{1} e s}{V_{k 1} s+k R T K_{p 1}}, \\
p_{2}=-\frac{k p_{k 2} A_{2} e s}{V_{k 2} s+k R T K_{p 2}} . \tag{12}
\end{gather*}
$$

Substituting the above equations into (11) produces

$$
\begin{equation*}
e=\frac{\left(V_{k 2} s+k R T K_{p 2}\right) A_{1} k R T K_{q 1} u}{C} \tag{13}
\end{equation*}
$$

where

$$
\begin{align*}
C= & m V_{k 1} V_{k 2} s^{4} \\
& +\left(m V_{k 1} k R T K_{p 2}+m V_{k 2} k R T K_{p 1}+b V_{k 1} V_{k 2}\right) s^{3} \\
& +\left(m k^{2} R^{2} T^{2} K_{p 2} K_{p 1}+b V_{k 1} k R T K_{p 2}+b V_{k 2} k R T K_{p 1}\right) s^{2} \\
& +\left(f V_{k 1} V_{k 2}+k p_{k 1} A_{1}^{2} V_{k 2}+k p_{k 2} A_{2}^{2} V_{k 1}\right) s^{2} \\
& +\left(b k^{2} R^{2} T^{2} K_{p 1} K_{p 2}+f V_{k 1} k R T K_{p 2}+f V_{k 2} k R T K_{p 1}\right) s \\
& +\left(k^{2} p_{k 1} A_{1}^{2} R T K_{p 2}+k^{2} p_{k 2} A_{2}^{2} R T K_{p 1}\right) s . \tag{14}
\end{align*}
$$

## 4. Fuzzy PID Control Algorithm

PID algorithm is the most used and useful control technique in mechatronics system. But the classical PID algorithm has its inherent shortcomings in practice because of the fixed parameters. For example, the fixed parameters cannot take into account the dynamic features and control requirements in both transient process and stable period. It often fails to achieve the ideal integrated control quality. So, in practice, PID algorithm is usually combined with other parameter adjusting methods, such as fuzzy logic and artificial neuro network.


Figure 2: Fuzzy PID control principle.

We integrate the classical PID algorithm and fuzzy logic, using fuzzy logic to adjust the PID control parameters according to the deviation and its gradient between the output and target. Thus we can control the cylinder position precisely. The basical control principle is shown in Figure 2.
4.1. Fuzzy Variables and Their Domain. The PID control input is

$$
\begin{equation*}
e(t)=r(t)-y(t) . \tag{15}
\end{equation*}
$$

And the output of the control module can be written as

$$
\begin{equation*}
u(t)=K_{P} e(t)+K_{I} \int e(t) \mathrm{d} t+K_{D} \frac{\mathrm{~d} e(t)}{\mathrm{d} t} \tag{16}
\end{equation*}
$$

The deviation $e\left(\right.$ dis $_{0}-$ dis) between the target position dis $_{0}$ and the actual displacement dis of the cylinder and its gradient $e c(\mathrm{~d} e / \mathrm{d} t)$ are the input variables for fuzzy logic. And the variations $\Delta K_{P}, \Delta K_{I}$, and $\Delta K_{D}$ of PID control parameters $K_{P}, K_{I}$, and $K_{D}$ are the output variables of the fuzzy logic. The cylinder position deviation $e$ and its gradient $e c$ are sampled and calculated in real time. And the output variables $\Delta K_{P}$, $\Delta K_{I}$, and $\Delta K_{D}$ are extracted from the fuzzy matrix table based on the fuzzy rules and reasoning. The PID control parameters are adjusted using $\Delta K_{P}, \Delta K_{I}$, and $\Delta K_{D}$, in order to realize the real-time dynamic control of the cylinder displacement. According to the cylinder position control requirement, the domain of the displacement deviation $e$ is set as $(-0.5,0.5)$, and the domain of the $e c$ is $(-0.1,0.1)$. The domains of $\Delta K_{P}$, $\Delta K_{I}$, and $\Delta K_{D}$ for PID parameters are ( $-1.2,1.2$ ), ( $-0.1,0.1$ ), and ( $-0.05,0.05$ ), respectively.
4.2. Fuzzy Logic Rules. The triangle membership function is adopted, and the membership function for $\Delta K_{P}$ is shown in Figure 3. The fuzzy logic rules are deduced, as listed in Tables 1, 2, and 3. In these tables, NB, NM, NS, ZO, PS, $\mathrm{PM}, \mathrm{PB}$ represent negative big, negative medium, negative small, zero, positive small, positive medium, and positive big, respectively.


Figure 3: Membership function for $\Delta K_{P}$.

Table 1: Fuzzy logic rule for $\Delta K_{P}$.

| $e c$ |  |  |  | $e$ |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | NB | NM | NS | ZO | PS | PM | PB |
| NB | PB | PB | PB | PM | PS | PS | ZO |
| NM | PB | PB | PM | PM | PS | ZO | NS |
| NS | PB | PM | PM | PS | PS | ZO | NS |
| ZO | PM | PM | PB | PS | PS | ZO | NM |
| PS | PM | PS | PS | ZO | ZO | NS | NM |
| PM | PS | PS | ZO | ZO | NS | NM | NB |
| PB | PS | ZO | ZO | NM | NS | NB | NB |

Table 2: Fuzzy logic rule for $\Delta K_{I}$.

| $e c$ |  |  |  | $e$ |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | NB | NM | NS | ZO | PS | PM | PB |
| NB | NB | NB | NM | NM | NS | NS | ZO |
| NM | NB | NM | NM | NS | NS | ZO | ZO |
| NS | NB | NM | NS | ZO | ZO | PS | PS |
| ZO | NM | NM | NS | ZO | PS | PM | PM |
| PS | NM | NS | ZO | PS | PS | PM | PB |
| PM | ZO | ZO | PS | PS | PM | PM | PB |
| PB | ZO | ZO | PS | PM | PM | PM | PB |

Using the above fuzzy logic rules, the PID control parameters can be adjusted as

$$
\begin{align*}
K_{P(n+1)} & =K_{P n}+\Delta K_{P} \\
K_{I(n+1)} & =K_{I n}+\Delta K_{I}  \tag{17}\\
K_{D(n+1)} & =K_{D n}+\Delta K_{D}
\end{align*}
$$

Table 3: Fuzzy logic rule for $\Delta K_{D}$.

| $e c$ |  |  |  | $e$ |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | NB | NM | NS | ZO | PS | PM | PB |
| NB | PS | NS | NB | NB | NM | NM | PS |
| NM | PS | NB | NB | NM | NM | NS | ZO |
| NS | ZO | NS | NM | NM | NS | NS | ZO |
| ZO | ZO | NS | NS | NS | NS | ZO | NS |
| PS | ZO | PM | PS | ZO | ZO | PS | ZO |
| PM | PS | PB | PS | PS | PB | PB | PS |
| PB | PB | PM | PM | PS | PB | PS | PS |

## Define

$$
\begin{align*}
R_{l} & =(e \text { and } e c) \longrightarrow K_{P} \\
& =\int_{e \times e c \times K_{P}} \frac{u(e) \Lambda u(e c) \Lambda u\left(K_{P n}\right)}{2}, \\
R_{m} & =(e \text { and } e c) \longrightarrow K_{I} \\
& =\int_{e \times e c \times K_{I}} \frac{u(e) \Lambda u(e c) \Lambda u\left(K_{I n}\right)}{2},  \tag{18}\\
R_{n} & =(e \text { and } e c) \longrightarrow K_{D} \\
& =\int_{e \times e c \times K_{D}} \frac{u(e) \Lambda u(e c) \Lambda u\left(K_{D n}\right)}{2},
\end{align*}
$$

where $l, m, n=1,2,3, \ldots, 25$.
Then the fuzzy relations of $K_{P}, K_{I}$, and $K_{D}$ are

$$
\begin{align*}
& R_{K_{P}}=\bigcup_{l=1}^{25} R_{l}, \\
& R_{K_{I}}=\bigcup_{m=1}^{25} R_{m},  \tag{19}\\
& R_{K_{D}}=\bigcup_{n=1}^{25} R_{n} .
\end{align*}
$$

4.3. Defuzzification. The outputs of the fuzzy logic rules are also fuzzy set. In practical digital control system, the parameters must be defuzzified, that is, converting the fuzzy set into exact values according to an appropriate algorithm.

We use conventional gravity center method to realize the defuzzification:

$$
\begin{equation*}
y^{*}=\frac{\int_{Y} y u_{c}(y) \mathrm{d} y}{\int_{Y} u_{c}(y) \mathrm{d} y} \tag{20}
\end{equation*}
$$

where $y^{*}$ is the center of the covered region by membership function $u_{c}(y)$ of fuzzy set $C$.

It is obvious that the calculating process needs certain time, which makes it difficult to be used in real-time control system. So, the calculating process is executed off-line in advance. Then the produced defuzzification decision tables are stored in the memory of the controller. In this way, the instantaneity of the control process can be enhanced.


Figure 4: Discrete-time Fuzzy PID controller.
4.4. Stability Analysis. Chen and Ying [10] theoretically proved the stability of nonlinear fuzzy PI controller, based on their previous work on fuzzy control theory [11]. After that, they continued to investigate the stability of nonlinear fuzzy PI $+D$ controller [12]. Their work offers a quite convenient and practical method to explore the stability of similar control algorithms.

As described in Section 2, the target cylinder system can be taken as a classical second order system. To interpret the stability of the proposed nonlinear system, we need to reconsider the fuzzy PID control principle shown in Figure 2, which can be rearranged as Figure 4 in discrete-time form, where $T$ is the sampling period, $T>0$. This diagram expresses the same meaning as Figure 2 and shows the simplified structure as a figure in [12].

The stability of the fuzzy PI controller and the fuzzy PD controller has been analyzed in [10, 13], respectively, according to the small gain theorem [14]. In our case, if we disconnect the fuzzy $D$ control component from Figure 4, we have the fuzzy PI control system, whose stability is completely proved in [10]. The stability conditions are as follows.

Theorem 1. A sufficient condition for the nonlinear fuzzy PI control system to be globally bounded-input and boundedoutput (BIBO) stable is that
(1) the given nonlinear system has a bounded norm (gain) $\|N\|<\infty$;
(2) the parameters of the fuzzy PI controller $K_{P}, K_{I}$, and $K_{u P I}$ satisfy

$$
\begin{equation*}
\frac{K_{u P I}\left(\gamma K_{P}+K_{I}\right) L}{T\left(2 L-K_{M}\right)}\|N\|<1 \tag{21}
\end{equation*}
$$

where $L$ is the domain boundary of fuzzy logic parameters, $\gamma=\max \{1, T\}$ and $K_{M}=\max \left\{K_{P} M_{P}, K_{I} M_{c}\right\}$, with $M_{P}=$ $\sup _{n \geq 0}|e(n T)|$ and $M_{c}=\sup _{n \geq 0}|e c(n T)| \leq(2 / T) M_{P}$.

In the same way, by disconnecting the fuzzy PI controller from Figure 4, we reduce the fuzzy PID control system to a simple fuzzy $D$ controller. This fuzzy $D$ control system is a special or simplified case of the fuzzy PD control system studied in [13], and hence its stability condition can be derived from that obtained in [13] by removing the fuzzy $P$ controller or just setting the output of fuzzy $P$ component as zero. The stability conditions can be derived as follows.


Figure 5: Equivalent closed-loop control system for the fuzzy PID controller.

Theorem 2. A sufficient condition for the fuzzy $D$ control system to be BIBO stable is that the given process has a bounded norm (gain) $\|N\|<\infty$ and the parameters of the fuzzy $D$ controller $K_{D}$ and $K_{u D}$ satisfy

$$
\begin{equation*}
\frac{\gamma K_{D} K_{u D}}{2 T\left(L-K_{D}\left(M_{D}+|r|\right)\right)}\|N\|<1 \tag{22}
\end{equation*}
$$

where $\gamma=\max \{1, L\}$.
Till now, we are sure that the fuzzy PI controller and fuzzy $D$ controller are stable according to Theorems 1 and 2, respectively. Then, we need to verify that the combined fuzzy PID controller is stable.

Again, the Fuzzy PID controller shown in Figure 2 can be redrawn as Figure 5. The fuzzy PID control systems shown in Figures 2, 4, and 5 are the same thing but in different forms, just for analysis convenience. In Figure 5(a), let the system model be denoted by $S_{1}$ and the fuzzy PID controller together be denoted by $S_{2}$, resulting in the new structure in Figure 5(b). Then, as discussed in [10, 13], we can obtain a sufficient condition for the BIBO stability of the overall fuzzy PID equivalent closed-loop control system from the bounds:

$$
\begin{align*}
& \left\|S_{1}\left(u_{\mathrm{PID}}\right)\right\| \leq M_{1}+L_{1}\left\|u_{\mathrm{PID}}\right\|, \\
& \left\|S_{2}\left(\left[\begin{array}{l}
y \\
e
\end{array}\right]\right)\right\| \leq M_{2}+L_{2}\left\|\left[\begin{array}{l}
y \\
e
\end{array}\right]\right\|, \tag{23}
\end{align*}
$$

where $M_{1}, M_{2}, L_{1}, L_{2}$ are constants, and $L_{1} L_{2}<1$.

## 5. Experiments and Analysis

5.1. Experimental System Design. The experimental system is composed of pneumatic servo control actuating mechanism, feedback units, loading module, and controller. The pneumatic servo control actuating mechanism is symmetrical cylinder system controlled by proportional flow valve. The feedback units include displacement transducer and the pressure


Figure 6: Pneumatic servo control system principle.
transducer for the cylinder chambers. The whole controller for the system includes industrial personal computer (shorted as IPC), A/D, and D/A board cards for data acquisition and output. The experimental system schematic diagram is shown in Figure 6 and the experimental platform is shown in Figure 7. The instruments used in the experiment are listed in Table 4.

The control software was developed based on MATLAB and LabVIEW. All the fuzzy logic and PID control algorithms were realized in MATLAB simulink toolbox and then compiled into real-time control program using RTW technique.

Table 4: Experimental instruments.

| Name | Model | Specification | Brand |
| :---: | :---: | :---: | :---: |
| Main cylinder | CA2WL40-500 | Ф32 mm, range: 500 mm | SMC |
| Flow proportional valve | MPYE-5-1/8-010B | Max flow: $700 \mathrm{~L} / \mathrm{min}$, response: 3 ms , lag: $0.3 \%$ | Festo |
| Pressure proportional valve | MPPE-5-1/8-010B | Max flow: $820 \mathrm{~L} / \mathrm{min}$, response: 3 ms , lag: $0.3 \%$ | Festo |
| Displacement transducer | MTS-500 | Range: 500 mm , resolution: 5 us, repeatability: $\pm 0.001 \%$ FS | MTS |
| Pressure transducer | JYB-KO-HVG | Accuracy: $0.25 \% \mathrm{FS}$, range: $0-1 \mathrm{Mpa}$, response: 30 ms , nonlinearity: $\pm 0.2 \%$ FS, repeatability: $\pm 0.1 \%$ FS | Kunlun Coast |
| Force transducer | BK-1 | Range: 1500 N, accuracy: $0.05 \%$ FS, nonlinearity: $0.05 \%$ FS, repeatability: $0.05 \%$ FS | Kunlun Coast |



Figure 7: Experimental system.

RTW is an important supplementary functional module for MATLAB graphic modeling and simulation module Simulink. Optimized, portable, and personalized codes can be directly generated from Simulink model with RTW tools. According to the specific target preparation, the generated codes can be compiled into program for a different rapid prototype real-time environment. RTW ensures us to focus on the model establishment and system design and release from the boring programming work. This kind of developing pattern is very suitable for laboratory experimental system design.

RTW technique has the following features: (1) it supports continuous, discrete, and hybrid time system, including conditioned executing system and nonvirtual system; (2) RTW seamlessly integrates the Run-Time Monitor with the realtime target, which provides an excellent signal monitor and parameters adjusting interface. The flow diagram of real-time control program developing using RTW technique is shown in Figure 8.

LabWindows/CVI is adopted to create the control program frame and user interface, shown in Figure 9.
5.2. Target Position Control Experiments. On the experimental platform, we set the target position of the cylinder as $250 \mathrm{~mm}, 300 \mathrm{~mm}$, and 350 mm , respectively. And the control results are shown in Figures 10, 11, and 12.

The rising times of the three experiments are $2.65 \mathrm{~s}, 4.3 \mathrm{~s}$, and 3.2 s , respectively, which indicates that long displacement


Figure 8: Working flow with RTW.

Table 5: Control errors of cylinder position (mm).

| Initial | Target | AE | RE |
| :--- | :---: | :---: | :---: |
| 100 | 250 | 0.2441 | $0.20 \%$ |
| 100 | 300 | 0.20 | $0.07 \%$ |
| 100 | 350 | 0.2441 | $0.09 \%$ |

AE represents absolute error and RE denotes relative error.
does not mean long corresponding time. During the motion, the proposed fuzzy PID controller can adjust the control parameters and change the behavior of the system to achieve the best performance. Also, the overshoots in Figures 10, 11, and 12 are $0.49 \mathrm{~mm}, 0.04 \mathrm{~mm}$, and 0 mm , respectively. Consulting the stable errors listed in Table 5, we can see that when the displacement becomes longer, the system hysteresis shows greater influence on the final error. To be more frank, long displacement has no overshoot but big negative error, while short displacement has big overshoot and positive error.

From the experimental data, three significant features can be drawn as follows.
(1) Dynamic quality: the proposed method has fuzzy logic virtues in the earlier stage of control that can actuate the cylinder to approximate the target position rapidly. And during the late stages of control, it has virtues of PID algorithm, which means that the PID parameters are adjusted to execute the cylinder to quickly reach the target position without overshoot.
(2) Stable quality: the analysis of stable error is listed in Table 5. From the error analysis, it can be seen that the proposed theoretical model, control method, and experimental system can guarantee that the absolute


Figure 9: LabVIEW control program diagram.


Figure 10: Response of target position 250 mm .
control error is around 0.24 mm . In addition, the error is independent of the target position. The robust of the control method is quite well.
(3) No creeping phenomenon: when the cylinder runs with quite low speed or stops in the middle, there will be creeping phenomenon because of the air pressure in both the chambers and friction. From the response data in Figures 10, 11, and 12, it can be concluded that the proposed method can control the cylinder to stay at any position without creeping phenomenon.
5.3. Compared with Classical PID. To show the advantages of the proposed cylinder position servo control method, an experiment was done to compare the classical PID controller and the developed one in this paper, with the target position 300 mm .


Figure 11: Response of target position 300 mm .

The stable state data and error data are shown in Figures 13 and 14. From the above two comparing data curves, it can be seen that the classical PID controller can achieve the destination, but has bigger error, error range, and overshoot, which are $0.78 \mathrm{~mm}, 0.25 \mathrm{~mm}$, and 0.78 mm , respectively. However, the proposed fuzzy PID controller has relative smaller error, error range, and overshoot, which are $0.20 \mathrm{~mm}, 0.24 \mathrm{~mm}$, and 0.04 mm , respectively.

## 6. Conclusions

(1) The nonlinear mathematical models of cylinder and its valve-control pneumatic system, that is, pressure differential equation, pressure-flow equation, and moment equilibrium equation, are proposed.
(2) The cylinder position servo controller based on the mathematical models and fuzzy PID algorithm is


Figure 12: Response of target position 350 mm .

Figure 13: Stable data of comparing experiment.
established and proved to be stable under specified conditions.
(3) Experimental results show that the absolute control error is less than 0.25 mm and the proposed fuzzy PID controller has better performance than classical PID. The dynamic and stable qualities of the controller are quite well.

## Conflict of Interests

The authors wish to confirm that there is no known conflict of interests associated with this paper and there is no conflict of interests for any of the authors.


Figure 14: Error of comparing experiment.

## Acknowledgments

This work is financially supported by the National Natural Science Fund of China (Grant no. 51275470), the open fund of Key Laboratory of E\&M (Zhejiang University of Technology), Ministry of Education \& Zhejiang Province (Grant no. 2011EM001), and fund of Zhejiang Educational Department (Grant no. Y201225592).

## References

[1] J. L. Shearer, "Study of pneumatic processes in the continuous control of motion with compressed air," Transactions of the ASME, vol. 2, pp. 233-242, 1956.
[2] S. Scavanda, A. Kellal, and E. Richard, "Linearized models for an electropneumatic cylinder servo valve system," in Proceedings of the 3rd International Conference in Advanced Robotics, pp. 13-15, Versailles, France, 1987.
[3] S. Liu and J. E. Bobrow, "Analysis of a pneumatic servo system and its application to a computer-controlled robot," Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME, vol. 110, no. 3, pp. 228-235, 1988.
[4] L. Baoren, Research on Digital Electronic/Pneumatic Pressure Control Valve and Electronic/Pneumatic Position Servo System, Harbin Institute of Technology, Harbin, China, 1995.
[5] D. Yunbo, Research on Pneumatic PWM Position Servo System, Harbin Institute of Technology, 1995.
[6] T. Guoliang, M. Wenjie, and W. Xuanyin, "Lab research for mechanism modeling of pneumatic servo system," Hydraulics Pneumatics and Seals, vol. 10, pp. 26-31, 1999.
[7] B. Wei, T. Guoliang, L. Bo, M. Deyuan, and Y. Yuefeng, "Modeling and characteristics investigation on a new pneumatic proportional pressure valve," Journal of Zhejiang University, vol. 46, no. 11, pp. 1953-1959, 2012.
[8] H. K. Lee, G. S. Choi, and G. H. Choi, "A study on tracking position control of pneumatic actuators," Mechatronics, vol. 12, no. 6, pp. 813-831, 2002.
[9] W. Haijiang, Design of Pneumatic System and Research on Its Position and Velocity Servo Control System, Xi'an Jiaotong University, 2004.
[10] G. Chen and H. Ying, "Stability analysis of nonlinear fuzzy PI control systems," in Proceedings of the 3rd International Conference on Fuzzy Logic Applications, pp. 128-133, Houston, Tex, USA, 1993.
[11] H. Ying, W. Siler, and J. J. Buckley, "Fuzzy control theory: a nonlinear case," Automatica, vol. 26, no. 3, pp. 513-520, 1990.
[12] D. Misir, H. A. Malki, and G. Chen, "Design and analysis of a fuzzy proportional-integral-derivative controller," Fuzzy Sets and Systems, vol. 79, no. 3, pp. 297-314, 1996.
[13] H. A. Malki, H. Li, and G. Chen, "New design and stability analysis of fuzzy proportional-derivative control systems," IEEE Transactions on Fuzzy Systems, vol. 2, no. 4, pp. 245-254, 1994.
[14] R. J. P. de Figueiredo and G. R. Chen, Nonlinear Feedback Control Systems: An Operator Theory Approach, Academic Press, New York, NY, USA, 1993.

## Research Article

# Parameter Matching Analysis of Hydraulic Hybrid Excavators Based on Dynamic Programming Algorithm 

Wei Shen, ${ }^{1}$ Jihai Jiang, ${ }^{1}$ Xiaoyu Su, ${ }^{2}$ and Hamid Reza Karimi ${ }^{3}$<br>${ }^{1}$ School of Mechatronics, Harbin Institute of Technology, Harbin 150080, China<br>${ }^{2}$ College of Automation, Harbin Engineering University, Harbin 150001, China<br>${ }^{3}$ Department of Engineering, Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Wei Shen; shenw@hit.edu.cn
Received 26 July 2013; Revised 2 September 2013; Accepted 3 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 Wei Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

In order to meet the energy saving requirement of the excavator, hybrid excavators are becoming the hot spot for researchers. The initial problem is to match the parameter of each component, because the system is tending to be more complicated due to the introduction of the accumulator. In this paper, firstly, a new architecture is presented which is hydraulic hybrid excavator based on common pressure rail combined switched function (HHES). Secondly, the general principle of dynamic programming algorithm (DPA) is explained. Then, the method by using DPA for parameter matching of HHES is described in detail. Furthermore, the DPA is translated into the M language for simulation. Finally, the calculation results are analyzed, and the optimal matching group is obtained.


## 1. Introduction

The demand for fuel efficient and low-emission hydraulic excavators has been increased due to the growing energy crisis and environmental deterioration recently. The appearance of hybrid excavator has the immense potential for reducing the fuel consumption, because it can eliminate the throttling loss theoretically and recover the braking or gravitational potential energy. Nevertheless, the system tends to be more complicated by introducing the hydraulic accumulator, which is used as another power source. The power flow is also changed due to the new power source and the recovery energy; hence, different parameters of the system units can result in different fuel consumption rate. It is important for improving the system efficiency and reducing the fuel rate of the hydraulic hybrid excavator by investigating the parameter matching method, which is also a good way to cut down the rated engine power and cost.

The parameter matching of power transmission system makes the parameters of the components in the system adjust to the working conditions by choosing the parameters of the components appropriately in the premise that
the system working correctly can guarantee the system in optimal working condition, and then the overall efficiency of the system is improved; the purpose of energy saving is reached [1-4]. Static matching is the main way in existing matching methods. In this method, the maximum values in the working process of all actuators are used to choose the parameters of components. However, the working characteristics of frequent and large-scale power changes when the excavator works, to some extent, lead to oversize of components. However, the excavator has characteristic that multiple actuators of the excavator act at the same time, so the working conditions and system dynamics under the condition of composite actions have to be considered to make various components work in high efficiency and reduce the fuel consumption of the engine. In the existing optimization algorithms, the Dynamic Programming Algorithm (DPA) can solve the optimizing problems of any complex systems in theory, so it has been widely used, but DPA algorithm is mainly used to solve the optimal trajectory of controlled variables to provide reference for designing suboptimal controller [5-9]. One of the earliest researchers in this regard is Filipi et al. [10], who proposed a design optimization process in


Figure 1: Example of DP.
two stages for a parallel hybrid medium truck. Then, Cross used this algorithm to extend the application in parameter matching [11]. In this work, the Hydraulic Hybrid Excavator based on CPR combined switched function (HHES) is investigated. CPR means Common Pressure Rail which is similar with the electric grid. It is divided into two lines including high and low pressure pipelines. All of the hydraulic actuators are connected with the two lines in parallel; it means that it is convenient to arrange the hydraulic components. Moreover, it not only eliminates the throttling loss in the theory aspect, but can also recover the braking or gravitational potential energy. Hence, applying this structure on the hydraulic excavator is a promising hydraulic architecture in the aspect of saving energy. However, because HHES is a new system, there are only a few relevant research papers published on parameter matching. In this paper, the optimal control principle based on DPA is first introduced to the parameters optimization matching research of HHES. The minimized engine fuel consumption in typical working condition is treated as the optimization goal. Considering the influence of the factors such as the efficiency of components and system dynamics, the minimum fuel consumption of various components parameters matching mode will be excavated most possibly by choosing a group of optimal parameters, and the method in this paper can guarantee that the fuel consumption of the different components parameters can be compared fairly without considering the influence of control method.

## 2. Basic Principle

2.1. Dynamic Program Algorithm Principle. DPA algorithm is an effective computing method combined with sorting decision method and optimization principle. In 1953, American mathematician Robert Bellman proposed the optimization principle in his writing "An optimal policy has the property that, whatever the initial state and optimal first decision may be, the remaining decisions constitute an optimal policy with regard to the state resulting from the first decision" [12]. According to this theory, the sorting decision can be applied in a complicated system, and "optimization procedure" is used at each level so as to achieve the overall optimization goal.

Now, the basic principle of sorting decision is simply illustrated by Figure 1. For Figure 1, numbers close to the connecting lines between two points are the distance of two
points. The red lines in Figure 1 show the trajectory between $A$ and $B$ :

$$
\begin{equation*}
J_{A B}^{*}=J_{A D}+J_{D B}^{*}, \tag{1}
\end{equation*}
$$

where $J_{A D}$ constitutes the initial control and $J_{D B}^{*}$ represents the shortest distance from $D$ to $B$. So we can calculate every possible route and compare to get the shortest distance. However, if the number of the points is large, it tends to be impossible to get the suitable result through the calculation process:

$$
\begin{equation*}
J_{D B}^{*}=\min \left(J_{D E B}, J_{D F B}, J_{D G B}, J_{D E B}, J_{D E F B}, J_{D G F B}\right), \tag{2}
\end{equation*}
$$

so we can get $J_{A B}^{*}=18$.
The application of optimization algorithm can reduce the number of trajectories to be considered, as shown in Figure 1. Taking the reverse calculation from point $B$ as an example, if optimal path passes state point $C$, the optimal path between $C$ and $B$ is from the above node to $B$ (the required time is $2+5=$ 7) instead of the path from the below node to $B$ (the required time is $6+6=12$ ), then the minimum cost and optimal path from this point to terminal point are determined. By repeating the calculation process to all stated points, the minimum costs and optimal paths for all state points can be calculated, and the optimal path of the whole process can be obtained until the calculation of point $A$ is finished. Because of the iteration method used in DPA, the main application background is for discrete system. For continuous system, it should be converted into discrete system, and the optimal solution can be solved after discretization.

For a given system, the system dynamics can be described as

$$
\begin{equation*}
\dot{X}_{k}=f\left(X_{k}, u_{k}, d_{k}\right), \tag{3}
\end{equation*}
$$

where $X$ is the state vector, $u$ is the control vector, $d$ is the disturbance vector, and the subscript $k$ is the time instant. Generally, to simplify the problem, the system dynamics can be described in a discrete domain; in other words, differential equations are replaced by difference equations:

$$
\begin{equation*}
X_{k+1}=X_{k}+f\left(X_{k}, u_{k}, d_{k}\right) \tag{4}
\end{equation*}
$$

Generalizing the principle of optimal control to discrete time systems results in [11],

$$
\begin{equation*}
C_{k N}^{*}\left(X_{k}, u_{k}\right)=J_{k, k+1}\left(X_{k}, u_{k}, d_{k}\right)+J_{k+1, N}^{*}(X(k+1)), \tag{5}
\end{equation*}
$$

where $C_{k N}^{*}$ is the minimum cost of operation from $k$ to $N$ for a specific state $x(k)$ and control $u(k)$. The minimum cost of operation for all combinations of control is calculated from

$$
\begin{equation*}
J_{k N}^{*}\left(X_{k}\right)=\min _{u(k)}\left[C_{k N}^{*}\left(X_{k}, u_{k}\right)\right] \tag{6}
\end{equation*}
$$

2.2. Hydraulic Hybrid Excavator Based on CPR Combined Switched Function. In CPR, the constant pressure variable pump and hydraulic accumulator constitute the high pressure line, and the low pressure line is connecting the oil tank directly. Multiple different loads connect in parallel between


Figure 2: Hydraulic hybrid excavator based on CPR combined switched function.
the two lines. The rotating loads can be controlled by regulating the displacement of hydraulic pump/motor, while the linear loads are actuated by hydraulic transformer because the hydraulic cylinders are hard to change displacement normally [13-17]. Since the system includes secondary components and accumulators, energy can be recovered when the actuator brakes or falls and then is stored in the accumulator. Hence, the excavator possesses two kinds of power source. The low fuel consumption can be obtained by using adopted appropriate control strategy. In this configuration, the former three fixed displacement motors, which are used for swinging and driving, respectively, should be replaced by three hydraulic pump/motors [18, 19]. However, the key component is not popular and expensive. We propose a new architecture which uses on-off valves to switch the hydraulic transformer control, and Figure 2 shows the schematic. The reason for this modification is the working condition of excavators, because the travel part and the arm cylinder or the bucket cylinder are not working at the same time. So the fixed displacement motors which are used for traveling in the original nonhybrid excavator can remain. There are travel 1 and arm cylinder in Group 1, and Group 2 includes travel 2 and bucket cylinder. Moreover, two sets of valves, in which there are four on-off valves, are used to switch the hydraulic transformer control motor or cylinder. Hence, not only the energy-saving characteristic is remained, but also the cost can be reduced because of the manipulation of the fixed displacement motor instead of variable displacement pump/motor. Furthermore, it is easier to modify based on the existing manufacture process.

## 3. Application of DPA for Hydraulic Hybrid Excavators

The purpose of this paper is to calculate the component parameter configuration that minimizes the fuel consumption in typical working condition of the excavator by DPA algorithm, and a 5 ton LS-control prototype is used as research object, and the existing components in prototype should be changed as less as possible to reduce the reform cost. The main components of the entire hydraulic

Table 1: Parameter names and their ranges.

| Parameter name | Unit | Range |
| :--- | :---: | :---: |
| Hydraulic accumulator $V_{0}$ | Volume (L) | $10 ; 16 ; 25 ; 40$ |
| Precharge pressure $P_{0}$ | Pressure (bar) | $50 ; 100 ; 150 ; 200 ; 250$ |
| Swing pump/motor $V_{2}$ | Volume (L) | $28 ; 40 ; 71$ |

system include constant pressure variable pump, hydraulic accumulator, and hydraulic transformer, and the actuators contain boom hydraulic cylinder, bucket hydraulic cylinder, arm hydraulic cylinder, swing motor, and travel motors. By using switch control principle, the actuators except for the quantitative swing motor are reserved, and the quantitative swing motor is replaced by variable pump/motor. Because of the limitation of current technical level, the hydraulic transformers have not been applied widely, and the displacement of hydraulic transformer is not a choice. In addition, the main pump of original system also has the function of electronically controlling variables, so it has been in use. Thus, Table 1 shows that the components parameters need to be optimized matching in the entire system.

Installation space of a 5 -ton excavator is limited, so the optional maximum volume of the hydraulic accumulator is determined as 40 L . The decision of swing motor mainly refers to the existing parameters of the hydraulic pump/motor. We need to know the relevant data of circulatory working condition when using DPA algorithm. In addition, we need to determine the state and controlled variables of the system, and the dynamic state equation also needs to be established.
3.1. Working Cycle. The standard working cycle is used for calculation. This cycle represents an excavator digging a load of dirt, rotating and releasing the load into a truck or onto a pile, and then returning to its initial position. It should be noticed that the travel part is not considered in this paper. This process is divided into four parts. Figure 3 shows the velocity of each actuator, respectively [20]. During the beginning part, the boom cylinder and the swing keep the position basically, but the arm cylinder and bucket cylinder move out to dig. Then, the boom cylinder extends, and the swing rotates to lift the dirt and prepare for dumping. Next, the bucket cylinder retracts to dump the dirt. Finally, the swing rotates back, and the boom cylinder retracts to go back to the initial status.
3.2. State Variables and Controls of the System. The critical state variables of the system can be selected by (7), and Table 2 shows the symbol and the meaning

$$
\begin{align*}
X=[ & n_{\mathrm{eng}}, n_{s}, p_{1 \_\mathrm{bm}}, p_{2 \_\mathrm{bm}}, p_{1 \_A}, p_{2 \_A}, p_{1 \_\mathrm{bk}}, p_{2 \_\mathrm{bk}}, p_{h},  \tag{7}\\
& \left.v_{\mathrm{bm}}, v_{A}, v_{\mathrm{bk}}\right] .
\end{align*}
$$

According to the DPA principle, if all the state variables in the state matrix we establish are unknown, then it is difficult to realize the optimization process because the calculation amount will increase rapidly [11]. Hence, according to


Figure 3: The velocity of the actuator during the working cycle.

Table 2: Meanings of the state variables.

| Symbol | Meaning | Unit |
| :--- | :--- | :---: |
| $n_{\text {eng }}$ | Engine speed | rpm |
| $n_{2}$ | Swing speed | rpm |
| $p_{1 \_ \text {bm }}$ | Boom cylinder bore side | Pa |
| $p_{1 \_A}$ | Arm cylinder bore side | Pa |
| $p_{1 \_b k}$ | Bucket cylinder bore side | Pa |
| $p_{h}$ | Pressure of high pressure pipe line in CPR | Pa |
| $v_{\text {bm }}$ | Boom cylinder speed | $\mathrm{m} / \mathrm{s}$ |
| $v_{A}$ | Arm cylinder speed | $\mathrm{m} / \mathrm{s}$ |
| $v_{\mathrm{bk}}$ | Bucket cylinder speed | $\mathrm{m} / \mathrm{s}$ |

the known working conditions, state variables can be divided into two categories, namely, state variables decided by working conditions and the optimal state trajectory calculated by DPA algorithm. Because there is no coupling relationship between the engine of HHE and the key state variables in system, the rotating speed of engine and the pressure of high pressure pipe line are selected to be the state variables for optimization. Some state variables are limited by working condition requirements; other state variables, such as the pressure between two chambers of actuators and the resultant torques (or resultant forces) of actuators calculated by the pressure between two chambers, can also be regarded as known in the calculation process

$$
\begin{equation*}
X=\left[n_{\text {eng }}, p_{h}\right] . \tag{8}
\end{equation*}
$$

In addition, the critical control of the system is

$$
\begin{equation*}
C=\left[u_{1}, \beta_{1}, \beta_{2}, \delta_{1}, \delta_{2}, \delta_{3}\right] . \tag{9}
\end{equation*}
$$

The controls can also be divided into two parts: one is being decided by the working cycle, and the other is the optimizing trajectory. In order to finish the working cycle, the torque and force requirement should be met. For instance, $\beta_{2}$ would
be decided during each step after the state variable $p_{h}$ is confirmed by the next equation:

$$
\begin{equation*}
\beta_{2}=\frac{2 \pi}{p_{h} \cdot V_{2}}\left(M_{r}+\operatorname{sign}\left(n_{2}\right) \cdot\left|M_{l}\right|\right), \tag{10}
\end{equation*}
$$

where $M_{r}$ is the requirement torque of the swing and $M_{l}$ is the torque loss.

Hence, the free controls are chosen as

$$
\begin{equation*}
X=\left[u_{1}, \beta_{1}\right] . \tag{11}
\end{equation*}
$$

3.3. Discretization of the System. After the state and controlled variables are determined, we need to ensure the scope of the state and controlled variables and perform the mesh generation. The rotating speed range of the engine is determined by the inherent curve of the original engine, and the maximum value of the high pressure pipe line is defined by the allowable maximum pressure 350 bar of the components. The interval of the engine rotating speed is 100 rpm , and the interval of the pressure in high pressure pipe line is 5 Mpa , both the range of the controlled variables $\mu_{1}, \beta_{1}$ being from 0 to 100. The grids are shown in Figure 4 [21].

Generally speaking, the more dense grids, the more accurate results, but the calculated amount will be greatly increased. The purpose of this paper is to obtain minimum fuel consumption in the same cycle. Dynamic performances of the variable displacement mechanism in pump have not been considered, so it is more reasonable to choose the similar time interval with the variable displacement mechanism, since the frequency of the variable displacement mechanism is 5 Hz , and dt is chosen as 0.2 s .
3.4. Optimizing Object. The fewest fuel consumption rate of the engine is the optimization objective for the hydraulic


Figure 4: Discretization of the system.
hybrid excavator deterministic dynamic programming simulation

$$
\begin{equation*}
J_{c}=\sum_{k=1}^{N} \dot{m}_{f}(k) \cdot \Delta t=\sum_{k=1}^{N} F\left(M_{\mathrm{eng}}, n_{e}\right) \cdot \Delta t . \tag{12}
\end{equation*}
$$

Due to the big difference among the different components, especially for the excavator which is used widely, we consider the cost combined with the object of optimal fuel consumption by using weight factor method,

$$
\begin{equation*}
F\left(V_{0}, V_{2}, p_{0}, p_{\max }\right)=\alpha_{1} \cdot \frac{J_{c}-J_{\min }}{J_{\max }-J_{\min }}+\alpha_{2} \cdot \frac{C_{c}-C_{\min }}{C_{\max }-C_{\min }}, \tag{13}
\end{equation*}
$$

where $C$ represents the additional cost for different components.

### 3.5. Equations of System Dynamics

3.5.1. Engine Dynamics. The engine dynamics is a complicated process. It is difficult to state the detailed procedure by using mathematical analysis, especially, how to model a model is not the object of this work. Hence, one effective method which is based on the experience data is adopted. It means that the main torque types such as friction torque and loss are obtained from the lookup table which is calculated from the exact speed and torque. For the HHEC, the only load torque of the engine is the torque of the main pump and the friction torque

$$
\begin{equation*}
\dot{n}_{\mathrm{eng}}=\frac{1}{J_{\mathrm{eng}}}\left[u_{1} \cdot M_{W O T}-M_{p}-M_{\mathrm{loss}}-M_{f}\right], \tag{14}
\end{equation*}
$$

where $M_{p}=\left(\left(p_{h} \cdot V_{1}\right) /(2 \cdot \pi)\right) \beta_{1}$ is the torque of the main pump, $M_{W O T}$ represents the maximum torque for different engine speed, $M_{\text {loss }}$ is the loss torque which is a lookup table by using the experimental dates, and $M_{f}$ is the friction torque.

A discrete difference equation is required by using DPA, so the continuous differential equations are approximated as

$$
\begin{align*}
& \Delta n_{\mathrm{eng}} \\
& =\frac{\Delta t}{J_{\mathrm{eng}}}\left[u_{1} \cdot M_{W O T}-\frac{p_{h} \cdot V_{1}}{2 \cdot \pi} \beta_{1}-f_{M_{\mathrm{loss}}}\left(n_{\mathrm{eng}}, p_{h}, \beta_{1}\right)-M_{f}\right] . \tag{15}
\end{align*}
$$

3.5.2. Pressure of the High Pressure Pipe. The pressure buildup equation describes the change of pressure in the system with respect to time.

Because all of the high pressure sides of components in CPR are connected together, every component flow rate should be considered to calculate the pressure change. In detail, the high pressure pipe contains a main pump, hydraulic accumulator, and the actuators which are depicted in Figure 5. The direction of the flow rate is defined by positive if coming from the component to the high pressure pipe and negative for the opposite direction. Then, the pressure is calculated by the following equation, and it is noticed that again travel motors are omitted in the part:

$$
\begin{equation*}
\dot{p}_{h}=\frac{Q_{1}-\sum_{i=1}^{3} Q_{\mathrm{HT} i}-Q_{2}+\sum_{i=1}^{3} Q_{A 2 i}-Q_{L}}{\left(1 / \beta_{e}\right)\left[\sum_{i=1}^{3} V_{i \_a}+V_{m \_a}+\sum_{i=1}^{3} A_{i 1} \cdot\left(H_{i \_\mathrm{sk}}-l_{i}\right)\right]+C_{\mathrm{accu}}}, \tag{16}
\end{equation*}
$$

where $i$ represents the index of each actuator, such as bucket, arm, and boom cylinders; $\sum_{i=1}^{3} V_{i \Omega}$ is the total capacity which includes each $A$ port of the HTs, every cylinder volume of the rod side, and the pipe line volume; the initial volume of the motor/pump is represented as $V_{m_{-} a}$; $H_{i_{-s k}}$ is the stroke of each cylinder and $l_{i}$ is the displacement of every cylinder; $Q_{p}$ represents the output flow rate of the main pump; $Q_{A 2 j}$ is the flow rate of the rod side of each cylinder; $Q_{2}$ is the flow rate which goes into the motor/pump and $Q_{L}$ is the total flow rate of leakage. $Q_{\mathrm{HT}_{-} i}$ is the flow rate that goes into the HT, respectively.


Figure 5: Schematic of the PHP.

Also, in the previous equation $C_{\text {accu }}$ is defined as the capacity of the accumulator which is the function [22]

$$
\begin{equation*}
C_{\mathrm{accu}}=\frac{V_{a}}{k}\left(\frac{P_{\mathrm{pre}}}{P_{h}^{k+1}}\right)^{1 / k} . \tag{17}
\end{equation*}
$$

Again, the discrete difference equation is as follows:

$$
\begin{equation*}
\Delta p_{h}=\frac{\Delta t \cdot\left(Q_{p}-\sum_{i=1}^{3} Q_{\mathrm{HT}_{\_} i}-Q_{2}+\sum_{i=1}^{3} Q_{A 2 i}-Q_{L}\right)}{\left(1 / \beta_{e}\right)\left[\sum_{i=1}^{3} V_{i \_a}+V_{m-a}+\sum_{i=1}^{3} A_{i 1} \cdot\left(H_{i \_\mathrm{sk}}-l_{i}\right)\right]+C_{\mathrm{accu}}}, \tag{18}
\end{equation*}
$$

where $Q_{A 2 \_i}$ is confirmed by the working cycle which equals the velocity times to the area of the rod side for each cylinder. However, the way to calculate $Q_{\mathrm{HT} i}$ should be pointed out. The SHT of boom cylinder is chosen to show the process. The method for the other two HTs is the same.

The boom cylinder is controlled by regulating the port plate angle of the HT in HHEC. Firstly, we define the transformer ratio, and the next equation is [23]

$$
\begin{align*}
\lambda= & \frac{p_{B}}{p_{A}} \\
= & \left(-\sin \frac{\alpha}{2} \cdot \sin \delta-\frac{p_{T}}{p_{A}} \cdot \sin \frac{\gamma}{2} \cdot \sin \left(\delta+\frac{\alpha}{2}+\frac{\gamma}{2}\right)\right) \\
& \times\left(\sin \frac{\beta}{2} \cdot \sin \left(\delta-\frac{\alpha}{2}-\frac{\beta}{2}\right)\right)^{-1}  \tag{19}\\
= & \frac{F_{\text {net_bm }}+p_{h} \cdot A_{2}}{p_{h} \cdot A_{1}},
\end{align*}
$$

where $F_{\text {net_bm }}=p_{1 \_\mathrm{bm}} \cdot A_{1 \_\mathrm{bm}}-p_{2 \_\mathrm{bm}} \cdot A_{2 \_\mathrm{bm}}$ means the net force of the boom cylinder because all of the pressure and the area are known according to the cycle data.

Moreover, the flow rate of $A$ and $B$ can be obtained by

$$
\begin{align*}
Q_{\mathrm{HT} \_\mathrm{bm}}=q_{A}= & \frac{\omega_{\mathrm{HT}} \cdot V_{\mathrm{HT}}}{2 \pi} \cdot \sin \frac{\alpha}{2} \cdot \sin \delta+L_{i m}\left(p_{A}-p_{B}\right) \\
& +L_{i m}\left(p_{A}-p_{T}\right)+L_{e m} p_{A}, \\
Q_{A 1 \_\mathrm{bm}}=q_{B}= & \frac{\omega_{\mathrm{HT}} \cdot V_{\mathrm{HT}}}{2 \pi} \cdot \sin \frac{\beta}{2} \cdot \sin \left(\delta-\frac{\alpha}{2}-\frac{\beta}{2}\right) \\
& -L_{i m}\left(p_{B}-p_{A}\right)-L_{i m}\left(p_{B}-p_{T}\right)-L_{e m} p_{B} . \tag{20}
\end{align*}
$$

After considering the leakage coefficient in total,

$$
\begin{equation*}
\frac{Q_{A 1 \_b m}}{Q_{H T \_b m}}=\frac{q_{B}}{q_{A}}=\frac{\sin (\delta-(\alpha / 2)-(\beta / 2))}{\sin \delta}=-\lambda, \tag{21}
\end{equation*}
$$

where $Q_{A 1 \_b m}$ equals the velocity times to the area of the bore side for boom cylinder, and it is also the known data.
3.6. Programming. Figure 6 shows the whole flow chart of the program [21]. The program can be divided into three loops, in which the inner is the control loop and the middle is the state loop; the outside ones are the district layers which are divided by district time dt. Then, every state in per layer should be calculated by using all of the controls through the dynamic equations. During the calculation, the control values result in the result which exceeds the state domain that should be abandoned, and the calculation should go on by using the next control values. For those accepted controls, the fuel consumption for that state and the controls should be added. After comparing all the controls in that state, the minimum one is stored. The middle loop includes the same cycle for each state.

Figure 7 shows the process in detail. $N$ represents the step. The calculation begins from the end. In fact, the dynamic programming is one type of iterative algorithms. It begins from the end; hence, the initial value must be given. In this work, the initial value $J$ and $u$ are set to 0 . Some states are unavailable, which are represented by red rectangles. The black cycles represent the minimum fuel consumption values corresponding to those states, respectively. And the blue triangle means the optimal value in the step. All of the fuel consumption values (matrix $J$ ) in each step should be used as


Figure 6: Flow chart of the program.
the initial value for calculating in the next step. For example, the matrix $J_{1}$ is used for $N-1$ step. It should be noticed that when calculating the new state by using the controls, the values may not fit well in the mesh grid. Hence, the bilinear interpolation algorithm is introduced to calculate the fuel consumption for $J_{x y}$ [24]:

$$
\begin{align*}
J_{x y}= & {\left[J\left(p_{m}, n_{m-1}\right)-J\left(p_{m}, n_{m-1}\right)\right] \cdot p_{x} } \\
& +\left[J\left(p_{m-1}, n_{m}\right)-J\left(p_{m-1}, n_{m-1}\right)\right] \cdot n_{y} \\
& +\left[J\left(p_{m}, n_{m}\right)+J\left(p_{m-1}, n_{m-1}\right)-J\left(p_{m}, n_{m-1}\right)\right.  \tag{22}\\
& \left.\quad-J\left(p_{m-1}, n_{m}\right)\right] \\
& \cdot p_{x} \cdot n_{y}+J\left(p_{m-1}, n_{m-1}\right)
\end{align*}
$$

where $J\left(p_{m}, n_{m-1}\right) J\left(p_{m-1}, n_{m-1}\right), J\left(p_{m}, n_{m}\right)$, and $J\left(p_{m-1}, n_{m}\right)$ are the fuel consumption which are coming from the former results.

## 4. Simulation Results

There are 60 combinations of the three parameters in total. Hence, the simulation runs 60 times for each group of parameters. It takes about 5 hours once by using a single core computer. In order to eliminate the influence of the initial state, 5 cycles are input into the simulation, but only the middle three are used to compare the fuel consumption.

Figure 8 shows the relationship among $V_{0}, V_{2}$, and $P_{0}$. It can be found the general tendency, with the increment of $V_{0}$, the fuel consumption decreases. However, the fuel consumption reduces slowly after $V_{0}$ approaches $40 \mathrm{~L} . V_{2}$ is not independent from the other parameters, but it is coupled with $V_{0}$ and $p_{0}$. In general, the fuel consumption reduces with the increment of $V_{2}$, and it shows the similar tendency with $V_{0}$; that is the fuel consumption reduces slowly after $V_{0}$ approaches 40 L . Furthermore, the precharge pressure is a key variable to impact the fuel consumption. The optimal pressure value locates from 100 bar to 150 bar normally according to the simulation results.

In order to state it in detail, different fuel consumption values corresponding to different precharge pressure values of the 16 L accumulator are plotted in Figure 9, which shows that the minimum fuel consumption appears in 150 bar. All of the simulation results show the similar trend. This is because the energy storage reaches the maximum around this pressure level

$$
\begin{align*}
E & =-\int_{V_{i}}^{V_{f}} p d V=\frac{p_{0} V_{0}}{n-1}\left[\left(\frac{V}{V_{0}}\right)^{1-n}-1\right] \\
& =\frac{p_{0} V_{0}}{n-1}\left[\left(\frac{p_{0}}{p}\right)^{(1-n) / n}-1\right] \tag{23}
\end{align*}
$$

To get the precharge pressure which results in the maximum energy, the derivation of $E$ is calculated as follows:

$$
\begin{gather*}
\frac{d E}{d p_{0}}=\frac{V_{0}}{n-1}\left[\frac{1}{n}\left(\frac{p_{0}}{p}\right)^{(1-n) / n}-1\right]=0  \tag{24}\\
\frac{p}{p_{0}}=n^{n /(n-1)} \tag{25}
\end{gather*}
$$

It means that if the maximum pressure and the parameter $n$ are decided, then the optimal precharge pressure can be obtained from (25). Hence, the same accumulator under the optimal precharge pressure can store the maximum energy, and then the fuel consumption can be reduced.

In general, large components have low fuel consumption under the same condition. Because in this algorithm, the minimum engine fuel consumption is taken as the optimizing objective, so every state pursues the highest efficiency. However, for key components, such as the axial piston type component, the efficiency gets lower with the pressure increasing, then when we want to achieve the same torque, large components can reach the purpose of efficiency improvement in smaller pressure conditions; however, we need to take into account the price growth of complete machine. After the comprehensive comparison, a set of parameters we choose are $V_{2}=40 \mathrm{~mL} / \mathrm{r}, P_{0}=15 \mathrm{Mpa}$, and $V_{0}=16 \mathrm{~L}$.


Figure 7: The detailed calculation process of the program.


Figure 8: Fuel consumption with the changing accumulator volume and precharge pressure.


Figure 9: Fuel consumption with the different $p_{0}$ under the same $V_{0}=16 \mathrm{~L}$.

## 5. Conclusion

Optimal parameter matching results for HHES were analyzed with the aim of reducing the fuel consumption and modification cost. Firstly, a new architecture HHES is presented which not only keeps the advantages of the hydraulic hybrid excavator but also reduces the modification cost. Then, the DPA was applied in the matching process successfully. The results show that the fuel consumption reduces with the increment of the $V_{0}$. And the similar tendency is obtained for the swing pump/motor. However, it is coupled with $V_{0}$ and $p_{0}$. The precharge pressure shows the independent relationship for the fuel consumption among other parameters. The optimal value is located around $10 \sim 15 \mathrm{Mpa}$ under the conditions that the maximum pressure is 35 Mpa and $n$ is 1.25 . By combining the cost factor, the optimal group is obtained which is $V_{2}=40 \mathrm{~mL} / \mathrm{r}, P_{0}=15 \mathrm{Mpa}$, and $V_{0}=16 \mathrm{~L}$. The future work will focus on the optimal trajectory of the state variable based on the dynamic programming result firstly. Then, design the suboptimal control strategy according to the optimal trajectory and test it in the real excavator.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors acknowledge the contribution of National Natural Science Foundation of China (50875054 and 51275123) and Open fund of State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University (GZKF-2008003).

## References

[1] Q. Xiao and Q. Wang, "Parameter matching method for hybrid power system of hydraulic excavator," China Journal of Highway and Transport, vol. 21, no. 1, pp. 121-126, 2008.
[2] T. Li, D. Zhao, H. Kang, and Z. Zhang, "Parameter matching of parallel hybrid power loaders," Journal of Jilin University, vol. 42, no. 4, pp. 916-921, 2012.
[3] H. Li, H. Liu, H. Gao, and P. Shi, "Reliable fuzzy control for active suspension systems with actuator delay and fault," IEEE Transactions on Fuzzy Systems, vol. 20, no. 2, pp. 342-357, 2012.
[4] H. Li, J. Yu, C. Hilton, and H. Liu, "Adaptive sliding mode control for nonlinear active suspension systems using T-S fuzzy model," IEEE Transactions on Industrial Electronics, vol. 60, no. 8, pp. 3328-3338, 2013.
[5] S. M. Aceves, J. R. Smith, L. J. Perkins et al., "Optimization of CNG series hybrid concept vehicle," SAE Paper 960234, SAE, Warrendale, Pa, USA, 1996.
[6] B. Wu, C. C. Lin, Z. Filipi, H. Peng, and D. Assanis, "Optimal power management for a hydraulic hybrid delivery truck," Vehicle System Dynamics, vol. 42, no. 1-2, pp. 23-40, 2004.
[7] A. brahma, Y. Guezennec, and G. Rizzoni, "Optimal energy management in series hybrid electric vehicles," Proceedings of American Control Conference, vol. 1, no. 6, pp. 60-64, 2000.
[8] C. Lin, Z. Filipi, L. Louca, H. Peng, D. Assanis, and J. Stein, "Modelling and control of a medium-duty hybrid electric truck," International Journal of Heavy Vehicle Systems, vol. 11, no. 3-4, pp. 349-370, 2004.
[9] B. Wu, C. Lin, Z. Filipi et al., "Optimization of power management strategies for a hydraulic hybrid medium truck," in Proceedings of the 2002 Advanced Vehicle Control Conference, Hiroshima, Japan, 2002.
[10] Z. Filipi, L. Louca, B. Daran et al., "Combined optimisation of design and power management of the hydraulic hybrid propulsion system for the $6 \times 6$ medium truck," International Journal of Heavy Vehicle Systems, vol. 11, no. 3-4, pp. 372-402, 2004.
[11] M. Cross and Z. Filipi, "Series hydraulic hybrid propulsion for a light truck optimizing the thermostatic power management," in Proceedings of the 8th International Conference on Engines for Automobile, SAE Technical Paper 2007-24-0080, Capris, Italy, 2007.
[12] D. Feng, Study on power management strategies for a series hydraulic hybrid vehicle [Doctor thesis], University of Electronic Science and Technology of China, Chengdou, China, 2009.
[13] T. J. Kim, Optimal control: an effective method for designing hydraulic hybrid vehicles [M.S. thesis], Purdue University, West Lafayette, Ind, USA, 2011.
[14] W. Shen and J. Jiang, "Analysis and development of the hydraulic secondary regulation system based on the CPR," in Proceedings of theInternational Conference on Fluid Power and Mechatronics (FPM '11), pp. 117-122, August 2011.
[15] P. A. J. Achten, F. Zhao, and G. E. M. Vael, "Transforming future hydraulics: a new design of a hydraulic transformer," in Proceedings of the 5th Scandinavian International Conference on Fluid Power, Linköping, Sweden, May 1997.
[16] W. Shen and J. Jiang, "Analysis of energy recovery efficiency of hydraulic hybrid excavator," Journal of South China University of Technology, vol. 40, no. 1, pp. 82-87, 2012.
[17] K. K. Ahn, T. H. Ho, and Q. T. Dinh, "A study on energy saving potential of hydraulic on energy saving potential of hydraulic control system using switching type closed loop constant pressure system," in Proceedings of the 7th JFPS International Symposium on Fluid Power, pp. 317-322, 2008.
[18] W. Shen, J. Jiang, and H. R. Karimi, "Observer-based robust control for hydraulic velocity control system," Mathematical Problems in Engineering, vol. 2013, Article ID 689132, 9 pages, 2013.
[19] W. Shen and J. Jiang, "Study on control performance and energy recovery efficiency of hydraulic hybrid excavator," ICIC Express Letters, vol. 5, no. 12, pp. 4439-4444, 2011.
[20] J. Zimmerman and M. Ivantysynova, "The effect of system pressure level on the energy consumption of displacement controlled actuator systems," in Proceedings of the 5th FPNI PhD Symposium, pp. 77-92, Krakow, Poland, 2008.
[21] J. Zimmerman, Toward optimal multi-actuator displacement controlled mobile hydraulic system [Ph.D. thesis], School of Mechanical Engineering, Purdue University, West Lafayette, Ind, USA, 2012.
[22] J. Zimmerman, R. Hippalgaonkar, and M. Ivantysynova, "Optimal control for the series-paralleld displacement controlled hydraulic hybrid excavator," in Proceedings of the ASME 2011 Dynamic Systems and Control Conference, vol. October 2011, pp. 1-8.
[23] P. A. J. Achten, G. E. M. Vael, and F. Zhao, "The Innas hydraulic transformer-the key to the hydrostatic common pressure rail," SAE Paper 2000-01-2561, SAE, Warrendale, Pa, USA.
[24] L. Chen and C. Gao, "Fast discrete bilinear interpolation algorithm," Computer Engineering and Design, vol. 28, no. 15, pp. 3787-3790, 2007.

## Research Article

# Neural Network Predictive Control for Vanadium Redox Flow Battery 

Hai-Feng Shen, ${ }^{1}$ Xin-Jian Zhu, ${ }^{2}$ Meng Shao, ${ }^{2}$ and Hong-fei Cao ${ }^{2}$<br>${ }^{1}$ Automation Department, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China<br>${ }^{2}$ Institute of Fuel Cell, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China<br>Correspondence should be addressed to Hai-Feng Shen; hfshen999@sjtu.edu.cn

Received 6 July 2013; Revised 23 September 2013; Accepted 25 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 Hai-Feng Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The vanadium redox flow battery (VRB) is a nonlinear system with unknown dynamics and disturbances. The flowrate of the electrolyte is an important control mechanism in the operation of a VRB system. Too low or too high flowrate is unfavorable for the safety and performance of VRB. This paper presents a neural network predictive control scheme to enhance the overall performance of the battery. A radial basis function (RBF) network is employed to approximate the dynamics of the VRB system. The genetic algorithm (GA) is used to obtain the optimum initial values of the RBF network parameters. The gradient descent algorithm is used to optimize the objective function of the predictive controller. Compared with the constant flowrate, the simulation results show that the flowrate optimized by neural network predictive controller can increase the power delivered by the battery during the discharge and decrease the power consumed during the charge.


## 1. Introduction

Because of the energy crisis, utilization of renewable energy sources such as wind and solar energy for electric power supply has received more and more attention in recent years. However, the intermittent nature of most renewable energy makes it highly dependent on reliable and economical energy storage systems. All-vanadium redox flow battery (VRB) is a promising candidate for the storage of renewable energy. Compared with other redox batteries such as zinc bromine battery and lead acid battery, VRB has many attractive features, including long cycle life, high energy conversion efficiency, flexible design, and low cost [1]. Moreover, the problem of electrolytes cross-contamination is avoided by using the same element in both half cells. The potential applications of VRB include load leveling, uninterruptible power supply (UPS), and renewable energy storage [2]. Thus, it has good application and development prospects.

The flowrate of the electrolyte is an important control mechanism in the operation of a vanadium redox flow battery
system. At low flowrates, the electrolyte is provided insufficiently for the chemical reaction and stagnant regions can form in the electrode. The higher electrolyte flowrate will increase the VRB performance. But on the other hand, if the flowrate is too high, there is a risk of leakage, and the pump consumption will increase, which will reduce the system efficiency $[3,4]$. In order to enhance system efficiency, the optimal electrolyte flow rate should be determined.

Until recently, most researches are focused on the key materials of VRB, and there is little information available in the literature about the optimization of the electrolyte flowrate. An optimal strategy of electrolyte flowrate is proposed in [3] to improve the system efficiency and keep the high capacity simultaneously. At the beginning of the charge/ discharge process, VRB operates at the lower flowrate, and then increases to higher flowrate when the voltage increases/ decrease to certain value. Energy efficiency, system efficiency, and capacity at different operating modes are compared and the optimal electrolyte flowrate is determined. A multiphysics model of the VRB is proposed in [5]. The battery power is
represented during the charge/discharge as a function of flow rate, states of charge (SOC), and the stack current. The optimal flow rates are obtained by maximizing the power delivered during the discharge and minimizing the power consumed during the charge. However, these optimal strategies suffer from a serious drawback in the form of deterioration in the performance when the system is operated under wide range operating conditions or subjected to disturbance. To overcome these drawbacks, controllers based on robust control techniques must have been used.

Model predictive control (MPC) is an application of optimal control theory. In model predictive control, process model is utilized to predict the future response of a plant. An optimal control sequence is determined by solving a finite horizon optimization problem online at each sampling instant and the first control in this sequence is applied to the plant [6]. Because of its ability to handle the multivariable/ nonlinear nature of the dynamics, constraints, and optimality in an integrated fashion [7], MPC technology can now be found in a wide variety of application areas including chemicals, food processing, automotive, and aerospace applications [8]. The performance of model predictive controller relies upon the accuracy of the model on which it is based. However, the VRB suffers aging, reactant crossover, and load disturbance that cause no well-known effects on the system dynamics; it is difficult to establish accurate mathematical model. Moreover, the mathematical model is too complex for online optimization, and a simpler model is therefore required. An attractive approach to tackle these problems is to use neural networks as nonlinear models of the dynamic behavior of the process [9]. This is because multilayer networks have a capability to learn and uniformly approximate nonlinear functions to a prospected accuracy [10].

In this paper, a nonlinear model predictive control scheme is proposed to maximize the power delivered by the battery during the discharge and minimize the power consumed during the charge.

## 2. VRB System Process Description

The VRB system consisted of two key elements: the cell stack, where electrochemical reaction occurred and the tanks of electrolytes, where energy is stored. The electrolytes were pumped from the tanks to the stack by a circulation system. A schematic diagram of a vanadium redox flow batter is given in Figure 1.

The main electrode reactions for the VRB are as follows:

$$
\begin{gather*}
\text { cathode: } \mathrm{V}^{3+}+\mathrm{e}^{-} \rightleftarrows \mathrm{V}^{2+}  \tag{1}\\
\text { anode: } \mathrm{VO}^{2+}+\mathrm{H}_{2} \mathrm{O} \rightleftarrows \mathrm{VO}_{2}^{+}+\mathrm{e}^{-}+2 \mathrm{H}^{+} \tag{2}
\end{gather*}
$$

A multiphysics model of a VRB system with 19 cells is introduced in [11], which is composed of the electrochemical model and the mechanical model.


Figure 1: A schematic diagram of a vanadium redox flow battery.
2.1. Electrochemical Model. The equilibrium potential of the individual cells can be approximated using the Nernst equation (assuming unit activity coefficients) as follows:

$$
\begin{equation*}
E_{\text {cell }}=E^{0}+\frac{R T}{F} \ln \left(\frac{C_{\mathrm{V}^{2+}} C_{\mathrm{V}^{5+}} C_{\mathrm{H}^{+}}^{2}}{C_{\mathrm{V}^{3+}} C_{\mathrm{V}^{4+}}}\right) \tag{3}
\end{equation*}
$$

where $E^{0}$ is the standard potential; $T$ is the cell temperature; $C_{i}$ is the molar concentration of species $i$ in the cells. For simplicity, they assuming that the concentration inside the cell and tank is uniform and the time delay of electrolyte flow is negligible, the concentration inside the cell and tank is given by [12]

$$
\begin{gather*}
\frac{d C_{i}}{d t}=\frac{b I(t)}{V_{\text {cell }} F}+\frac{Q(t)}{V_{\text {cell }}}\left(C_{\text {tank }_{i}}-C_{i}\right), \\
\frac{d C_{\text {tank }_{i}}}{d t}=\frac{1}{V_{\text {tank }}}\left(-V_{\text {cell }} \frac{d C_{i}}{d t}+\frac{b I(t)}{F}\right), \tag{4}
\end{gather*}
$$

where $C_{\text {tank }}$ is the concentration inside the tank, $V_{\text {cell }}$ is the volume of the cell, $V_{\text {tank }}$ is the volume of the tank, $I(t)$ is the current, $Q(t)$ is the electrolyte flowrate, and $b$ is a sign factor that depends on the considered vanadium species $i$ ( -1 for $\mathrm{V}^{2+}$ and $\mathrm{V}^{5+}$ ions and 1 for $\mathrm{V}^{3+}$ and $\mathrm{V}^{4+}$ ions).

The $\mathrm{H}^{+}$quantity in the catholyte increases by 1 M (after the migration) when 1 M of vanadium $\mathrm{V}^{5+}$ is produced. So, the $\mathrm{H}^{+}$concentration in the catholyte at any state of charge is

$$
\begin{equation*}
C_{\mathrm{H}^{+}}=C_{\mathrm{H}^{+} \text {, discharged }}+C_{\mathrm{VO}_{2}^{+}}, \tag{5}
\end{equation*}
$$

where $C_{\mathrm{H}^{+} \text {,discharged }}$ is the protons concentration when the electrolyte is completely discharged.

Assuming that each individual cell composing the stack has the same charging characteristics, the equilibrium voltage $U_{\text {eq }}$ of stack can be written as follows:

$$
\begin{equation*}
U_{\text {eq }}=E_{\text {cell }} \cdot N_{\text {cell }}, \tag{6}
\end{equation*}
$$

where $N_{\text {cell }}$ is the number of cells.
The stack voltage $U_{\text {stack }}$ is decreased when current flows through the stack because of several types of internal losses, such as activation, concentration, and Ohmic losses. But these internal losses are difficult to measure; here, we replace them with equivalent resistance $R_{\text {eq, ch/disch }}$ :

$$
\begin{equation*}
U_{\mathrm{loss}}=\eta_{\mathrm{act}}+\eta_{\mathrm{conc}}+\eta_{\mathrm{ohm}}=I \cdot R_{\mathrm{eq}, \mathrm{ch} / \mathrm{disch}} . \tag{7}
\end{equation*}
$$

So stack voltage $U_{\text {stack }}$ is given by

$$
\begin{equation*}
U_{\text {stack }}=U_{\text {eq }}-U_{\text {loss }} \tag{8}
\end{equation*}
$$

Then the power of stack can be calculated as

$$
\begin{equation*}
P_{\text {stake }}=U_{\text {stack }} \cdot I . \tag{9}
\end{equation*}
$$

2.2. Mechanical Model. The circulation system pumps the electrolytes from the tanks through the stack and back in the tanks. The power consumed by pumps is expressed as follows:

$$
\begin{equation*}
P_{\text {mech }}=2 \frac{\left(\Delta P_{\text {pipes }}+\Delta P_{\text {stack }}\right) Q(t)}{\eta_{\text {pump }}} \tag{10}
\end{equation*}
$$

where $\eta_{\text {pumps }}$ is the pump efficiency, $\Delta P_{\text {pipes }}$ is the pressure drop in the pipes which can be obtained from the extended Bernoulli equation. The pressure drop in the stack $\Delta P_{\text {stack }}$ is proportional to the flowrate $Q(t)$ :

$$
\begin{equation*}
\Delta P_{\text {stack }}=Q(t) \widetilde{R} \tag{11}
\end{equation*}
$$

where $\widetilde{R}$ is the hydraulic resistance obtained from FEM simulations [13].
2.3. Battery Power. In practice, $P_{\text {mech }}$ is provided from the external power source during the charge and from the stack during the discharge [5]. By convention, the stack current is defined as positive during the discharge and negative during the charge. Thus, the battery power $P_{\text {VRB }}$ is given by

$$
\begin{equation*}
P_{\mathrm{VRB}}=P_{\text {stake }}-P_{\text {mech }} . \tag{12}
\end{equation*}
$$

## 3. Design of Nonlinear Model Predictive Controllers

The schematic of the neural network predictive control (NNPC) system developed in this research is shown in Figure 2. The main steps of the NNPC algorithm are listed as follows.
(1) Measure the input and output of the VRB system.
(2) Use the previous calculated control inputs and the neural network identifier to compute the cost function.


Figure 2: Schematics of the NNPC system.
(3) Use the optimization algorithm to calculate a new control vector.
(4) Repeat steps (2) and (3) till the desired optimal result is achieved.
(5) Apply the first element of the control vector to the VRB system.
(6) Update the parameters of the NN with the new training set.
(7) Repeat steps (1)-(6) for each time step.
3.1. Predictive Model Based on RBF Neural Network. According to previous section, the battery power can be expressed as follows

$$
\begin{equation*}
P_{\mathrm{VRB}}=g(Q, I, T, t) . \tag{13}
\end{equation*}
$$

Suppose the stack current and temperature keep constant for a certain amount of time. So, there is only one control variable: the flowrate $Q$. The following NARX model can be used to represent the VRB system:

$$
\begin{equation*}
y(t)=f(y(t-1), y(t-2), u(t-1), u(t-2)) \tag{14}
\end{equation*}
$$

where $y$ is the battery power, $u$ is the flowrate, and $f(\cdot)$ is an unknown nonlinear function that needs to be identified. Radial basis function (RBF) networks having one hidden layer were proven to be universal approximator [14]. Because of the advantages of easy design and good generalization, a RBF network is used to identify the nonlinear function $f(\cdot)$ in this paper. The structure of the RBF network is shown in Figure 3.

A Gaussian function is used as the activation function. So at the hidden layer, the output of RBF unit $i$ is

$$
\begin{equation*}
\varphi_{i}(x)=\exp \left(-\frac{\left\|x-c_{i}\right\|^{2}}{2 \sigma_{i}^{2}}\right) \quad(i=1,2, \ldots, 5) \tag{15}
\end{equation*}
$$

where $x(t)=[y(t-1), y(t-2), u(t-1), u(t-2)]^{T}$ is the input of RBF network. $c_{i}$ and $\sigma_{i}$ are the center and width of the $i$ th unit, respectively.


Figure 3: The structure of the RBF network.

The network output is calculated by

$$
\begin{equation*}
\widehat{y}=\sum_{i=1}^{5} w_{i} \varphi_{i}(x) \tag{16}
\end{equation*}
$$

where $w_{i}$ is the weight value on the connection between RBF unit $i$ and network output. The one-step ahead prediction is given by

$$
\begin{equation*}
\widehat{y}(t+1)=f_{N N}(y(t), y(t-1), u(t), u(t-1)) . \tag{17}
\end{equation*}
$$

The $j$-step ahead prediction of the system's output is calculated by feeding back the model outputs (instead of the future system's outputs which do not exist) to the input nodes of the network [15].

Consider the following:

$$
\begin{array}{r}
\widehat{y}(t+j)=f_{N N}(\hat{y}(t+j-1), \hat{y}(t+j-2) \\
u(t+j-1), u(t+j-2)) . \tag{18}
\end{array}
$$

The computational burden of the optimization problem showed in next subsection increases with the complexity of RBF network structure. In order to simplify the RBF network structure and simultaneously ensure the approximation accuracy, in this study, genetic algorithm (GA) is adopted to obtain the optimum initial values of the RBF network parameters before training the RBF network. These parameters include the output weights, the centers, and widths of the hidden unit.
3.2. The Objective Function Optimization Algorithm. There are different forms of the objective function under different control requirements. In this study, our purpose is to maximize the power delivered by the battery during the discharge
and minimize the power consumed during the charge while ensuring the control signal is smooth. Noticing that the battery power is positive during the discharge and negative during the charge, the objective function is given as follows:

$$
\begin{equation*}
\min J(t)=-\sum_{j=1}^{n} \widehat{y}(t+j)+\frac{1}{2} \sum_{i=1}^{m} \lambda \Delta u^{2}(t+i-1) \tag{19}
\end{equation*}
$$

subject to constraints

$$
\begin{align*}
& u_{\min } \leq u(t+i-1) \leq u_{\max } \quad(i=1,2, \ldots, m), \\
& y_{\min } \leq \widehat{y}(t+j) \leq y_{\max } \quad(j=1,2, \ldots, n), \tag{20}
\end{align*}
$$

where $\Delta u(t+i-1)=u(t+i-1)-u(t+i-2), \lambda>0$ is weight coefficient, and $n$ and $m$ are the predictive horizon and control horizon, respectively. The vector of the control variables is obtained from the minimization of the objective function over the specified horizon. The control vector is available only within the control horizon and maintains constant afterward, that is, $u(t+i)=u(t+m-1)$ for $i=m, \ldots, n-1$. Only the first element of the optimized control sequence is implemented on the process.

Since the function $\varphi$ is nonlinear, an analytical solution of the objective function is not possible. Stochastic optimization algorithms such as genetic algorithm and simulated annealing suffer from the drawback of slow convergence, which make them not suitable for online control. Since the objective function surface is simple, the gradient based method is an appropriate choice. Based on the gradient based method, for a given iterative step $i$, the control vector can be calculated as follows:

$$
\begin{gather*}
u^{k}(t+i-1)=u^{k-1}(t+i-1)+\Delta u^{k}(t+i-1) \\
(i=1,2, \ldots, m), \\
\Delta u^{k}(t+i-1)=-\eta \frac{\partial J}{\partial u(t+i-1)}+\alpha \Delta u^{k-1}(t+i-1), \tag{21}
\end{gather*}
$$

where $\eta$ is the learning rate and $\alpha \Delta u^{k-1}(t+i-1)$ is referred to as the additional momentum term. The initial value of $u(t+$ $i-1)$ in the iteration at each sampling period is defined as

$$
\begin{equation*}
u^{0}(t+i-1)=u(t-1) \tag{22}
\end{equation*}
$$

Constraints on control sequence can be handled as follows: when any one of the $u(t+i)$ reaches its limit, this control input is then set to be equal to its limit [16].

The derivative of the objective function at time $t+h-1$, $h=1,2, \ldots, m$ can be written as follows:

$$
\begin{align*}
\frac{\partial J}{\partial u(t+h-1)}= & -\sum_{j=1}^{n} \frac{\partial \widehat{y}(t+j)}{\partial u(t+h-1)} \\
& +\sum_{i=0}^{m} \lambda \Delta u(t+i) \frac{\partial \Delta u(t+i)}{\partial u(t+h-1)} \tag{23}
\end{align*}
$$

Table 1: The characteristics of the VRB stack.

| Name | Value |
| :--- | :---: |
| Number of cells $N_{\text {cells }}$ | 19 |
| $R_{\text {charge }}$ | $0.037 \Omega$ |
| $R_{\text {discharge }}$ | $0.039 \Omega$ |
| Electrolyte vanadium concentration | $2 \mathrm{~mol} / \mathrm{L}$ |
| Initial $H^{+}$concentration | $5 \mathrm{~mol} / \mathrm{L}$ |
| Tank size $V_{\text {tk }}$ | 83 L |
| Flow resistance $\widetilde{R}$ | $14186843 \mathrm{~Pa} / \mathrm{m}^{3}$ |
| Cell temperature $T$ | 298 K |
| Standard potential $E^{0}$ | 1.255 V |

The partial derivative can be calculated by the chain rule:

$$
\begin{align*}
& \frac{\partial \hat{y}(t+j)}{\partial u(t+h-1)} \\
& = \begin{cases}\frac{\partial f_{N N}}{\partial x_{3}}, & j=h, \\
\frac{\partial f_{N N}}{\partial x_{1}} \frac{\partial x_{1}}{\partial u(t+h-1)} \\
+\frac{\partial f_{N N}}{\partial x_{4}}, & j=h+1, \\
\frac{\partial f_{N N}}{\partial x_{1}} & \frac{\partial x_{1}}{\partial u(t+h-1)} \\
+\frac{\partial f_{N N}}{\partial x_{2}} \frac{\partial x_{2}}{\partial u(t+h-1)}, & h+2 \leq j \leq n,\end{cases} \tag{24}
\end{align*}
$$

where $x(t+j)=[\hat{y}(t+j-1), \hat{y}(t+j-2), u(t+j-1)$, $u(t+j-2)]^{T}$ is the input vector at time $t+j$ :

$$
\begin{equation*}
\frac{\partial f_{N N}}{\partial x_{l}}=\sum_{i=1}^{5} w_{i} \exp \left(-\frac{\left\|x-c_{i}\right\|^{2}}{2 \sigma_{i}^{2}}\right) \frac{c_{i l}-x_{l}}{\sigma_{i}^{2}}, \quad l=1,2,3,4 \tag{25}
\end{equation*}
$$

where $x_{l}$ represents the network input vector of $y$ and $u$. $\partial \Delta u(t+i) / \partial u(t+h-1)$ can be given by

$$
\frac{\partial \Delta u(t+i-1)}{\partial u(t+h-1)}= \begin{cases}1, & i=h  \tag{26}\\ -1, & i=h+1 \\ 0, & \text { else }\end{cases}
$$

## 4. Simulation

To investigate the performance of the proposed controller, a 19 cells, $2.5 \mathrm{~kW}, 6 \mathrm{kWh}$ VRB is simulated. Its main characteristics are listed in Table 1 [5].
4.1. Identification. In order to reduce the online computing time, the RBF network was trained offline before being applied to online control. The multiphysics model developed in Section 2 was used for train data generation. An input-output data set to train the RBF network was obtained by randomly changing the manipulated variable, $Q$, within the range of $0.05-0.7$ and normalized between -1 and +1 . The sampling


## Physical model <br> _ RBF model

Figure 4: Physical model and RBF model outputs for battery power during the discharge at 100 A .
time is set as 5 s .1026 samples were used for the training, while 513 samples were used for validation. The initial values of the RBF network parameters were optimized by GA. After the optimum initial values were obtained, the LevenbergMarquardt algorithm was used as training algorithm to adjust the network parameters. Root mean square error (RMSE) was employed to evaluate the accuracy of RBF network model. The training was terminated after 500 iterations; the obtained value of RMSE is 1.6591 . Figure 4 shows the validation results. From the results of Figure 4, it can be observed that the RBF network can accurately represent the VRB dynamics.

The RBF network trained offline works well when there are no disturbances. However, it can not accurately represent the VRB dynamics when VRB system is subjected to uncertainty. So, the RBF network requires to train online to adapt with the change in the process. Newest 100 samples were used for training.
4.2. Control Results. Normally, in a charge-discharge cycle, the battery is charged at constant current, the battery SoC increases from $2.5 \%$ (discharged) to $97.5 \%$, and then it is discharged at constant current until it reached its initial SoC [11]. The predictive horizon and the control horizons for NMPC are chosen as 4 and 1 , respectively. The parameter $\lambda$ is set to 10000. The lower limit and upper limit of flowrate are $0.05 \mathrm{~L} / \mathrm{s}$ and $2 \mathrm{~L} / \mathrm{s}$, respectively. In normal working condition, the battery is charged/discharged at constant current. Assuming at $t=5000 \mathrm{~s}$, a disturbance on generator speed causes the charge current to change from 100 A to 95 A , and at $t=12000 \mathrm{~s}$, a load disturbance causes the discharge current to change from 100 A to 110 A . Figure 5 shows the battery power during a charge-discharge cycle when influenced by a series of step changes in stack current. The corresponding optimal flowrate that is shown in Figures 6 and 7 shows the comparison of battery power during a charge-discharge cycle at different flowrate. Compared with the battery power at $Q=0.3$, the average


Figure 5: Battery power during a charge-discharge cycle.


Figure 6: Optimal flowrate during a charge-discharge cycle.
power consumed during the charge at optimal flowrate decreased by 10.80 W , and the average power delivered by the battery during the discharge increased by 10.62 W .

## 5. Conclusions

The electrolyte flowrate of VRB system was optimized online using model predictive control based on artificial neural networks. An RBF network is built to predict the future battery power. In order to reduce the computational burden of the optimization problem, the hidden layer nodes were chosen as 5. The RBF network model was found to be valid for wide flowrate variation with random load disturbances. The gradient descent algorithm method is used to realize the optimization procedure. Simulation result at different flowrate indicates that the proposed controller can enhance the output power of battery during the discharge and reduce the operating cost during the charge. Future works will focus on control strategy for VRB and wind farm combined system.

## References

[1] K.-L. Huang, X.-G. Li, S.-Q. Liu, N. Tan, and L.-Q. Chen, "Research progress of vanadium redox flow battery for energy


Figure 7: Comparison of battery power at different flowrate.
storage in China," Renewable Energy, vol. 33, no. 2, pp. 186-192, 2008.
[2] M. Vynnycky, "Analysis of a model for the operation of a vanadium redox battery," Energy, vol. 36, no. 4, pp. 2242-2256, 2011.
[3] X. Ma, H. Zhang, C. Sun, Y. Zou, and T. Zhang, "An optimal strategy of electrolyte flow rate for vanadium redox flow battery," Journal of Power Sources, vol. 203, pp. 153-158, 2012.
[4] H. Al-Fetlawi, A. A. Shah, and F. C. Walsh, "Non-isothermal modelling of the all-vanadium redox flow battery," Electrochimica Acta, vol. 55, no. 1, pp. 78-89, 2009.
[5] C. Blanc and A. Rufer, "Optimization of the operating point of a vanadium redox flow battery," in Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE '09), pp. 2600-2605, San Jose, Calif, USA, September 2009.
[6] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: stability and optimality," Automatica, vol. 36, no. 6, pp. 789-814, 2000.
[7] T. Salsbury, P. Mhaskar, and S. Joe Qinc, "Predictive control methods to improve energy efficiency and reduce demand in buildings," Computers and Chemical Engineering, vol. 51, pp. 7785, 2013.
[8] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," Control Engineering Practice, vol. 11, no. 7, pp. 733-764, 2003.
[9] A. Draeger, S. Engell, and H. Ranke, "Model predictive control using neural networks," IEEE Control Systems Magazine, vol. 15, no. 5, pp. 61-66, 1995.
[10] M. T. Hagan and H. B. Demuth, "Neural networks for control," in Proceedings of the American Control Conference (ACC '99), pp. 1642-1656, June 1999.
[11] C. Blanc and A. Rufer, "Multiphysics and energetic modeling of a vanadium redox flow battery," in Proceedings of the IEEE International Conference on Sustainable Energy Technologies (ICSET '08), pp. 696-701, Singapore, November 2008.
[12] M. Li and T. Hikihara, "A coupled dynamical model of redox flow battery based on chemical reaction, fluid flow, and electrical circuit," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences A, vol. 91, no. 7, pp. 1741-1747, 2008.
[13] C. Blanc, Modeling of a Vanadium Redox Flow Battery Electricity Storage System, Ecole Polytechnique Fédérale de Lausanne, 2009.
[14] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," Neural Computation, vol. 3, no. 2, pp. 246-257, 1991.
[15] P. H. Sørensen, M. Nørgaard, O. Ravn, and N. K. Poulsen, "Implementation of neural network based non-linear predictive control," Neurocomputing, vol. 28, no. 1-3, pp. 37-51, 1999.
[16] C. Venkateswarlu and K. Venkat Rao, "Dynamic recurrent radial basis function network model predictive control of unstable nonlinear processes," Chemical Engineering Science, vol. 60, no. 23, pp. 6718-6732, 2005.

## Research Article

# Tree-Based Backtracking Orthogonal Matching Pursuit for Sparse Signal Reconstruction 

Yigang Cen, ${ }^{1}$ Fangfei Wang, ${ }^{1}$ Ruizhen Zhao, ${ }^{1}$ Lihong Cui, ${ }^{2}$ Lihui Cen, ${ }^{3}$ Zhenjiang Miao, ${ }^{1}$ and Yanming Cen ${ }^{4}$<br>${ }^{1}$ School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China<br>${ }^{2}$ Department of Mathematics, Beijing University of Chemical Technology, Beijing 100029, China<br>${ }^{3}$ School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China<br>${ }^{4}$ Polytechnic College, Guizhou Minzu University, Guiyang, Guizhou 550025, China

Correspondence should be addressed to Lihui Cen; lhcen@csu.edu.cn
Received 17 July 2013; Accepted 5 September 2013
Academic Editor: Dewei Li
Copyright © 2013 Yigang Cen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Compressed sensing (CS) is a theory which exploits the sparsity characteristic of the original signal in signal sampling and coding. By solving an optimization problem, the original sparse signal can be reconstructed accurately. In this paper, a new Tree-based Backtracking Orthogonal Matching Pursuit (TBOMP) algorithm is presented with the idea of the tree model in wavelet domain. The algorithm can convert the wavelet tree structure to the corresponding relations of candidate atoms without any prior information of signal sparsity. Thus, the atom selection process will be more structural and the search space can be narrowed. Moreover, according to the backtracking process, the previous chosen atoms' reliability can be detected and the unreliable atoms can be deleted at each iteration, which leads to an accurate reconstruction of the signal ultimately. Compared with other compressed sensing algorithms, simulation results show the proposed algorithm's superior performance to that of several other OMP-type algorithms.


## 1. Introduction

Compressive sensing (CS) $[1,2]$ aims to recover sparse or compressible signal with low amount of information and high probability. It breaks the traditional rule of Nyquist sampling theorem, which states that a signal's information is preserved if it is uniformly sampled at a rate at least two times faster than its Fourier bandwidth. By this state-of-the-art signal compression and processing theory, the signal sampling frequency, the cost of processing time, data storage, and transmission can be greatly reduced.

For a given orthogonal basis $\Psi=\left\{\psi_{1}, \ldots, \psi_{N}\right\}$, the signal $x \in R^{N \times 1}$ can be represented in terms of the coefficient vector $\alpha$ as

$$
\begin{equation*}
x=\sum_{k=1}^{N} \psi_{k} \alpha_{k}=\Psi \alpha . \tag{1}
\end{equation*}
$$

The corresponding inverse transformation is $\alpha=\Psi^{H} x$, $\Psi \Psi^{H}=\Psi^{H} \Psi=I$, and $\Psi \in C^{N \times N}$. Here, $I$ is the identity
matrix. We say that $x$ is $K$-sparse under the orthogonal basis $\Psi$ if only $K \ll N$ coefficients $\alpha_{k}$ of $x$ are nonzero.

Usually, the signal is not sparse but its coefficient can be considered to be sparse or compressible after some transformations, such as the wavelet transformation.

Suppose that a matrix $\Phi$ represents the $M \times N$ measurement matrix. Then $\alpha$ is accomplished by collecting a measurement vector $y$ of dimension $M$ with $M \ll N . y$ can be expressed as $y=\Phi \alpha$. Then, (1) becomes

$$
\begin{equation*}
y=\Phi \alpha=\Phi \Psi^{H} x \tag{2}
\end{equation*}
$$

$\Phi$ is called as the CS measurement matrix and its columns are called atoms. The matrix $\Phi$ is rank deficient and hence loses information in general. The CS reconstruction problem wishes to recover the coefficient vector $\alpha$ from the set of $M$ linear measurements $y$. Since $M<N$, the reconstruction of $\alpha$ from $y$ is generally ill-posed.

The two major algorithmic approaches to sparse recovery are methods based on $\left(l_{1}\right)$ minimization and iterative
methods (matching pursuits). We now briefly describe these methods, as follows.
1.1. $\left(l_{1}\right)$ Minimization. The sparse recovery of this approach can be stated as the problem of finding the sparsest signal $\alpha=$ $\Psi^{H} x$ with the given measurements $y$ :

$$
\begin{gather*}
\left(l_{0}\right): \min \left\|\Psi^{H} x\right\|_{l_{0}}  \tag{3}\\
\text { s.t. } y=\Phi \Psi^{H} x .
\end{gather*}
$$

Donoho and his associates advocated the principle that for some measurement matrices $\Phi$, the highly nonconvex combinatorial optimization problem $\left(l_{0}\right)$ should be equivalent to its convex relaxation:

$$
\begin{array}{ll}
\left(l_{1}\right): \min \left\|\Psi^{H} x\right\|_{l_{1}}  \tag{4}\\
\text { s.t. } y=\Phi \Psi^{H} x .
\end{array}
$$

Reference [3] showed that if the measurement matrix satisfies the restricted isometry property (RIP), then a $K$-sparse signal can be recovered exactly; that is,

$$
\begin{equation*}
\left(1-\delta_{K}\right)\|x\|_{2}^{2} \leq\|\Phi x\|_{2}^{2} \leq\left(1+\delta_{K}\right)\|x\|_{2}^{2} \tag{5}
\end{equation*}
$$

$\delta_{K}$ is called as the Restricted Isometry Constant of $\Phi$. It has been shown that $\left(l_{1}\right)$ minimization can recover a sparse signal exactly under various conditions on restricted isometry constants, see $[4,5]$. Then, the convex problem $\left(l_{1}\right)$ can be solved using method of convex and even linear programming.
1.2. Orthogonal Matching Pursuit (OMP). An alternative approach to sparse recovery is via iterative algorithms, which find the support of the $K$-sparse signal $\alpha$ progressively. Once $S=\operatorname{supp}(\alpha)$ is found correctly, it is easy to compute the signal $\alpha$ from its measurements $y$ as $\alpha=\left(\Phi_{S}\right)^{-1} y$, where $\Phi_{S}$ denotes the measurement matrix $\Phi$ restricted to columns indexed by $S$.

A basic iterative algorithm is Orthogonal Matching Pursuit (OMP) [6]. OMP recovers the support of $\alpha$, one index at a time, in $n$ steps. Under a hypothetical assumption that is an isometry, that is, the columns of $\Phi$ are orthogonal, the signal $\alpha$ can be exactly recovered from its measurements $y$ as $\alpha=$ $\Phi^{*} y$.

The problem is that the $M \times N$ matrix $\Phi$ is never an isometry in the interesting range where the number of measurements $M$ is smaller than the ambient dimension $N$. Even though the matrix is not an isometry, one can still use the notion of coherence in recovery of sparse signals. In that setting, greedy algorithms are used with incoherent dictionaries to recover such signals, see [7, 8]. In our setting, for the commonly used random matrices, one expects the columns to be approximately orthogonal, and the observation vector $\alpha=\Phi^{*} y$ to be a good approximation to the original signal $\alpha$.

Tropp and Gilbert [6] analyzed the performance of OMP for Gaussian measurement matrices $\Phi$; a similar result holds for general sub-gaussian matrices. They proved that, for every
fixed $K$-sparse $N$-dimensional signal $\alpha$ and a random Gaussian measurement matrix $\Phi$, OMP recovers (the support of) $\alpha$ from the measurements $y$ correctly with high probability, provided the number of measurements is $M \sim K \log N$.

The $\left(l_{1}\right)$-minimization method has the strongest known guarantees of sparse recovery. Once the measurement matrix $\Phi$ satisfies the Restricted Isometry Condition, this method works correctly for all sparse signals $\alpha$. $\left(l_{1}\right)$-minimization is based on linear programming, which has its advantages and disadvantages. One thinks of linear programming as a black box and any development of fast solvers will reduce the running time of the sparse recovery method. On the other hand, it is not very clear what this running time is, as there is no strongly polynomial time algorithm in linear programming yet. All known solvers take time polynomial not only in the dimension of the program $N$ but also on certain condition numbers of the program. While for some classes of random matrices the expected running time of linear programming solvers can be bounded, estimating condition numbers is hard for specific matrices. For example, there is no result yet showing that the Restricted Isometry Condition implies that the condition numbers of the corresponding linear program is polynomial in $N$.

OMP is quite fast, both theoretically and experimentally. It makes $n$ iterations, where each iteration amounts to a multiplication by a $N \times M$ matrix $\Phi^{*}$ (computing the observation vector $\alpha$ ) and solving a least squares problem in dimensions at most $M \times n$. This yields strongly polynomial running time. In practice, OMP is observed to perform faster and is easier to implement than $\left(l_{1}\right)$-minimization. For more details, see [6]. OMP is quite transparent; at each iteration, it selects a new coordinate from the support of the signal $\alpha$ in a very specific and natural way. In contrast, the known $\left(l_{1}\right)$-minimization solvers, such as the simplex method and interior point methods, compute a path toward the solution. However, the geometry of $\left(l_{1}\right)$ is clear, whereas the analysis of greedy algorithms can be difficult simply because they are iterative.

On the other hand, OMP has weaker guarantees of exact recovery. Unlike $\left(l_{1}\right)$-minimization, the guarantees of OMP are nonuniform: for each fixed sparse signal $\alpha$ and not for all signals, the algorithm performs correctly with high probability. Rauhut has shown that uniform guarantees for OMP are impossible for natural random measurement matrices [9].

Moreover, OMP's condition on measurement matrices given in [6] is more restrictive than the Restricted Isometry Condition. In particular, it is not known whether OMP succeeds in the important class of partial Fourier measurement matrices.

These open problems about OMP, first stated in [6] and often reverberated in the Compressed Sensing community, motivated the recent works on the modified OMP algorithms, such as the model-based Compressive Sensing [10], TreeBased Orthogonal Matching Pursuit [11], Compressive Sampling Matching Pursuit (CoSaMP) [12], Regularized Orthogonal Matching Pursuit (ROMP) [13], and BacktrackingBased Matching Pursuit (BAOMP) [14]. ROMP and CoSaMP require the sparsity level as an input parameter. However, in the most practical applications, this information may not be known before reconstruction. Although the sparsity level
is not required for the OMP and BAOMP, they do not use the characteristics of the sparse representation, such as the tree structure of wavelet transform. In this paper, a new Treebased Backtracking Orthogonal Matching Pursuit (TBOMP) algorithm is presented based on the tree model in wavelet domain. Our algorithm converts the wavelet tree structure to the corresponding relations of candidate atoms without the prior information of signal sparsity level. Also, combing with the backtracking algorithm, the unreliable atoms can be deleted. Compared with OMP, ROMP, and BAOMP algorithms, the atom selection process will be more traceable, normalizable, and structural.

## 2. Tree-Based Backtracking Orthogonal Matching Pursuit (TBOMP) Algorithm

In this section, we will first review the wavelet tree structure. Second, the proposed TBOMP algorithm will be presented in detail.
2.1. Wavelet Tree Structure. Consider a signal $x$ of length $N=$ $2^{L}$, after $L$-level wavelet transformations, the set of $K$-tree sparse signals is defined as

$$
\begin{equation*}
\Gamma_{k}=\left\{x=v_{L} v+\sum_{i=L}^{1} \sum_{j=1}^{2^{L-i}} \omega_{i, j} \psi_{i, j}:\left.\omega\right|_{\Omega^{C}}=0,|\Omega|=K\right\} \tag{6}
\end{equation*}
$$

where $\nu$ is the scaling function and $\psi_{i, j}$ is the wavelet function at scale $i$ and offset $j$. The wavelet transform consists of the scale coefficient $v_{L}$ and wavelet coefficients $\omega_{i, j}$ at scale $i, 1 \leq$ $i \leq L$, and position $j$, with $1 \leq j \leq 2^{L-i}$.

Suppose that $\alpha=\left[v_{L}, \omega_{L, 0}, \omega_{L-1,0}, \omega_{L-1,1}, \omega_{L-2,0}, \ldots\right]^{T}$ is the vector of the scaling and wavelet coefficients of $x$ with the maximum decomposition level $L$. Also, it is a set of wavelet coefficients $\Omega$ forms a connected subtree [10]. The set $\Omega$ defines a subspace of signals whose support is contained in $\Omega$, which means that all wavelet coefficients outside $\Omega$ are approximately zero. The nested structure of wavelet coefficients creates a parent/child relationship between wavelet coefficients at different scales. We say that $\omega_{i+1,\lfloor j / 2\rfloor}(\lfloor\cdot\rfloor$ denotes rounded down) is the parent of $\omega_{i, j}$. Also, $\omega_{i-1,2 j}$ and $\omega_{i-1,2 j+1}$ are the children of $\omega_{i, j}$. These relations can be expressed graphically by the wavelet coefficient tree in Figure 2(a). The relationship between the parent and child nodes is that the index value of the parent node in a level is $1 / 2$ times the index of the child node.

A kind of tree structure (greedy tree) was proposed in [15]. For the greedy tree, if a coefficient is significant then its child and all of its grandchildren are likely significant [11]. Figure 1 depicts two cases of greedy tree approximation. The number of each node is the wavelet coefficient modulus. Nodes not labeled depict zeros. In the first case, the wavelet coefficients decay monotonically along the tree branches toward the leaves. Suppose that the wavelet tree $\Omega$ containing $P$ wavelet coefficients; that is, $|\Omega|=P$. The $P$-term greedy tree approximation (here, we assume that $P=4$ ) can be proceeded in three steps: (1) find the $p, p \leq 4$ largest wavelet coefficient terms; (2) form the smallest connected rooted subtree
that contains all of these $p$ coefficients; and then (3) increase $p$ until $|\Omega|=4$.

Initializing $p=2$, two coefficients 10 and 8 will be found and will form a minimum, connected subtree $\Omega$. Gradually increase $p$ until $p=4$, the greedy tree approximation forms the connected rooted subtree $\Omega, 10-8-4-3$, with 4 nodes that maximize the sum of the wavelet coefficients in the subtree. This process was shown in Figure 1(a), the error is small. Another case was shown in Figure 1(b), when the wavelet coefficients do not decay monotonically along the tree branches toward the leaves, an isolated significant coefficient away from the root will be selected, either of its all ancestor coefficients. These ancestor coefficients may be very small, which will increase the approximation error. For example, initializing $p=2$, then two coefficients 10 and 8 will be found and the resulted subtree is $10-0-0-8$ with $p=|\Omega|=4$. Obviously, the error is large.

We can see that the process of greedy tree approximation is simple, but when the tree includes isolated large coefficients far from the tree root, the approximation error will be increased. Thus, backtracking is imposed to deleting the wrong nodes selected by the greedy tree. This will be illustrated in the Section 2.2.
2.2. Tree-Based Backtracking Orthogonal Matching Pursuit (TBOMP) Algorithm. Our proposed Tree-based Backtracking Orthogonal Matching Pursuit (TBOMP) is as follows.

Algorithm 1 (TBOMP).

## Symbol Description

$\omega$-wavelet high frequency coefficient vector;
$\widehat{\omega}$-reconstruction wavelet high frequency coefficient vector;
$A$-measurement matrix, $y=A \omega$;
$a_{i}$-the $i$ th column vector of $A, 1 \leq i \leq N$;
$\mu_{1}, \mu_{2}$-parameters of thresholds, $\mu_{1}, \mu_{2} \in[0,1]$;
$\Lambda_{n}$-index set, $\Lambda$ denotes the index set of all columns $\left\{a_{i}\right\}$ of matrix $A$;
$n_{\max }$ - number of maximum iterations allowed;
$\Gamma_{n}$-atom-deleting set in the $n$th iteration;
$C_{n}$-candidate set of the root atoms in the $n$th iteration;
$F_{n}$-family set that consists of the subtrees corresponding to the root nodes in $C_{n}$.
Initialization. $r_{0}=y$ (initial residual), $\Lambda_{0}=\emptyset, \Gamma_{0}=\emptyset$, and $C_{0}=\emptyset$.
Loop
(1) Initial selection: select the candidate set $C_{n}$ with absolute values of correlations satisfying:

$$
\begin{gather*}
\left|\left\langle r_{n-1}, a_{C_{n}}\right\rangle\right| \geq \mu_{1} \cdot \max _{i \in \Lambda_{n}}\left|\left\langle r_{n-1}, a_{i}\right\rangle\right|,  \tag{7}\\
\Lambda_{n}=\Lambda \backslash \Lambda_{n-1}
\end{gather*}
$$



Figure 1: Greedy tree search.


Figure 2: Wavelet tree structure.
(2) According to the 2-times relationship of wavelet tree node indices, find the wavelet tree rooted at each node in $C_{n}$. Then the family set $F_{n}$ consists of the atoms indexed by $C_{n}$ and all of their families can be found.
For example, assume that $C_{n}=\left\{c_{n}^{1}, c_{n}^{2}, \ldots, c_{n}^{\mathrm{Q}}\right\}$, then the wavelet subtrees rooted at $c_{n}^{1}, c_{n}^{2}, \ldots, c_{n}^{\mathrm{Q}}$ will be found, respectively, in this step. The index sets of these $Q$ trees are denoted as $F_{n}^{1}, F_{n}^{2}, \ldots, F_{n}^{\mathrm{Q}}$.
(3) Compute $\widehat{\omega}_{F_{n}^{q}}^{n}=\left(A_{F_{n}^{q}}^{H} A_{F_{n}^{q}}\right)^{-1} A_{F_{n}^{q}}^{H} y, 1 \leq q \leq Q$.
(4) Find $F_{n}^{\widetilde{q}}$ such that $\omega_{F_{n}^{\tilde{q}}}^{n}$ minimizing the residual as follows:

$$
\begin{equation*}
\widetilde{q}=\arg \min _{1 \leq q \leq Q}\left\|y-A_{F_{n}^{q}} \widehat{\omega}_{F_{n}^{q}}^{n}\right\|_{2} . \tag{8}
\end{equation*}
$$

(5) Select atom deleting index set $\Gamma_{n}$ satisfying

$$
\begin{equation*}
\left|\widehat{\omega}_{\Lambda_{n-1} \cup F_{n}^{\tilde{q}}}\right| \leq \mu_{2} \cdot \max \left|\widehat{\omega}_{F_{n}^{\bar{q}}}^{n}\right| \tag{9}
\end{equation*}
$$

(6) Set $\Lambda_{n}=\left\{\Lambda_{n-1} \cup F_{n}^{\tilde{q}}\right\} \backslash \Gamma_{n}, a_{\left\{i: i \in \Lambda_{n}\right\}}=\emptyset$, and update the residual as follows:

$$
\begin{equation*}
r_{n}=y-A_{\Lambda_{n}} \widehat{\omega}_{\Lambda_{n}}^{n} \tag{10}
\end{equation*}
$$

(7) If $\left\|r_{n}\right\|_{2}<\varepsilon$ or if $n=n_{\max }$, quit the iteration; otherwise, set $n=n+1$, go to step 1 .

End Loop.
Output. the estimated support set $\Lambda_{n}$ and the nonzero values $\widehat{\omega}_{\Lambda_{n}}=\left(A_{\Lambda_{n}}^{H} A_{\Lambda_{n}}\right)^{-1} A_{\Lambda_{n}}^{H} y$.

As seen in the above algorithm, we combined the characteristics of tree structure and the BAOMP algorithm. In the first step, TBOMP selects candidate set $C_{n}$ whose correlations
between the columns of $\Phi_{\Lambda_{n}}$ and the residual $r_{n-1}$ are not smaller than $\mu_{1} \cdot \max _{i \in \Lambda_{n}}\left|\left\langle r_{n-1}, a_{i}\right\rangle\right|, \Lambda_{n}=\Lambda \backslash \Lambda_{n-1}$. Here, the constant $\mu_{1}$ is used to adaptively decide how many atoms are chosen at each time. Then the atoms corresponding to the elements of $C_{n}$ are set as the root nodes of subtrees. As we mentioned in Section 2.1, due to the 2-times relationship between the indices of parent and child nodes, the subtree of each atom corresponding to an index in $C_{n}$ can be found to form the family set $F_{n}^{q}$, which consists of the indices of the family atoms in the $q$ th subtree. In the third step, least square method is applied to obtain the reconstruction wavelet high frequency coefficients $\widehat{\omega}_{F_{n}^{q}}^{n}$ corresponding to the atoms indexed by $F_{n}^{q}$.
Then the optimal subtree indexed by $F_{n}^{\widetilde{q}}$ will be selected according to step (4). In this step, there may exist insignificant atoms in $a_{F_{n}^{\tilde{q}}}$. This is because that we only simply applied the 2-times relationship discipline in the searching processing of subtrees. Thus, the backtracking deleting method is introduced in the algorithm to identify the true support set of $F_{n}^{\widetilde{q}}$. The backtracking deleting set $\Gamma_{n}$ consists of the indices corresponding to all the reconstructed coefficients satisfying (9). Then, the index set is updated by $\Lambda_{n}=\left\{\Lambda_{n-1} \cup F_{n}^{\tilde{q}}\right\} \backslash \Gamma_{n}$ at this iteration. According to the atoms corresponding to the indices in the set $\Lambda_{n}$, the reconstruction coefficients $\widehat{\omega}_{\Lambda_{n}}^{n}$ can be computed. Finally, update the residual by (10) and go to the next iteration. If $\left\|r_{n}\right\|_{2}<\varepsilon$ or $n=n_{\text {max }}$, quit the iteration.

In the TBOMP, the process of tree nodes selection was shown in Figure 2; the first step of the algorithm is to select candidate set $C_{n}$ by (7). For example, suppose that $C_{1}=$ $\left\{\omega_{L, 0}, \omega_{L-2,3}\right\}$ was chosen at the first iteration. The nodes of subtree (2) rooted at $\omega_{L, 0}$ and the family nodes rooted at $\omega_{L-2,3}$ are the significant coefficients needed to be found. According to the 2-times relationship of wavelet tree node indices and Figure 2(b), $\omega_{L-1,0}$ and $\omega_{L-2,0}$ are the child and grandchild nodes of $\omega_{L, 0}$. Thus, subtree (1) rooted at the node $\omega_{L, 0}$ will be found in the first iteration.


Figure 3: Reconstruction signal by TBOMP algorithm.


Figure 4: Reconstruction of TBOMP algorithm.

Now we assume that the subtree (1) is the optimal tree corresponding to $\omega_{L, 0}$. At the end of this iteration, the backtracking algorithm will remove the node $\omega_{L-2,0}$ according to step 5 of the TBOMP algorithm described above. In the remaining iteration, node $\omega_{L-2,1}$ will be choosen as the child node of $\omega_{L-1,0}$. Ultimately, subtree (2) will be found accurately. Analogously, the searching process of the subtree rooted at the node $\omega_{L-2,3}$ is the same, and it can be proceeded simultaneously.

These characteristics of tree structure provide a new way for the study of reconstruction algorithm. Thanks to the tree structure of wavelet coefficients, when the signal is sparsely represented by the wavelet transform, it also provides a clew for the selection of atoms in the reconstruction algorithm. This will greatly improve the reliability of the atom selection.

The coefficients of wavelet decomposition include lowfrequency coefficients and high-frequency coefficients (scaling coefficients and wavelet coefficients in $\alpha$ ). The more levels of wavelet decomposition, the less low-frequency coefficients, and more important information is reserved in the highfrequency coefficients. Compared with the high-frequency coefficients, the number of low-frequency coefficients are much less if the decomposition level is big enough. Since the low-frequency coefficients play an important role in the wavelet reconstruction, in our proposed algorithm, only the high-frequency coefficients are measured by measurement
matrix. For the reconstruction, we combine the reconstructed high-frequency coefficients $\widehat{\omega}$ and the unprocessed lowfrequency coefficients. Then the inverse wavelet transform is applied to obtain a reconstructed $\widehat{x}$ of the original signal $x$.

## 3. Simulation Results

In this section, several experiments will be given for the TBOMP algorithm. In the first experiment, the original signal $x$ is a one-dimensional blocks signal with length $N=256$. It was recovered from $M=64$ measurements by using the Gaussian random measurement matrix. The wavelet decomposition level is 4 and the wavelet function is Dbl. Figure 3 shows the reconstruction result of 7 th iterations by using the TBOMP algorithm.

In the first iteration of the TBOMP algorithm, according to the parent-child relations of wavelet tree, some unreliable atoms will be chosen, which leads to a wrong reconstruction result. As marked by the cycles in Figure 4(a). Then according to the backtracking deleting method, the wrong selected atoms can be deleted. After the second and the third iterations, some atoms are still not found. After the 7th iteration, the reconstruction result (Figure 3(c)) with TBOMP algorithm is exactly same as the original wavelet coefficients shown in Figure 3(b).


Figure 5: Reconstruction results of Doppler and HeavySine signal by TBOMP and OMP algorithms.

Similar results can be obtained for other signals. Reconstruction results of Doppler and Heavysine signals by using our TBOMP algorithm are shown in Figure 5. Here, we compared our reconstruction results with the classical OMP algorithm, $M / N=1 / 4$.

In the next experiment, we will compare the TBOMP with some popular algorithms such as OMP, ROMP, and BAOMP. Here, only the high frequency coefficients are measured; the low-frequency coefficients will not be processed [16]. The wavelet function is choosen as the "sym8" in MATLAB. The decomposition level is 5 for these four algorithms. Define $\mathrm{SNR}=20 \log _{10}(\operatorname{std}(x) / \operatorname{std}(\hat{x}-x))$, where std denotes
the standard deviation. Because of the randomness of the sensing matrix, numerical result at each time is different. Hereafter, we use the same sensing matrix in one experiment for these four algorithms.

We use the Bumps signal of length $N=2048$ and change the values of $M$ simultaneously in order to guarantee the same experiment condition. After 5 layers of wavelet decompositions, there are 64 low-pass coefficients in the 5th decomposition layer and total 1984 high-pass coefficients in the 5 decomposition layers. In order to obtain a fair comparison, in the Figure 6, the measurements number used in these four algorithms is $500-64=436$. For sake of simplicity, when


Figure 6: Comparison signal of TBOMP and BAOMP in time domain.


Figure 7: SNR comparison for different values of $M$.
we mention that $M$ measurements in the TBOMP, we means that $M$ is the sum of the low-pass coefficient number and the measurement number of the high-pass coefficients.

When $M=500$, the compression ratio is about $1 / 4$. The reconstruction results of Bumps signal of TBOMP and BAOMP are shown Figure 6. The SNR of TBOMP is about 1.8 dB higher than the BAOMP.

Since ROMP requires the sparsity level $K$ to be known for exact recovery, in the experiments, the best sparsity value $K$ of the wavelet coefficients can be estimated according to repeated experiments and then used in the simulations. Figure 7 shows the SNR comparison results for different values of $M$. The values of $M$ are selected as $200,500,800,1100$, and 1400 , respectively. For each $M$, we conduct the experiment 10 independent trials and calculate the average SNR. It is obviously that the reconstruction result of TBOMP algorithm is superior to others.

## 4. Conclusion

Sparse reconstruction algorithm is one of the three core problems (signal sparse representation, measurement matrix design, and reconstruction algorithm design) of CS. The existed sparse reconstruction algorithms such as ROMP and

CoSaMP algorithms employ the sparsity $K$ as the prior knowledge for exact recovery, which has many limitations for the realistic applications. However, although the sparsity level are not required for OMP and BAOMP algorithms, they do not use the characteristics of special sparse basis to improve the performance of the algorithms. In this paper, a new Treebased Backtracking Orthogonal Matching Pursuit (TBOMP) algorithm was proposed based on the tree model in wavelet domain. Our algorithm can convert the wavelet tree structures to the corresponding relations of candidate atoms without any prior information of signal sparsity level. Moreover, the unreliable atoms can be deleted according to the backtracking algorithm. Compared with other compressive sensing algorithms (OMP, ROMP, and BAOMP), the signal reconstruction results of TBOMP outperform the above mentioned CS algorithms.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grants nos. 61272028, 61104078, 61073079, and 61273274); the Fundamental Research Funds for the Central Universities of China (Grant no. 2013JBZ003); the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grants nos. 20110162120045 and 20120009110008); the Program for New Century Excellent Talents in University (Grant no. NCET-12-0768); the National Key Technology R\&D Program of China (Grant no. 2012BAH01F03); the National Basic Research (973) Program of China (Grant no. 2011CB302203); and the Foundation of Key Laboratory of System Control and Information Processing (Grant no. SCIP2011009).

## References

[1] E. Candès, "Compressive sampling," in Proceedings of the International Congress of Mathematicians, pp. 1433-1452, Madrid, Spain, 2006.
[2] D. L. Donoho, "Compressed sensing", IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306, 2006.
[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489-509, 2006.
[4] T. T. Cai and A. Zhang, "Sharp RIP bound for sparse signal and low-rank matrix recovery," Applied and Computational Harmonic Analysis, vol. 35, pp. 74-93, 2012.
[5] Y. G. Cen, R. Z. Zhao, Z. J. Miao, and L. H. Cen, "A new approach of conditions on $\delta_{2 s}(\phi)$ for $s$-sparse recovery," Science in China Series F, 2013.
[6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Transactions on Information Theory, vol. 53, no. 12, pp. 4655-4666, 2007.
[7] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Transactions on Information Theory, vol. 52, no. 1, pp. 6-18, 2006.
[8] D. L. Donoho, M. Elad, and V. N. Temlyakov, "On Lebesguetype inequalities for greedy approximation," Journal of Approximation Theory, vol. 147, no. 2, pp. 185-195, 2007.
[9] H. Rauhut, "On the impossibility of uniform sparse reconstruction using greedy methods," Sampling Theory in Signal and Image Processing, vol. 7, no. 2, pp. 197-215, 2008.
[10] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Modelbased compressive sensing," IEEE Transactions on Information Theory, vol. 56, no. 4, pp. 1982-2001, 2010.
[11] C. La and M. N. Do, "Tree-based orthogonal matching pursuit algorithm for signal reconstruction," in Proceedings of the IEEE International Conference on Image Processing (ICIP '06), pp. 1277-1280, October 2006.
[12] D. Needell and J. Tropp, "CoSaMP, iterative signal recovery from incomplete and inaccurate samples," Tech. Rep., California Institute of Technology, Pasadena, Calif, USA, 2008.
[13] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," Foundations of Computational Mathematics, vol. 9, no. 3, pp. 317-334, 2009.
[14] H. Huang and A. Makur, "Backtracking-based matching pursuit method for sparse signal reconstruction," IEEE Signal Processing Letters, vol. 18, no. 7, pp. 391-394, 2011.
[15] R. Baraniuk, "Optimal tree approximation with wavelets," in Proceedings of the Wavelet Applications in Signal and Image Processing VII, pp. 196-207, July 1999.
[16] Y.-G. Cen, X.-F. Chen, L.-H. Cen, and S.-M. Chen, "Compressed sensing based on the single layer wavelet transform for image processing," Journal on Communications, vol. 31, no. 8, pp. 5355, 2010 (Chinese).

# Nonlinear $H_{\infty}$ Optimal Control Scheme for an Underwater Vehicle with Regional Function Formulation 

Zool H. Ismail ${ }^{1}$ and Matthew W. Dunnigan ${ }^{2}$<br>${ }^{1}$ Centre for Artificial Intelligence \& Robotics, Universiti Teknologi Malaysia, Jalan Semarak, 54100 Kuala Lumpur, Malaysia<br>${ }^{2}$ School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

Correspondence should be addressed to Zool H. Ismail; zool@fke.utm.my
Received 19 July 2013; Accepted 19 September 2013
Academic Editor: Xianxia Zhang
Copyright © 2013 Z. H. Ismail and M. W. Dunnigan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

A conventional region control technique cannot meet the demands for an accurate tracking performance in view of its inability to accommodate highly nonlinear system dynamics, imprecise hydrodynamic coefficients, and external disturbances. In this paper, a robust technique is presented for an Autonomous Underwater Vehicle (AUV) with region tracking function. Within this control scheme, nonlinear $H_{\infty}$ and region based control schemes are used. A Lyapunov-like function is presented for stability analysis of the proposed control law. Numerical simulations are presented to demonstrate the performance of the proposed tracking control of the AUV. It is shown that the proposed control law is robust against parameter uncertainties, external disturbances, and nonlinearities and it leads to uniform ultimate boundedness of the region tracking error.


## 1. Introduction

A valuable robotic system for the ocean environment is known as an Autonomous Underwater Vehicle (AUV). It has been used for many years in the oil and gas industry to obtain detailed maps of the ocean floor as well as to supervise pipeline activities [1]. The ongoing research on AUVs has given attention to the improvement of navigation and tracking control schemes. The conventional control methodologies are not the most suitable choice and they cannot guarantee the required tracking performance since an underwater vehicle exhibits inherent highly nonlinear system dynamics, imprecise hydrodynamic coefficients, and external disturbances. On the other hand, sliding mode control, due to its robustness against modelling inaccuracies and external disturbances, has been demonstrated to be a very attractive approach to cope with these problems [2-6]. However, a well-known drawback of conventional sliding mode controllers is the chattering effect. Therefore, to overcome the undesired effects of the control chattering, the authors in $[7,8]$ proposed a saturation function rather than a sign function. This substitution can minimize or, when desired, even completely eliminate chattering, but the
trajectory tracking error is uniformly ultimately bounded (UUB), which in fact means that a steady-state error will always remain. In order to enhance the tracking performance inside the boundary layer, some adaptive strategy should be used for uncertainty/disturbance compensation.

Recently, a nonlinear $H_{\infty}$ optimal control scheme was adopted for an underwater robotic system as an external tracking control loop and a disturbance observer was used as an internal disturbance compensation loop [9, 10]. The resultant control obtained by combining these two controls is then derived. A brief review of $H_{\mathbf{\infty}}$ optimality control has been presented in [11]. Moreover, the disturbance observer [12] is chosen, so that the $L_{2}$-gain conditions of the nonlinear $H_{\mathbf{o}}$ optimal control are relaxed, the magnitude of extended disturbances is reduced, and the robustness of the resulting control is improved without increasing the control input beyond that of the nonlinear $H_{\infty}$ optimal control alone. By using this control, the underwater robotic system can successfully follow the given trajectories, even when uncertainties and disturbances exist. An adaptive region tracking control was presented in [13] for an AUV where a region is used rather than a point due to minimizing the control effort


Figure 1: An illustration of an underwater robotic system which performs tracking task in a spherical region.
to track the region. Note that the total potential energy of the desired region is a summation of the potential energy associated with each region. Inspired from [13], some related research works such as in $[14,15]$ have been carried out to ensure that the marine robotic systems can cope with the underwater conditions and missions.

In this paper, a nonlinear $H_{\infty}$ optimal control with region tracking function is proposed for an underwater vehicle. The proposed dynamic region control, where it is formulated in task space, aims to reduce the energy consumed by vehicle thrusters. Within the region function formulation, the controller activates and sends commands to the thrusters only when the AUV is outside the desired region, and hence it significantly reduces energy consumption. However, the disturbances such as ocean currents may pull the underwater vehicle out of its desired region. This is likely to occur when the AUV navigates near to the boundary as illustrated in Figure 1. Hence, a nonlinear $H_{\infty}$ optimal control is proposed in this paper to counteract this problem. The performance of conventional region tracking control and region function adopted with nonlinear $H_{\infty}$ optimal control law can be observed with respect to the existence of unidirectional and bounded ocean current. The rest of the paper is organized as follows: Section 2 describes the kinematic and dynamic properties of an AUV. In Section 3, the nonlinear $H_{\infty}$ optimal control with region function formulation is briefly explained. The stability analysis using a Lyapunov-like function is also given in this section. In Section 4, numerical simulation results are provided to demonstrate the performance of the proposed control. Finally, the paper is concluded with some remarks in Section 5.

## 2. Kinematic and Dynamic Model of an AUV

2.1. Kinematic Model. The relationship between inertial and body-fixed vehicle velocity can be described using the Jacobian matrix $J\left(\eta_{2}\right)$ in the following form:

$$
\left[\begin{array}{l}
\dot{\eta}_{1}  \tag{1}\\
\dot{\eta}_{2}
\end{array}\right]=\left[\begin{array}{cc}
J_{1}\left(\eta_{2}\right) & 0_{3 \times 3} \\
0_{3 \times 3} & J_{2}\left(\eta_{2}\right)
\end{array}\right]\left[\begin{array}{l}
v_{1} \\
v_{2}
\end{array}\right] \Longleftrightarrow \dot{\eta}=J\left(\eta_{2}\right) v,
$$

where $\eta_{1}=\left[\begin{array}{lll}x & y & z\end{array}\right]^{T} \in \mathbb{R}^{3}$ and $\eta_{2}=\left[\begin{array}{lll}\phi & \theta & \psi\end{array}\right]^{T} \epsilon$ $\mathbb{R}^{3}$ denote the position and the orientation of the vehicle,
respectively, expressed in the inertial fixed frame. $J_{1}$ and $J_{2}$ are the transformation matrices expressed in terms of the Euler angles. The linear and angular velocity vectors, $v_{1}=$ $\left[\begin{array}{lll}u & v & w\end{array}\right]^{T} \in \mathbb{R}^{3}$ and $v_{2}=\left[\begin{array}{ccc}p & q & r\end{array}\right]^{T} \in \mathbb{R}^{3}$, respectively, are described in terms of the body-fixed frame.
2.2. Dynamic Model. Let the velocity state vector with respect to the body-fixed frame be defined by $v \in \mathbb{R}^{6}$; and the underwater vehicle dynamic equation can be expressed in closed form as [16]

$$
\begin{equation*}
M \dot{v}+C(v) v+g(\eta)+F_{\mathrm{ext}}=\tau \tag{2}
\end{equation*}
$$

where $M$ and $C(v)$ represent the inertia matrix and the Coriolis and centripetal forces matrix including the effects of added mass and hydrodynamic damping by body motion and $g(\eta)$ is the restoring force. $F_{\text {ext }}$ contains the effects of external disturbances and the effects of added mass and hydrodynamic damping by body motion in static water. The dynamic (2) preserves the following properties [16, 17].

Property 1. The inertia matrix $M$ is symmetric and positive definite such that $M=M^{T}>0$ and $\gamma I \leq M \leq \Upsilon I$.

Property 2. $C(v)$ is the skew-symmetric matrix such that $C(v)=-C^{T}(v)$.

In the Property $1, \gamma$ and $\Upsilon$ denote the minimum and maximum eigenvalues of the inertia matrix, respectively. The matrix $I$ is the identity matrix that has suitable dimension.

## 3. Nonlinear $H_{\infty}$ Optimal Control Law with Region Formulation

In the region-based control framework, the desired moving target is specified by a region at the desired trajectory. A robust nonlinear $H_{\infty}$ optimal control for AUV proposed in this paper is formulated as follows.

First, the vehicle needs to converge into a region with specific shape. The objective function for this region is defined by the following:

$$
\begin{equation*}
f\left(\delta \eta_{B}\right) \leq 0 \tag{3}
\end{equation*}
$$

where $\delta \eta_{B}=B\left(\eta-\eta_{d}\right) \in \mathbb{R}^{6}$ are the continuous first partial derivatives of the dynamic region; $\eta_{d}(t)$ is the time-varying reference point inside the geometric shape and $B(t)$ is a timevarying and nonsingular scaling factor. It is assumed that $\eta_{d}(t)$ and $B(t)$ are bounded functions of time. To achieve the scaling formation, that is, if the scaling factor increases, then the size of a desired region also increases, a nonsingular matrix is defined as follows:

$$
B=\left[\begin{array}{cc}
B_{1} & 0  \tag{4}\\
0 & B_{2}
\end{array}\right]
$$

where $B_{1}$ is the scaling matrix of $\eta_{1}$ and $B_{2}$ is the scaling matrix of $\eta_{2}$. This function is useful when the AUV needs to adapt the moving region, depending on the situation and environment.

The corresponding potential energy function for the desired region described in (3) can be specified as

$$
P\left(\delta \eta_{B}\right)= \begin{cases}0, & f\left(\delta \eta_{B}\right) \leq 0  \tag{5}\\ \frac{k_{p}}{2} f^{2}\left(\delta \eta_{B}\right), & f\left(\delta \eta_{B}\right)>0\end{cases}
$$

such that

$$
\begin{equation*}
P\left(\delta \eta_{B}\right)=\frac{k_{p}}{2}\left[\max \left(0, f\left(\delta \eta_{B}\right)\right)\right]^{2} \tag{6}
\end{equation*}
$$

where $k_{p}$ is a positive scalar. Differentiating (6) with respect to $\delta \eta_{B}$ gives

$$
\begin{equation*}
\left(\frac{\partial P\left(\delta \eta_{B}\right)}{\partial \eta_{B}}\right)^{T}=k_{p} \max \left(0, f\left(\delta \eta_{B}\right)\right)\left(\frac{\partial f\left(\delta \eta_{B}\right)}{\partial \eta_{B}}\right)^{T} \tag{7}
\end{equation*}
$$

Now, let (7) be represented as the region error $\widetilde{e}_{B}$ in the following form:

$$
\begin{equation*}
\tilde{e}_{B}=\max \left(0, f\left(\delta \eta_{B}\right)\right)\left(\frac{\partial f\left(\delta \eta_{B}\right)}{\partial \eta_{B}}\right)^{T} \tag{8}
\end{equation*}
$$

If $B$ is set to an identity matrix, then a useful vector $v_{r}$ is defined as

$$
\begin{equation*}
v_{r}=J^{-1} \dot{\eta}_{d}-\alpha J^{-1} \widetilde{e}_{B}-\beta J^{-1} \int \tilde{e}_{B} d t \tag{9}
\end{equation*}
$$

where $\alpha$ and $\beta$ are arbitrary positive constants. The matrix $J^{-1}$ represents the inverse of the Jacobian matrix. From the arguments of trigonometric functions, this matrix is bounded. Based on the structure of (8) and (9) and the subsequent stability analysis, a filtered tracking error vector for an underwater vehicle is defined as

$$
\begin{equation*}
r(t)=v-J^{-1} \dot{\eta}_{d}+\alpha J^{-1} \widetilde{e}_{B}+\beta J^{-1} \int \widetilde{e}_{B} d t \tag{10}
\end{equation*}
$$

From the definition of $r$ in (10), the control law for an AUV can be proposed in the following form:

$$
\begin{equation*}
\tau=-K_{v} r+\widehat{M} \dot{v}_{r}+\widehat{C}(v) v_{r}+\widehat{g}(\eta), \tag{11}
\end{equation*}
$$

where $K_{v}=K+\left(1 / \kappa^{2}\right) I ; K$ and $I$ are the positive definite matrix and identity matrix, respectively. $\widehat{M}, \widehat{C}(v)$, and $\widehat{g}(\eta)$ are the nominal matrices and vectors of $M, C(v)$, and $g(\eta)$, respectively. The derivative of $v_{r}$ in (9) is given as

$$
\begin{align*}
\dot{v}_{r}= & \dot{J}^{-1} \dot{\eta}_{d}+J^{-1} \ddot{\eta}_{d}-\alpha \dot{J}^{-1} \widetilde{e}_{B}-\alpha J^{-1} \dot{\tilde{e}}_{B} \\
& -\beta \dot{J}^{-1} \int \widetilde{e}_{B} d t-\beta J^{-1} \widetilde{e}_{B} \tag{12}
\end{align*}
$$

where $\dot{\eta}_{d}(t), \ddot{\eta}_{d}(t)$, and $\dot{j}^{-1}(t)$ are all assumed to be bounded functions of time. Substituting (11) into (2) produces a closedloop dynamic equation for $r(t)$ as follows:

$$
\begin{equation*}
M \dot{r}+C(v) r+K_{v} r-\omega=0 \tag{13}
\end{equation*}
$$

where $\omega$ is the extended disturbance vector which is defined in the following form:

$$
\begin{equation*}
\omega=\widetilde{M} \dot{v}_{r}+\widetilde{C}(v) v_{r}+\widetilde{g}(\eta)-F_{\mathrm{ext}} \tag{14}
\end{equation*}
$$

where $\widetilde{(\cdot)}=\widehat{(\cdot)}-(\cdot)$ denotes the parameter estimation error. The modeling error acts as a disturbance in (14) when the AUV is in motion. Note that $\kappa$ is the $L_{2}$ gain for disturbance attenuation satisfying the following condition:

$$
\begin{equation*}
\int_{0}^{T} z^{T} z d t=\int_{0}^{T} \omega^{T} \omega d t \tag{15}
\end{equation*}
$$

where $z^{T} z$ is defined as the weighted sum of the quadratic forms of the error states and the control input. Since nonlinear $H_{\infty}$ optimal control scheme is based on feedback tracking errors, $z^{T} z$ can be approximated up to magnitude of these errors.

Remark 1. Equation (13) can be represented in state space such that the nonlinear $H_{\infty}$ optimality satisfies [18]

$$
\begin{equation*}
\int_{0}^{T}\left\{x^{T} Q x+u^{T} R u\right\} d t=\kappa^{2} \int_{0}^{T} \omega^{T} \omega d t \tag{16}
\end{equation*}
$$

with $\alpha^{2}>2 \beta . x$ and $u$ in (16) denote the state and input variables, respectively. Meanwhile, the matrices $Q$ and $R$ are state weighting and input weighting matrices, respectively, and they are determined by inverse optimal problem with respect to specific $L_{2}$ attenuation gain, $\kappa$.

Theorem 2. Let the filtered tracking error vector $r$ be upper bounded as the following form:

$$
\begin{equation*}
\|r\| \leq \sqrt{\frac{\Upsilon}{\gamma}} \frac{\kappa^{2}}{\sqrt{2 k_{m} \kappa^{2}+1}}\|\omega\|_{\infty} \tag{17}
\end{equation*}
$$

where $k_{m}$ is scalar constant and $\|\omega\|_{\infty}$ denotes an infinity norm of $\omega$ for a given time interval. Then, the control law (11) above is continuous and the closed-loop system is uniformly ultimately bounded (u.u.b) as defined in [11].

Proof. The following nonnegative function is introduced to analyze the stability of the proposed control law:

$$
\begin{equation*}
V=\frac{1}{2} r^{T} M r . \tag{18}
\end{equation*}
$$

Differentiating $V$ with respect to time and utilizing (10) and (14), a closed-loop dynamic (13) yields

$$
\begin{equation*}
\dot{V}=-r^{T} C(v) r-r^{T} K_{v} r+r^{T} \omega \tag{19}
\end{equation*}
$$

Simplifying (19) leads to

$$
\begin{align*}
\dot{V}= & -r^{T} K_{v} r+r^{T} \omega \\
= & -r^{T}\left(K+\left(\frac{1}{\kappa^{2}}\right) I\right) r+r^{T} \omega \\
= & -r^{T}\left(K+\left(\frac{1}{2 \kappa^{2}}\right) I\right) r  \tag{20}\\
& -\frac{\kappa^{2}}{2}\left|\frac{1}{\kappa^{2}} r-\omega\right|^{2}+\frac{\kappa^{2}}{2}\|\omega\|_{\infty}^{2}
\end{align*}
$$



Figure 2: The desired lawnmower trajectory where " $x$ " marks the initial position of the AUV.
where Property 2 is used. Let $k_{m}$ be the minimum diagonal element of gain matrix $K$ utilizing the worst case disturbance [11] to yield the following inequality:

$$
\begin{equation*}
\dot{V} \leq-\left(k_{m}+\frac{1}{2 \kappa^{2}}\right)\|r\|^{2}+\frac{\kappa^{2}}{2}\|\omega\|_{\infty}^{2} . \tag{21}
\end{equation*}
$$

From (21), it is necessary to choose sufficiently large value of $k_{m}$ to ensure the negative definiteness of $\dot{V}$. Therefore, implying the results and terminology of [19], the ultimate boundedness of $\|r\|$ can be obtained as in (17).

Remark 3. It is assumed that the norm of extended disturbance which includes tracking errors is not deviated largely, when a control input (11) is used in (2). Thus, the control gain can be changed according to (17), so that the satisfactory performance of proposed control law with region function formulation can be achieved.

## 4. Simulation Results

In this section, simulation studies are carried out to assess the effectiveness of the proposed nonlinear $H_{\infty}$ optimal control law with region function formulation for an underwater vehicle. The performance of conventional tracking control and the proposed technique is observed concerning two cases: the first case is the conventional region tracking control and the second case is where the region function is adopted with nonlinear $H_{\infty}$ optimal control law. Both control laws are observed with respect to the existence of random disturbances and bounded ocean current. The ODIN AUV [20, 21] that is known as a near-spherical omnidirectional vehicle equipped with four horizontal thrusters and four vertical thrusters is chosen as the Autonomous Underwater


Figure 3: Three-dimensional view for conventional region tracking control.

Table 1: Simulation result: vehicle forces for four vertical thrusters (N).

|  | Set-point tracking <br> control | Nonlinear $H_{\infty}$ optimal <br> controller |
| :--- | :---: | :---: |
| Thrusters 1 | 45.20 | 42.20 |
| Thrusters 2 | 71.38 | 66.80 |
| Thrusters 3 | 45.38 | 42.78 |
| Thrusters 4 | 66.96 | 65.69 |
| Total input | 116.97 | 111.30 |

Vehicle model in these numerical simulations. The following inequality function is defined as

$$
\begin{align*}
f\left(\delta \eta_{B}\right)= & s_{x}\left(x-x_{0}\right)^{2}+s_{y}\left(y-y_{0}\right)^{2} \\
& +s_{z}\left(z-z_{0}\right)^{2}+s_{\phi}\left(\phi-\phi_{0}\right)^{2}  \tag{22}\\
& +s_{\theta}\left(\theta-\theta_{0}\right)^{2}+s_{\psi}\left(\psi-\psi_{0}\right)^{2} \leq \kappa_{r}^{2}
\end{align*}
$$

where the element of $\left\{s_{x}, s_{y}, s_{z}, s_{\phi}, s_{\theta}, s_{\psi}\right\}$ is the component of the time-varying scaling matrix $B$ and $\kappa_{r}$ is a scalar tolerance. In these simulations, the matrix $B$ is defined as the identity matrix and $\kappa_{r}$ is set to 0.25 . Note that (22) can also be represented as the root mean square error for all axes. In Table 1, the norm values of required forces for four vertical thrusters are presented. The total control input is included to signify the overall energy needed for the system to maintain at depth -1.2 meter. Notice that when the proposed controller is utilized, the energy requirement is reduced as compared with set-point tracking method.

The underwater vehicle is required to track a predefined trajectory as illustrated in Figure 2 where the green (crosssection) path is the horizontal basis position initialized at the position $\left[\begin{array}{lll}1.5 & 0 & -1.2\end{array}\right]^{T} \mathrm{~m}$. Moreover, the vehicle is initialized at the same position $\eta_{1}(0)=\left[\begin{array}{lll}1.5 & 0 & -1.2\end{array}\right]^{T} \mathrm{~m}$ while its attitude is kept constant during simulation and the initial values are $\eta_{2}(0)=\left[\begin{array}{lll}0 & 0 & 0\end{array}\right]^{T}$ degrees. From Figures 3, 4, 5 , and 6 , it has been shown that the proposed control scheme


Figure 4: Planar view for conventional region tracking control.


Figure 5: Three-dimensional view for the nonlinear $H_{\infty}$ optimal control law with region function formulation.
exhibited a more robust tracking performance than the conventional region control, when parameter uncertainties, current effects, and disturbances exist. In Figure 4, the acute fluctuations in the early stages of the simulation were mainly caused by parameter uncertainties in the restoring force and moment. However, as long as the AUV is inside the desired region, the control input is turned off, and when the disturbances pull the vehicle out, the control input is applied to navigate the AUV back into the region.

## 5. Conclusion

A new nonlinear $H_{\infty}$ optimal control law with region function formulation for a hovering underwater vehicle with four horizontal and four vertical thrusters has been presented in this paper. Two cases have been considered: the first case is the conventional region tracking control and the second case is where the region function is adopted with the nonlinear $H_{\infty}$ optimal control law. Both control laws are observed with respect to the existence of unidirectional and bounded


Figure 6: Planar view for nonlinear $H_{\infty}$ optimal control law with region function formulation.
ocean currents. Although the underwater disturbances exist during task execution, the AUV is still able to track a desired moving region. A Lyapunov-like function has been proposed for stability analysis. Simulation results have been presented to demonstrate the performance of the proposed controller.

## Appendix

An omnidirectional intelligent navigator (ODIN) is a nearspherical AUV designed in the University of Hawaii. The dynamic model of ODIN is given by [20,21]

$$
\begin{equation*}
\left[M_{\mathrm{RB}}+M_{\mathrm{A}}\right] \dot{v}+\left[C_{\mathrm{RB}}(v)+C_{\mathrm{A}}(v)\right] v+D(v) v+g(\eta)=\tau, \tag{A.1}
\end{equation*}
$$

where the subscripts RB and A represent the rigid body and added mass terms of the relevant parameters, respectively. The numerical values for the matrices of the vehicle dynamic equation (A.1) are given as

$$
M_{\mathrm{RB}}=\left[\begin{array}{cccccc}
m & 0 & 0 & 0 & m z_{G} & 0  \tag{A.2}\\
0 & m & 0 & -m z_{G} & 0 & 0 \\
0 & 0 & m & 0 & 0 & 0 \\
0 & -m z_{G} & 0 & I_{x x} & 0 & 0 \\
m z_{G} & 0 & 0 & 0 & I_{y y} & 0 \\
0 & 0 & 0 & 0 & 0 & I_{z z}
\end{array}\right],
$$

where $I_{x x}=I_{y y}=I_{z z}=I=(8 / 15) \pi \rho_{v} r_{\text {ODIN }}^{5}$ are the moments of inertia about the principle axes.

Consider

$$
\begin{align*}
& M_{\mathrm{A}}=\left[\begin{array}{cccccc}
X_{u d} & 0 & 0 & 0 & 0 & 0 \\
0 & Y_{v d} & 0 & 0 & 0 & 0 \\
0 & 0 & Z_{w d} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right], \\
& C_{\mathrm{RB}}(v)=\left[\begin{array}{cccccc}
0 & 0 & 0 & m z_{G} u_{6} & m u_{3} & -m u_{2} \\
0 & 0 & 0 & -m u_{3} & m z_{G} u_{6} & m u_{1} \\
0 & 0 & 0 & m\left(u_{2}-z_{G} u_{4}\right) & -m\left(u_{1}+z_{G} u_{5}\right) & 0 \\
-m z_{G} u_{6} & m u_{3} & -m\left(u_{2}-z_{G} u_{4}\right) & 0 & I u_{6} & -I u_{5} \\
-m u_{3} & -m z_{G} u_{6} & m\left(u_{1}+z_{G} u_{5}\right) & -I u_{6} & 0 & I u_{4} \\
m u_{2} & -m u_{1} & 0 & I u_{5} & -I u_{4} & 0
\end{array}\right], \\
& C_{\mathrm{A}}(v)=\left[\begin{array}{cccccc}
0 & 0 & 0 & 0 & m u_{3} & -m u_{2} \\
0 & 0 & 0 & -m u_{3} & 0 & m u_{1} \\
0 & 0 & 0 & m u_{2} & -m u_{1} & 0 \\
0 & m u_{3} & -m u_{2} & 0 & 0 & 0 \\
-m u_{3} & 0 & m u_{1} & 0 & 0 & 0 \\
m u_{2} & -m u_{1} & 0 & 0 & 0 & 0
\end{array}\right],  \tag{A.3}\\
& D(v)=\left[\begin{array}{cccccc}
-d_{t 1}\left|u_{1}\right| & 0 & 0 & 0 & 0 & 0 \\
0 & -d_{t 1}\left|u_{2}\right| & 0 & 0 & 0 & 0 \\
0 & 0 & -d_{t 1}\left|u_{3}\right| & 0 & 0 & 0 \\
0 & 0 & 0 & -d_{r 1}\left|u_{4}\right|-d_{r 2} & 0 & 0 \\
0 & 0 & 0 & 0 & -d_{r 1}\left|u_{5}\right|-d_{r 2} & 0 \\
0 & 0 & 0 & 0 & 0 & -d_{r 1}\left|u_{6}\right|-d_{r 2}
\end{array}\right], \\
& g(\eta)=\left[\begin{array}{c}
\left(m g-\frac{4}{3} \pi r^{3} \rho g\right) \sin (\theta) \\
-\left(m g-\frac{4}{3} \pi r^{3} \rho g\right) \cos (\theta) \sin (\phi) \\
-\left(m g-\frac{4}{3} \pi r^{3} \rho g\right) \cos (\theta) \cos (\phi) \\
z_{G} m g \cos (\theta) \sin (\phi) \\
z_{G} m g \sin (\theta) \\
0
\end{array}\right] .
\end{align*}
$$

Provided that $r_{\text {ODIN }}=0.31 \mathrm{~m}$ is the radius of ODIN, $m=$ 125.0 kg is the mass of ODIN, $z_{G}=0.05 \mathrm{~m}$ is the distance of the center of gravity from the geometric center, $\rho_{v}=$ $965 \mathrm{~kg} / \mathrm{m}^{3}$ is the average density of the ODIN AUV, $\rho=$ $1000 \mathrm{~kg} / \mathrm{m}^{3}$ is the density of fresh water, and $g=9.81 \mathrm{~m} / \mathrm{s}^{2}$. The hydrodynamic derivatives are given by $X_{u d}=Y_{v d}=$ $Z_{w d}=(2 / 3) \pi \rho r_{\text {ODIN }}^{3}$, the translational quadratic damping factor $d_{t 1}=-248 \mathrm{~N}(\mathrm{~s} / \mathrm{m})^{2}$, the angular quadratic damping factor $d_{r 1}=-280 \mathrm{Ns}^{2} / \mathrm{m}$, and the angular linear damping factor $d_{r 2}=-230 \mathrm{Ns}^{2} / \mathrm{m}$.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to thank Offshore Geo-Surveys (OGS) Sdn. Bhd. for invaluable help and advice on the underwater vehicle. This work was supported in part by the Ministry of Higher Education, Malaysia, and Universiti Teknologi Malaysia under Grant no. R.J130000.7823.4L049.

## References

[1] T. Ura, "Autonomous underwater vehicle," Journal of the Robotics Society of Japan, vol. 18, no. 7, pp. 933-936, 2000 (Japanese).
[2] D. R. Yoerger and J. E. Slotine, "Robust trajectory control of underwater vehicles," IEEE Journal of Oceanic Engineering, vol. 10, no. 4, pp. 462-470, 1985.
[3] R. Cristi, F. A. Papoulias, and A. J. Healey, "Adaptive sliding mode control of autonomous underwater vehicles in the dive plane," IEEE Journal of Oceanic Engineering, vol. 15, no. 3, pp. 152-160, 1990.
[4] A. J. Healey and D. Lienard, "Multivariable sliding mode control for autonomous diving and steering of unmanned underwater vehicles," IEEE Journal of Oceanic Engineering, vol. 18, no. 3, pp. 327-339, 1993.
[5] J. Guo, F.-C. Chiu, and C.-C. Huang, "Design of a sliding mode fuzzy controller for the guidance and control of an autonomous underwater vehicle," Ocean Engineering, vol. 30, no. 16, pp. 21372155, 2003.
[6] A. Pisano and E. Usai, "Output-feedback control of an underwater vehicle prototype by higher-order sliding modes," Automatica, vol. 40, no. 9, pp. 1525-1531, 2004.
[7] J. J. E. Slotine, "Sliding controller design for nonlinear systems," International Journal of Control, vol. 40, no. 2, pp. 421-434, 1984.
[8] J. J. E. Slotine and W. Li, Applied Nonlinear Control, PrenticeHall, New Jersey, NJ, USA, 1991.
[9] J. Han and K. C. Wan, "Coordinated motion control of underwater vehicle-manipulator system with minimizing restoring moments," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '08), pp. 3158-3163, Nice, France, September 2008.
[10] J. Han, J. Park, and W. K. Chung, "Robust coordinated motion control of an underwater vehicle-manipulator system with minimizing restoring moments," Ocean Engineering, vol. 38, no. 10, pp. 1197-1206, 2011.
[11] Y. Choi and W. K. Chung, PID Trajectory Tracking Control for Mechanical Systems (Lecture Notes in Control and InFormation Sciences), Springer, Berlin, Germany, 2004.
[12] B. Kim, H. Choi, W. K. Chung, and I. H. Suh, "Analysis and design of robust motion controllers in the unified framework," ASME Journal of Dynamic Systems, Measurement and Control, vol. 124, no. 2, pp. 313-321, 2002.
[13] X. Li, S. P. Hou, and C. C. Cheah, "Adaptive region tracking control for autonomous underwater vehicle," in Proceedings of the 11th International Conference on Control, Automation, Robotics and Vision (ICARCV '10), pp. 2129-2134, Singapore, December 2010.
[14] Z. H. Ismail, B. M. Mokhar, and M. W. Dunnigan, "Tracking control for an autonomous underwater vehicle based on multiplicative potential energy function," in Proceedings of the OCEANS 2012 MTS/IEEE Conference, Yeosu, Republic of Korea, 2012.
[15] Z. H. Ismail, N. Sarman, and M. W. Dunnigan, "Dynamic region boundary-based control scheme for multiple autonomous underwater vehicles," in Proceedings of the OCEANS 2012 MTS/IEEE Conference, Yeosu, Republic of Korea, 2012.
[16] T. I. Fossen, Guidance and Control of Ocean Vehicles, John Wiley and Sons, New York, NY, USA, 1st edition, 1994.
[17] G. Antonelli, Underwater Robots: Motion and Force Control of Vehicle-Manipulator Systems, Springer, Berlin, Germany, 2003.
[18] J. Park, W. Chung, Y. Youm, and M. Kim, " $H_{\infty}$ robust motion control of kinematically redundant manipulators," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '98), pp. 330-335, Victoria, Canada, October 1998.
[19] H. K. Khalil, Nonlinear Systems, Prentice-Hall, Berlin, Germany, 3rd edition, 2001.
[20] T. K. Podder and N. Sarkar, "Fault-tolerant control of an autonomous underwater vehicle under thruster redundancy," Robotics and Autonomous Systems, vol. 34, no. 1, pp. 39-52, 2001.
[21] S. Choi, J. Yuh, and N. Keevil, "design of omni-directional underwater robotic vehicle," in Proceedings of the Conference on Engineering in Harmony with Ocean (Oceans '93), pp. I192-I197, October 1993.

## Research Article

# Parametric Approach to Trajectory Tracking Control of Robot Manipulators 

Shijie Zhang and Yi Ning<br>Research Institute of Robotics, Henan University of Technology, Zhengzhou 450001, China<br>Correspondence should be addressed to Shijie Zhang; zhangshijie1@gmail.com

Received 7 July 2013; Revised 8 September 2013; Accepted 9 September 2013
Academic Editor: Weihai Zhang
Copyright © 2013 S. Zhang and Y. Ning. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The mathematic description of the trajectory of robot manipulators with the optimal trajectory tracking problem is formulated as an optimal control problem, and a parametric approach is proposed for the optimal trajectory tracking control problem. The optimal control problem is first solved as an open loop optimal control problem by using a time scaling transform and the control parameterization method. Then, by virtue of the relationship between the optimal open loop control and the optimal closed loop control along the optimal trajectory, a practical method is presented to calculate an approximate optimal feedback gain matrix, without having to solve an optimal control problem involving the complex Riccati-like matrix differential equation coupled with the original system dynamics. Simulation results of 2-link robot manipulator are presented to show the effectiveness of the proposed method.


## 1. Introduction

Trajectory tracking problem is the most significant and fundamental task in control of robotic manipulator. Motivated by requirements such as a high degree of automation and fast speed operation from industry, various control methods are used such as PID control, adaptive control, variable structure control, neural networks control, and fuzzy control [1-5].

In the past two decades, the optimal control schemes for manipulator arms have been actively researched because the optimal motions that minimize energy consumption, error trajectories, or motion time yield high productivity, efficiency, smooth motion, durability of machine parts, and so forth [6-11]. Various types of methods have been developed to solve the robotic manipulator optimal control schemes.

By the application of the optimal control theory, Pontryagin's maximum principle leads to a two-point boundary value problem. Although this theory and its solutions are rigorous, it has been used to solve equations for the motions of 2-link or at most 3-link planar manipulators due to the complexity and the nonlinearity of the manipulator dynamics [6]. Approximation methods have been studied to obtain the solutions for three or more DOF spatial manipulators. However, the solutions obtained have not been proved to be
optimal. These approximation methods are roughly divided into two groups depending on whether or not they utilize gradients [11]. Recently, the applications of intelligent control techniques (such as fuzzy control or neural network control) with optimal algorithm to the motion control of robot manipulators have received considerable attention [12-17]. But sometimes these methods take quite a long time to find a coefficient that satisfies the requirement of the controlling task. In addition, lack of theoretical analysis and stability security makes industrialists wary of using the results in real industrial environments.

This paper is concerned with the nonlinear optimal feedback control of robot manipulator trajectory tracking. The energy consumption and error trajectories are minimized as performance index in the optimal control problem. An optimal open loop control is first obtained by using a time scaling transform [18] and the control parameterization technique [19]. Then, we derive the form of the optimal closed loop control law, which involves a feedback gain matrix, for the optimal control problem. The optimal feedback gain matrix is required to satisfy a Riccati-like matrix differential equation. Then, the third order $B$-spline function, which has been proved to be very efficient for solving optimal approximation and optimal control problems, is employed


Figure 1: Two-link (RR) robot manipulator.
to construct the components of the feedback gain matrix. By virtue of the relationship between the optimal open loop control and the optimal closed loop control along the optimal trajectory, a practical computational method is presented for finding an approximate optimal feedback gain matrix, without having to solve an optimal control problem involving the complex Riccati-like matrix differential equation coupled with the original system dynamics [20].

## 2. Robot Manipulators Dynamics

2.1. Models of Robot Dynamics. Consider the dynamic equation of a robot manipulator

$$
\begin{equation*}
M(q) \ddot{q}+C(q, \dot{q}) \dot{q}+g(q)=u(t) \tag{1}
\end{equation*}
$$

where $q, \dot{q}, \ddot{q} \in \mathbb{R}^{n}$ are the vectors of the generalized joint coordinates, velocity, and acceleration; $M(q) \in \mathbb{R}^{n \times n}$ denotes a symmetric positive definite inertia matrix; $C(q, \dot{q}) \in \mathbb{R}^{n \times n}$ stands for the Coriolis and centrifugal torques; $g(q) \in \mathbb{R}^{n}$ models the gravity forces; and $u(t) \in \mathbb{R}^{n}$ is the torque input. Some useful properties of robot dynamic are as follows.

Property 1. Matrix $M(q)$ is symmetric and positive definite.
Property 2. Matrix $\dot{M}(q)-2 C(q, \dot{q})$ is skew symmetric and satisfies that

$$
\begin{equation*}
\dot{q}^{T}[\dot{M}(q)-2 C(q, \dot{q})] \dot{q}=0 \tag{2}
\end{equation*}
$$

Property 3. The robot dynamics are passive in open loop, from torque input to velocity output, with the Hamiltonian as its storage function. If viscous friction was considered, the energy dissipates and the system is strictly passive.

The two-link revolute (RR) robot manipulator is shown in Figure 1. The masses of both links and actuators are denoted by $m_{1}$ and $m_{2}$ with $I_{1}$ and $I_{2}$ as mass moment of inertia. $a_{1}$ and $a_{2}$ denote the length; $u_{1}$ and $u_{2}$ are joints torques. The joints positions of the two links are defined by $\theta_{1}$ and $\theta_{2}$.

The dynamic equations of 2 -link RR robot are written in state space form as

$$
\begin{equation*}
\dot{x}=f(x)+B(x) u(t), \tag{3}
\end{equation*}
$$

where $x=\left[q^{T}, \dot{q}^{T}\right]^{T}$ is the system state, $q=\left[\theta_{1}, \theta_{2}\right]^{T}$, and

$$
\begin{gather*}
f(x)=\left[\begin{array}{c}
\dot{q} \\
-M^{-1}(q)(C(q, \dot{q}) \dot{q}+g(q))
\end{array}\right] \\
B(x)=\left[\begin{array}{c}
0 \\
M^{-1}(q)
\end{array}\right] \tag{4}
\end{gather*}
$$

where

$$
\begin{gather*}
M(q)=\left[\begin{array}{ll}
c_{11} & c_{12} \\
c_{21} & c_{22}
\end{array}\right], \\
c_{11}=\left(m_{1}+m_{2}\right) a_{1}^{2}+m_{2} a_{2}^{2}+2 m_{2} a_{1} a_{2} \cos \theta_{2}, \\
c_{12}=c_{21}=m_{2} a_{2}^{2}+m_{2} a_{1} a_{2} \cos \theta_{2}, \\
c_{22}=m_{2} a_{2}^{2}  \tag{5}\\
C(q, \dot{q}) \dot{q}=\left[\begin{array}{c}
-m_{2} a_{1} a_{2}\left(2 \dot{\theta}_{1} \dot{\theta}_{2}+\dot{\theta}_{2}^{2}\right) \sin \theta_{2} \\
m_{2} a_{1} a_{2} \dot{\theta}_{1}^{2} \sin \theta_{2}
\end{array}\right], \\
g(q)=\left[\begin{array}{c}
\left(m_{1}+m_{2}\right) g a_{1} \cos \theta_{1}+m_{2} g a_{2} \cos \left(\theta_{1}+\theta_{2}\right) \\
m_{2} g a_{2} \cos \left(\theta_{1}+\theta_{2}\right)
\end{array}\right] .
\end{gather*}
$$

Define

$$
\begin{gather*}
N(q, \dot{q})=C(q, \dot{q}) \dot{q}+g(q)=\left[\begin{array}{l}
N_{1}(q, \dot{q}) \\
N_{2}(q, \dot{q})
\end{array}\right],  \tag{6}\\
M^{-1}(q)=\left[\begin{array}{ll}
M_{11} & M_{12} \\
M_{21} & M_{22}
\end{array}\right]
\end{gather*}
$$

then,

$$
f(x)=\left[\begin{array}{c}
\theta_{2}  \tag{7}\\
\dot{\theta}_{2} \\
\Theta \dot{\theta}_{1} \\
\Xi \theta_{1}
\end{array}\right], \quad B(x)=\left[\begin{array}{cc}
0 & 0 \\
0 & 0 \\
M_{11} & M_{12} \\
M_{21} & M_{22}
\end{array}\right] ;
$$

here

$$
\begin{align*}
& \Theta=\frac{M_{22}\left(-N_{2}(q, \dot{q})\right)+M_{12}\left(-N_{1}(q, \dot{q})\right)}{x_{1}(t)},  \tag{8}\\
& \Xi=\frac{M_{22}\left(-N_{2}(q, \dot{q})\right)+M_{11}\left(-N_{1}(q, \dot{q})\right)}{x_{2}(t)} .
\end{align*}
$$

2.2. Problem Statement. The purpose of control is to determine an optimal closed loop control signal so that the robot manipulator tracks the desired trajectory with minimal energy consumption, The optimal control problem can be formulated as follows.

Given system (3), find a closed loop control $u(t) \in \mathbb{R}^{n}$ such that the cost function

$$
\begin{equation*}
J=\alpha_{1} \Phi_{0}(x(T))+\alpha_{2} \int_{0}^{T} u^{T} R u d t \tag{9}
\end{equation*}
$$

is minimized, where $\Phi_{0}(x(T))=\left(x(T)-x_{d}\right)^{T} Q\left(x(T)-x_{d}\right), T$ is the free terminal time, $x_{d}$ is the desired trajectory, $\alpha_{1}$ and $\alpha_{2}$
are the weighting parameters, and $Q \in \mathbb{R}^{2 n \times 2 n}$ and $R \in \mathbb{R}^{n \times n}$ are symmetric positive semidefinite and symmetric positive definite weighting matrices, respectively.

We refer to the above problem as problem ( $P$ ). This optimal close loop control problem is very difficult to be solved directly. In this paper, we derive the form of the optimal closed loop control law after obtaining an optimal open loop control by using a time scaling transform and the control parameterization technique. Then the difficulty of the problem is transformed to find a feedback gain matrix which is involved in the optimal closed loop control law. A practical computational method is presented in [20] for finding an approximate optimal feedback gain matrix, without having to solve an optimal control problem involving the complex Riccati-like matrix differential equation coupled with the original system dynamics.

## 3. Parametric Approach to the Optimal Controller Design

By using a time scaling transform and the control parameterization technique, the above problem is solved as an optimal open loop control problem firstly. An optimal open loop control and the corresponding optimal trajectory will be provided.

Let the time horizon $[0, T]$ be partitioned into $p$ subintervals as follows:

$$
\begin{equation*}
0=t_{0} \leq t_{1} \leq \cdots \leq t_{p}=T . \tag{10}
\end{equation*}
$$

The switching times $t_{i}, 1 \leq i \leq p$, are regarded as decision variables. Employing the time scaling transform introduced in [19] to map these switching times into a set of fixed time points $\theta_{i}=i / p, i=1, \ldots, p$, on a new time horizon $[0,1]$. Then the following differential equation is achieved:

$$
\begin{equation*}
\frac{d t(s)}{d s}=v^{p}(s), \quad s \in[0,1] \tag{11}
\end{equation*}
$$

where

$$
\begin{equation*}
v^{p}(s)=\sum_{i=1}^{p} \xi_{i} \chi_{\left[\theta_{i-1}, \theta_{i}\right]}(s), \tag{12}
\end{equation*}
$$

where $\chi_{I}(s)$ denotes the indicator function of $I$ defined by

$$
\chi_{I}(s)= \begin{cases}1, & s \in I  \tag{13}\\ 0, & \text { elsewhere }\end{cases}
$$

and $\xi_{i} \geq 0, \sum_{i=1}^{p} \xi_{i}=T$.
For $s \in\left[\theta_{l-1}, \theta_{l}\right]$, we have

$$
\begin{equation*}
t(s)=\sum_{i=1}^{l-1} \xi_{i}+\xi_{l}\left(s-\theta_{l-1}\right) p \tag{14}
\end{equation*}
$$

where $l=1, \ldots, p$. Clearly,

$$
\begin{equation*}
t(1)=\sum_{i=1}^{p} \xi_{i}=T \tag{15}
\end{equation*}
$$

Then after the time scaling transform, system (3) can be converted into the following form:

$$
\begin{equation*}
\widehat{x}(s)=v^{p}(s)[f(\widehat{x}(s))+B(\widehat{x}(s), s) \widetilde{u}(s)], \tag{16}
\end{equation*}
$$

where $\widehat{x}(s)=\left[\widetilde{x}(s)^{T}, t(s)\right]^{T}, \widetilde{x}(s)=x(t(s))$, and $\widetilde{u}(s)=u(t(s))$.
Now we apply the control parameterization technique to approximate the control $\widetilde{u}(s)$ as follows:

$$
\begin{equation*}
\tilde{u}_{i}^{p}(s)=\sum_{k=-1}^{p+1} \sigma_{k}^{i} \Omega\left(\left(\frac{1}{p}\right) s-k\right), \quad i=1, \ldots, n \tag{17}
\end{equation*}
$$

where

$$
\Omega(\kappa)= \begin{cases}0, & |\kappa|>2  \tag{18}\\ -\frac{1}{6}|\kappa|^{3}+\kappa^{2}-2|\kappa|+\frac{4}{3}, & 1 \leq|\kappa| \leq 2 \\ \frac{1}{2}|\kappa|^{3}-\kappa^{2}+\frac{2}{3}, & |\kappa|<2\end{cases}
$$

is the cubic spline basis function.
Define $\sigma^{i}=\left[\sigma_{-1}^{i}, \ldots, \sigma_{p+1}^{i}\right]^{T}, i=1, \ldots, n$, and $\sigma=$ $\left[\left(\sigma^{1}\right)^{T}, \ldots,\left(\sigma^{n}\right)^{T}\right]^{T}$; let $\Pi$ denote the set containing all $\sigma$. Then $\widetilde{u}^{p}(s)=\left[\widetilde{u}_{1}^{p}(s), \ldots, \widetilde{u}_{n}^{p}(s)\right]^{T}$ is determined uniquely by the switching vector $\sigma \in \Pi$. Thus, it can be written as $\widetilde{u}^{p}(\cdot \mid \sigma)$. Now the optimal parameterization selection problem, which is an approximation of problem $(P)$, can be stated as follows.

Problem (Q). Given system (16), find a combined vector $(\sigma, \xi)$, such that the cost function

$$
\begin{align*}
J(\sigma)= & \alpha_{1} \widehat{\Phi}_{0}(\widehat{x}(1 \mid \sigma)) \\
& +\alpha_{2} \int_{0}^{1} v^{p}(s \mid \xi) \widetilde{u}^{p}(s \mid \sigma)^{T} R \widetilde{u}^{p}(s \mid \sigma) d s \tag{19}
\end{align*}
$$

is minimized, where $\widehat{\Phi}_{0}(\widehat{x}(1 \mid \sigma))=\left(\widehat{x}(1 \mid \sigma)-\widehat{x}_{d}\right)^{T} \widehat{S}(\widehat{x}(1 \mid$ $\sigma)-\widehat{x}_{d}$ ) and $\widehat{x}_{d}$ is the desired trajectory.

Now problem ( $P$ ) is approximated by a sequence of optimal parameter selection problems, each of which can be viewed as a mathematical programming problem and hence can be solved by existing gradient-based optimization methods. Here, our controls are approximated in terms of cubic spline basis functions, and thus they are smooth. problem $(Q)$ can be solved easily by the use of the optimal control software package MISER 3.3 [21].

Suppose that $\left(\widetilde{u}^{p *}, \widehat{x}^{*}\right)$ is the optimal solution of problem $(Q)$. Then it follows that the optimal solution to problem $(P)$ is ( $u^{*}, x^{*}, T^{*}$ ), where $u^{*}$ is the optimal open loop control, $x^{*}$ is the corresponding optimal state vector, and $T^{*}$ is the optimal terminal time. For the computation of the optimal closed loop control problem, we have the following theorem.

Theorem 1. The optimal closed loop control $\bar{u}^{*}$ for problem $(P)$ is given by

$$
\begin{equation*}
\bar{u}^{*}(t)=\frac{1}{2 \alpha_{2}} R^{-1} B^{T} K(t) f\left(x^{*}(t), t\right), \tag{20}
\end{equation*}
$$



Figure 2: Position tracking of the end joint.
where $x^{*}$ is the optimal state and $K(t)$ is the solution of the following Riccati-like differential equation

$$
\begin{equation*}
\left(\dot{K}+K F+F^{T} K+\frac{1}{2} K F B R^{-1} B^{T} K\right) f+K D=0 \tag{21}
\end{equation*}
$$

where

$$
\begin{gather*}
F=\frac{\partial f}{\partial x}, \quad D=\frac{\partial f}{\partial t}, \\
K(T) f(x(T), T)=\alpha_{1} \frac{\partial \Phi_{0}(x(T))}{\partial x(T)}=2 \alpha_{1}\left(x(T)-x_{d}\right) S . \tag{22}
\end{gather*}
$$

The proof is similar to that given for Theorem 3.1 in [22]. Details can refer to this literature.

Problem (R). Subject to the dynamical system (1), with $\bar{u}$ given by Theorem 1 , find a $K(t)$ such that the cost function (17) also with $\bar{u}$ is minimized.

By Theorem 1, although the form of the optimal closed loop control law is given, the matrix function $K(t)$ is still required to be obtained. The solving process involves solving a new optimal control problem denoted as follow. Using the method proposed in [20], problem $(R)$ could be solved well.

In [15], an alternative approach was proposed to construct an approximate optimal matrix function $K^{*}(t)$ without having to solve this complicated optimal control problem $(R)$. The basic idea is explained as follows. Suppose that $u^{*}$ is an optimal open loop control of problem ( $P$ ) and that $x^{*}$ is the corresponding optimal state. We now consider problem $(P)$ with $x=x^{*}$, that is, along the optimal open loop path, and our task is to find a $K^{*}(t)$ such that $\breve{u}^{*}=$ $\left(1 / 2 \alpha_{2}\right) R^{-1} B^{T} K^{*}(t) f\left(x^{*}(t), t\right)$ best approximates the control $\bar{u}^{*}$ in the mean square sense. Then $\breve{u}^{*}$ can be regarded as a good approximate optimal feedback control for problem $(P)$.

The calculation steps of solving $K^{*}(t)$ are as follows.
Step 1. The time horizon $\left[0, T^{*}\right]$ is partitioned into $p$ equal subintervals:

$$
\begin{equation*}
0=t_{0} \leq t_{1} \leq \cdots \leq t_{p} \leq t_{p+1}=T^{*} \tag{23}
\end{equation*}
$$

Step 2. Let

$$
\begin{equation*}
\left[K(t)_{i, j}\right] \approx \sum_{k=-1}^{p+1}\left(c_{i, j, k}\right) \Omega\left(\left(\frac{T^{*}}{p}\right) t-k\right) \tag{24}
\end{equation*}
$$

where $c_{i, j, k}, i, j=1,2, \ldots, n, k=-1,0, \ldots, p+1$, are real constant coefficients that are to be determined. $p$ is the number of equality subintervals on $\left[0, T^{*}\right]$, and $p+3$ is the total number of cubic spline basis functions used in the approximation of each $\left[K(t)_{i, j}\right]$.

Step 3. Let

$$
\begin{equation*}
\Upsilon(K)=\int_{0}^{T^{*}}\left\|u^{*}(t)-\breve{u}(t)\right\| d t \tag{25}
\end{equation*}
$$

where

$$
\begin{equation*}
\breve{u}(t)=\frac{1}{2 \alpha_{2}} R^{-1} B^{T} K(t) f\left(x^{*}(t), t\right) . \tag{26}
\end{equation*}
$$

Step 4. Find coefficients $c_{i, j, k}$ such that the cost function (25) is minimized. These optimal coefficients can be obtained by solving the following optimality conditions:

$$
\begin{equation*}
\Lambda=\frac{\partial \Upsilon(K)}{\partial c_{i, j, k}}=0 \tag{27}
\end{equation*}
$$

We can see that these are linear equations and hence are easy to solve.

## 4. Simulation

In this section, the simulations of the nonlinear optimal control for the 2-link RR-robot manipulator are performed to show the efficiency of the proposed method.

Assuming that the friction is negligible, two-link robot manipulators is simulated with following parameters:

$$
\begin{align*}
& m_{1}=1 \mathrm{Kg} \\
& m_{2}=1 \mathrm{Kg} \\
& a_{1}=1 \mathrm{~m} \\
& a_{2}=1 \mathrm{~m}  \tag{28}\\
& g=9.8 \mathrm{~m} / \mathrm{s}^{2} \\
& x_{0}=[1,1]^{T} \\
& \dot{x}_{0}=[1,1]^{T}
\end{align*}
$$



Figure 3: Velocity tracking of the end joint.


Figure 4: The control inputs of the end joint.

The control objective is to track the desired trajectory given by

$$
\begin{align*}
& q_{1 d}=1+0.2 \cos (\pi t), \\
& q_{2 d}=1+0.2 \sin (\pi t) \tag{29}
\end{align*}
$$

The evolution of tracking errors are as follows

$$
e=\left[\begin{array}{ll}
e_{1} & e_{2}
\end{array}\right]^{T}=\left[\begin{array}{ll}
q_{1}-q_{1 d} & q_{2}-q_{2 d} \tag{30}
\end{array}\right]^{T}
$$

In the simulation, the time horizon $[0, T]$ is partitioned into 20 subintervals. $\alpha_{1}=3, \alpha_{2}=1$, and $S$ and $R$ are unit matrices of proper dimension. We first use the time scaling transform and the control parameterization method to construct the corresponding approximated problem (Q). Then, MISER 3.3 is utilized to solve it, giving rise to an optimal open loop control and the corresponding optimal trajectory. Then the feedback gain matrix $K^{*}(t)$ is obtained by the above calculation steps.

Simulation results are shown in Figures 2 to 4. The position tracking and the velocity tracking of the end joint are shown in Figures 2 and 3, and the control input of the end joint is shown in Figure 4.

## 5. Conclusions

A parametric approach to trajectory tracking control of robot manipulators is studied in this paper, in which an optimal open loop control is obtained firstly by using the control parametrization method and the time scaling transform. Then, we obtained the form of the optimal closed loop control law, where the feedback gain matrix is required to satisfy a Riccati-like matrix differential equation, and a practical method was proposed to calculate the feedback gain matrix. The simulation results demonstrate the validity of the proposed method.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61075083), the Natural Science Foundation of Henan Education Department (13A120204), the Technology Innovation Talents Plan of Henan University of Technology (2012CXRC14), and the Science Foundation of Henan University of Technology (2012BS043).

## References

[1] S. Mahmoudzadeh and H. Mojallali, "An optimized PID for legless capsubots using modified imperialist competitive algorithm," in Proceedings of the 1st RSI/ISM International Conference on Robotics and Mechatronics, pp. 194-199, 2013.
[2] J. P. Ferreira, M. M. Crisóstomo, and A. P. Coimbra, "Adaptive PD controller modeled via support vector regression for a biped robot," IEEE Transactions on Control Systems Technology, vol. 21, no. 3, pp. 941-949, 2013.
[3] J. Kasac, B. Novakovic, D. Majetic, and D. Brezak, "Passive finite-dimensional repetitive control of robot manipulators," IEEE Transactions on Control Systems Technology, vol. 16, no. 3, pp. 570-576, 2008.
[4] M. Asgari and M. A. Ardestani, "Dynamics and control of a novel 3-DoF spatial parallel robot," in Proceedings of the 1st RSI/ISM International Conference on Robotics and Mechatronics, pp. 183-188, 2013.
[5] R.-J. Wai and R. Muthusamy, "Fuzzy-neural-network inherited sliding-mode control for robot manipulator including actuator dynamics," IEEE Transactions on Neural Networks and Learning Systems, vol. 24, pp. 274-287, 2013.
[6] E.-B. Meier and A. E. Bryson Jr., "Efficient algorithm for timeoptimal control of a two-link manipulator," Journal of Guidance, Control, and Dynamics, vol. 13, no. 5, pp. 859-866, 1990.
[7] D. Constantinescu and E. A. Croft, "Smooth and time-optimal trajectory planning for industrial manipulators along specified paths," Journal of Robotic Systems, vol. 17, pp. 233-249, 2000.
[8] M. Galicki and D. Uciński, "Time-optimal motions of robotic manipulators," Robotica, vol. 18, no. 6, pp. 659-667, 2000.
[9] C. W. J. Hol, L. G. Van Willigenburg, E. J. Van Henten, and G. Van Straten, "A new optimization algorithm for singular and non-singular digital time-optimal control of robots," in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '01), pp. 1136-1141, May 2001.
[10] J.-K. Park and J. E. Bobrow, "Reliable computation of minimumtime motions for manipulators moving in obstacle fields using a successive search for minimum-overload trajectories," Journal of Robotic Systems, vol. 22, no. 1, pp. 1-14, 2005.
[11] Z. Shiller, "On singular time-optimal control along specified paths," IEEE Transactions on Robotics and Automation, vol. 10, no. 4, pp. 561-566, 1994.
[12] E. Bobaşu and D. Popescu, "On modelling and multivariable adaptive control of robotic manipulators," WSEAS Transactions on Systems, vol. 5, no. 7, pp. 1579-1586, 2006.
[13] F. M. Raimondi and M. Melluso, "A neuro fuzzy controller for planar robot manipulators," WSEAS Transactions on Systems, vol. 3, pp. 2991-2996, 2004.
[14] D. Popescu, "Neural control of manipulators using a supervisory algorithm," in Proceedings of the International Conference on Automation and Quality Control, pp. A576-A581, ClujNapoca, Romania, 1998.
[15] S.-J. Huang and J.-S. Lee, "A stable self-organizing fuzzy controller for robotic motion control," IEEE Transactions on Industrial Electronics, vol. 47, no. 2, pp. 421-428, 2000.
[16] Y. H. Kim and F. L. Lewis, "Optimal design of CMAC neuralnetwork controller for robot manipulators," IEEE Transactions on Systems, Man and Cybernetics Part C, vol. 30, no. 1, pp. 22-31, 2000.
[17] F. L. Lewis, A. Yesildirek, and K. Liu, "Multilayer neural-net robot controller with guaranteed tracking performance," IEEE Transactions on Neural Networks, vol. 7, pp. 388-396, 1996.
[18] K. L. Teo, L. S. Jennings, H. W. J. Lee, and V. Rehbock, "The control parameterization enhancing transform for constrained optimal control problems," Australian Mathematical Society B, vol. 40, no. 3, pp. 314-335, 1999.
[19] K. L. Teo, C. J. Goh, and K. H. Wong, A Uniled Computational Approach to Optimal Control Problems, John Wiley and Sons Press, New York, NY, USA, 1991.
[20] J. Zhou, D. Zhou, K. L. Teo, and G. Zhao, "Nonlinear optimal feedback control for lunar module soft landing," in Proceedings of the IEEE International Conference on Automation and Logistics (ICAL '09), pp. 681-685, Shenyang, China, August 2009.
[21] L. S. Jennings, M. E. Fisher, K. L. Teo, and C. J. Goh, MISER3. 3 Optimal Control Software Version 3. 3: Theory and User Manual. Centre for Applied Dynamics and Optimization, The University of Western Australia, 1996.
[22] H. W. J. Lee, K. L. Teo, and W. Y. Yan, "Nonlinear optimal feedback control law for a class of nonlinear systems," Neural, Parallel \& Scientific Computations, vol. 4, no. 2, pp. 157-178, 1996.

# Research Article 

# Generalized Kalman-Yakubovich-Popov Lemma Based I-PD Controller Design for Ball and Plate System 

Shuichi Mochizuki ${ }^{1}$ and Hiroyuki Ichihara ${ }^{2}$<br>${ }^{1}$ Graduate School of Science and Technology, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan<br>${ }^{2}$ Department of Science and Engineering, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan

Correspondence should be addressed to Hiroyuki Ichihara; ichihara@messe.meiji.ac.jp
Received 23 June 2013; Accepted 9 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 S. Mochizuki and H. Ichihara. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The ball-on-plate balancing system has a camera that captures the ball position and a plate whose inclination angles are limited. This paper proposes a PID controller design method for the ball and plate system based on the generalized Kalman-Yakubovich-Popov lemma. The design method has two features: first, the structure of the controller called I-PD prevents large input signals against major changes in the reference signal; second, a low-pass filter is introduced into the feedback loop to reduce the influence of the measurement noise produced by the camera. Both simulations and experiments are used to evaluate the effectiveness of the design method.


## 1. Introduction

The ball and plate [1] is an unstable underactuated nonlinear system that has double integrators at the origin and that has two control inputs against four degrees of freedom (DOF). A camera located above the plate captures the position of the ball, and two motors manipulate the inclination angles of the plate to keep the ball on the plate. The ball and plate system is an extension of the ball and beam system [2] from one to two dimensions. The system has demonstrated various controller design methods for positioning and trajectory tracking of the ball: proportional integral derivative (PID) control [3], fuzzy control [4], neural network control [3,5], and model predictive control [6]. In particular, PID control has the benefits of simple implementation and fewer hardware requirements, and it has been applied in many successful designs [7]. Because PID control enables us a limited performance, optimizing the parameters in a PID controller satisfying design specifications is an important subject for study. In the controller design of the ball and plate, it requires to consider limitation of the inclination angles with good transient and steady-state responses. Although proportional and derivative controllers are required to improve transient responses, a jump in the reference signal generates a large input signal
that reaches the limitation angle that degrades the transient responses [8]. In addition, ball position data from the camera include measurement noise that also degrades the steadystate responses.

To overcome the above issues, this paper proposes a PID controller design method for the ball and plate system by open-loop shaping based on the generalized Kalman-Yakubovich-Popov (GKYP) lemma [9]. The GKYP lemma is a generalization of the standard KYP lemma [10], which establishes the equivalence between a frequency domain inequality (FDI) for a transfer function and a linear matrix inequality (LMI) associated with its state-space realization. The standard KYP lemma is available for the infinite frequency range while the generalized one can limit the frequency range to be (semi) finite. By introducing the GKYP lemma to PID controller design, design specifications by FDIs in the finite frequency ranges for the open-loop transfer function result in LMIs [9]. The GKYP lemma gives a systematic openloop shaping design method through optimization to realize desirable transient and steady-state responses. In this visual feedback system, we introduce a low-pass filter. Since the filter gives freedom in optimization, it allows better steadystate responses and reduces the influence of the measurement noise. To prevent large input signals from P and D controllers,


Figure 1: 2D Ball Balancer.
we adopt the I-PD (integral-proportional derivative) structure, whose design is still in the framework of the GKYP lemma, because the open-loop transfer functions of PID and I-PD structures are fundamentally the same.

The paper is organized as follows. The description of the ball and plate system, including its modeling and the measurement noise, is presented in Section 2. The GKYP lemma based I-PD controller design method with a low-pass filter is provided in Section 3. The design of the I-PD controller is described in Section 4. Simulation and experimental results are presented in Sections 5 and 6, respectively. Finally, in Section 7, we present our conclusions.

The notation used is standard. For a matrix $M$, the transpose and complex conjugate transpose are denoted by $M^{\top}$ and $M^{*}$, respectively. For a Hermitian matrix $M, M>$ $(\succeq) 0$ and $M \prec(\preceq) 0$ denote positive (semi) definiteness and negative (semi) definiteness, respectively. The symbol $\mathbf{H}_{n}$ stands for the set of $n \times n$ Hermitian matrices. The subscript $n$ is omitted if $n=2$. The real and imaginary parts of $M$ are denoted by $\mathfrak{R}(M)$ and $\mathfrak{J}(M)$. For matrices $\Phi$ and $P$, $\Phi \otimes P$ denotes the Kronecker product. $\mathscr{L}\{x(t)\}$ represents the Laplace transform of a signal $x(t)$.

## 2. Ball and Plate System

The ball and plate, a QUANSER 2D Ball Balancer, is shown in Figure 1. The system consists of a plate, a ball, an overhead camera, and two servo units. The plate is allowed to swivel in both the $X$ - and $Y$-directions. The overhead CMOS digital camera, a Point Grey Research Inc. FFMV-03M2C-CS, measures the position of the ball. The two servo units located under the plate are QUANSER SRV02 devices, each of which has a peak time of approximately 200 ([ms]) and an overshoot of approximately $5 \%$. Each of the devices is connected to a side of the plate through a two DOF gimbal. The sampling time of the control system and the frame rate provided by the camera are $1([\mathrm{~ms}])$ and $60([\mathrm{fps}])$, respectively. Thus the


Figure 2: The ball and plate system.
image information is renewed approximately every 17 ([ms]). There is a constant time delay of less than 60 ([ms]) between the measurement of the ball position and the manipulation of the servo units in the visual feedback system.
2.1. Modeling. The $X$-direction of the ball and plate system is illustrated in Figure 2. We assume that the ball is completely symmetric and homogeneous and does not slip on the plate and that all frictions are neglected. The plate rotates in the $X Y$-Cartesian coordinates with the origin at the center of the plate. The equations of motion are

$$
\begin{align*}
& \left(m_{b}+\frac{I_{b}}{r_{b}^{2}}\right) \ddot{x}_{b}-m_{b}\left(x_{b} \dot{\alpha}^{2}+y_{b} \dot{\alpha} \dot{\beta}\right)+m_{b} g \sin \alpha=0 \\
& \left(m_{b}+\frac{I_{b}}{r_{b}^{2}}\right) \ddot{y}_{b}-m_{b}\left(y_{b} \dot{\beta}^{2}+x_{b} \dot{\alpha} \dot{\beta}\right)+m_{b} g \sin \beta=0 \tag{1}
\end{align*}
$$

where $\left(x_{b}, y_{b}\right)$ is the position of the ball on the plate, $\alpha$ and $\beta$ are the inclination angles of the plate to the $X$ - and $Y$-axis, respectively, $m_{b}$ is the mass of the ball, $r_{b}$ is the radius of the ball, $g$ is the gravitational acceleration, and $I_{b}$ is the inertia of the ball. In Figure 2, $\theta_{x}$ represents the angle of the load gear. The relationship between $\alpha$ and $\theta_{x}$ is as follows:

$$
\begin{equation*}
\sin \alpha=\frac{2 \sin \theta_{x} r_{\mathrm{arm}}}{L_{\mathrm{tbl}}} \tag{2}
\end{equation*}
$$

where $L_{\mathrm{tbl}}$ is the length of the side of the plate and $r_{\text {arm }}$ is the length between the joint and the center of the load gear. The relationship of $\beta$ and $\theta_{y}$ is the same as (2), since both gear systems have the same hardware and the plate is symmetrical. The numerical values of the constant parameters in the equations of motion and (2) are shown in Table 1. Since $\theta_{x}$ and $\theta_{y}$ are limited as

$$
\begin{equation*}
-30\left[^{\circ}\right] \leq\left\{\theta_{x}, \theta_{y}\right\} \leq 30\left[{ }^{\circ}\right], \tag{3}
\end{equation*}
$$

from (2), the working ranges of $\alpha$ and $\beta$ are

$$
\begin{equation*}
-5.3\left[^{\circ}\right] \leq\{\alpha, \beta\} \leq 5.3\left[^{\circ}\right] \tag{4}
\end{equation*}
$$

If the angular velocities $\dot{\alpha}$ and $\dot{\beta}$ are relatively low, the approximations

$$
\begin{equation*}
\dot{\alpha} \dot{\beta}=0, \quad \dot{\alpha}^{2}=0, \quad \dot{\beta}^{2}=0 \tag{5}
\end{equation*}
$$

Table 1: Parameters of the ball and plate system.

| Parameters | Numerical values |
| :--- | :---: |
| $m_{b}$ | $0.0252[\mathrm{~kg}]$ |
| $r_{b}$ | $0.0170[\mathrm{~m}]$ |
| $g$ | $9.81\left[\mathrm{~m} / \mathrm{s}^{2}\right]$ |
| $I_{b}$ | $2.89 \times 10^{-6}[\mathrm{~kg} \cdot \mathrm{~m}]$ |
| $L_{\text {tbl }}$ | $0.275[\mathrm{~m}]$ |
| $r_{\text {arm }}$ | $0.0254[\mathrm{~m}]$ |

are often used. Linearizing the equations of motion at $\theta_{x}=0$ and $\theta_{y}=0$, we have

$$
\begin{align*}
& \left(m_{b}+\frac{I_{b}}{r_{b}^{2}}\right) \ddot{x}_{b}+\frac{2 m_{b} g r_{\mathrm{arm}}}{L_{\mathrm{tbl}}} \theta_{x}=0,  \tag{6}\\
& \left(m_{b}+\frac{I_{b}}{r_{b}^{2}}\right) \ddot{y}_{b}+\frac{2 m_{b} g r_{\mathrm{arm}}}{L_{\mathrm{tbl}}} \theta_{y}=0 .
\end{align*}
$$

Since the axes are independent of each other, we can focus on one axis, for example, the $X$-axis. For the input $\theta_{x}$ and the output $x_{b}$, the transfer function is given by

$$
\begin{equation*}
P(s)=\frac{X_{b}(s)}{\Theta_{x}(s)}=\frac{K_{\text {bap }}}{s^{2}} \tag{7}
\end{equation*}
$$

where $X_{b}(s)=\mathscr{L}\left\{x_{b}(t)\right\}, \Theta_{x}(s)=\mathscr{L}\left\{\theta_{x}(t)\right\}$, and

$$
\begin{equation*}
K_{\text {bap }}=-\frac{2 m_{b} g r_{\mathrm{arm}} r_{b}^{2}}{L_{\text {tbl }}\left(m_{b} r_{b}^{2}+I_{b}\right)} \tag{8}
\end{equation*}
$$

2.2. Measurement Noise. In this visual feedback system, there is inevitable noise from the camera. To examine the noise level and frequencies, we observed the error signal between a fixed ball position and a measurement signal. The results are shown in Figure 3, where the upper part represents a time history of the error signal including noise and the lower part represents the fast Fourier transform (FFT) analysis of the error signal. The noise level in the error signals is relatively high at frequencies over $20([\mathrm{rad} / \mathrm{s}])$.

## 3. I-PD Control by GKYP Lemma

This section describes an I-PD controller design method based on the GKYP lemma. The feedback control system is shown in Figure 4, where a filter is introduced into the control system.
3.1. Low-Pass Filter. In the previous section, we showed that the measurement noise degrades the control performance. To reduce the influence of the noise, a low-pass filter is available in the controller design. According to the noise properties that we observed, it is sufficient to introduce a first-order lowpass filter into the output of the measurement, such that

$$
\begin{equation*}
F(s)=\frac{\mu}{s+\mu} \tag{9}
\end{equation*}
$$

where $\mu([\mathrm{rad} / \mathrm{s}])$ is the cut-off frequency.


Figure 3: Time history and FFT analysis results of measument signal.


Figure 4: Feedback control system of the ball and plate.
3.2. I-PD Controller. In the standard PID control, major changes in the reference signals cause large input signals to be generated by the proportional and derivative actions in the controller that saturate the actuator. Indeed, it is difficult to tune the parameters in the PID controller (Figure 4) such that the actuator in this visual feedback system is not saturated. The control system with the I-PD controller (Figure 4) has a structure whose inner loop includes the proportional and derivative actions $[7,8]$. In this structure, the integral action alone acts on the error signal and prevents large signals being input to the actuator. The control input $u$ can be written as

$$
\begin{equation*}
u=-K_{p} y+\frac{K_{i}}{s}(r-y)-\frac{K_{d} s}{\tau s+1} y \tag{10}
\end{equation*}
$$

where $\tau(>0)$ is the parameter to approximate the differentiator by a proper transfer function. $K_{p}, K_{i}$, and $K_{d}$ represent the proportional, integral, and derivative gains, respectively. The open-loop transfer function is

$$
\begin{equation*}
L(s)=F(s) P(s) K(s), \tag{11}
\end{equation*}
$$

where

$$
\begin{equation*}
K(s)=K_{p}+\frac{K_{i}}{s}+\frac{K_{d} s}{\tau s+1}=K_{p}\left(1+\frac{1}{T_{i} s}+\frac{T_{d} s}{\tau s+1}\right) . \tag{12}
\end{equation*}
$$

$T_{i}$ and $T_{d}$ are the integral time and derivative time, respectively. It should be noted that the open-loop transfer function
of the I-PD structure is the same as that of the standard PID structure. To tune the parameters in the I-PD controller, we employ an open-loop shaping that realizes a desirable frequency response of the closed-loop system.
3.3. Generalized KYP Lemma. It is known that design specifications for an open-loop transfer function can be reduced to LMIs through the GKYP lemma [9]. We briefly review this lemma in the case of continuous-time systems.

The design specification consists of a frequency range and a desired property in that range. The frequency range can be represented by

$$
\begin{equation*}
\Lambda(\Phi, \Psi):=\{s \in \mathbb{C} \mid \sigma(s, \Phi)=0, \sigma(s, \Psi) \geq 0\} \tag{13}
\end{equation*}
$$

where $\Phi, \Psi \in \mathbf{H}$,

$$
\sigma(s, \Phi):=\left[\begin{array}{ll}
s^{*} & 1
\end{array}\right] \Phi\left[\begin{array}{l}
s  \tag{14}\\
1
\end{array}\right]=0 .
$$

The equality constraint in (13) distinguishes between continuous-time and discrete-time specifications. Since we address continuous-time systems in this paper, we use $\Phi$ such that

$$
\Phi:=\left[\begin{array}{ll}
0 & 1  \tag{15}\\
1 & 0
\end{array}\right] .
$$

The inequality constraint $\sigma(s, \Psi) \geq 0$ in (13) sets a frequency range $\Omega$. For example, a low frequency range is written as

$$
\begin{equation*}
\Omega=\left\{\omega \mid \omega \leq \omega_{l}\right\}=\{\omega \mid \sigma(j \omega, \Psi) \geq 0\} \tag{16}
\end{equation*}
$$

where

$$
\Psi=\left[\begin{array}{cc}
0 & -j  \tag{17}\\
j & 2 \omega_{l}
\end{array}\right]
$$

Table 2 presents a summary of the choice of $\Psi$ versus a type of the frequency range $\Omega$, where $\omega_{\ell}, \omega_{h}, \omega_{1}$, and $\omega_{2}$ are real positive numbers, and $\omega_{c}:=\left(\omega_{1}+\omega_{2}\right) / 2$. On the other hand, the desired property in a specific frequency range can be represented by

$$
[L(j \omega) I] \Pi\left[\begin{array}{c}
L^{*}(j \omega)  \tag{18}\\
I
\end{array}\right] \prec 0
$$

where

$$
\Pi=\left[\begin{array}{ll}
\Pi_{11} & \Pi_{12}  \tag{19}\\
\Pi_{21} & \Pi_{22}
\end{array}\right] \in \mathbf{H}_{m+p}, \quad \Pi_{11} \in \mathbf{H}_{p}, \quad \Pi \succeq 0
$$

$m$ and $p$ are the input and output numbers of $L(s)$, respectively. For SISO systems, consider the requirement that $L(j \omega)$ in a frequency range is on the half plane under a straight line. That is, $L(j \omega)$ is under the straight line, such that

$$
\begin{equation*}
a \Re[L(j \omega)]+b \mathfrak{F}[L(j \omega)]<c \tag{20}
\end{equation*}
$$

that is equivalent to (18) with

$$
\Pi=\left[\begin{array}{cc}
0 & a-j b  \tag{21}\\
a+j b & -2 c
\end{array}\right] \in \mathbf{H}_{2}
$$

TABLE 2: Relation between $\Psi$ and $\Omega$ for continuous time.

| $\Psi$ | $\Omega$ |
| :--- | :---: |
| $\left[\begin{array}{cc}0 & -j \\ j & 2 \omega_{\ell}\end{array}\right]$ | $\omega \leq \omega_{\ell}$ (low) |
| $\left[\begin{array}{cc}-1 & j \omega_{c} \\ -j \omega_{c} & \omega_{1} \omega_{2}\end{array}\right]$ | $\omega_{1} \leq \omega \leq \omega_{2}$ (middle) |
| $\left[\begin{array}{cc}0 & j \\ -j & -2 \omega_{h}\end{array}\right]$ | $\omega_{h} \leq \omega$ (high) |

This requirement is designed to reduce sensitivity in a low frequency range. Another requirement is that $L(j \omega)$ in a frequency range is in the interior of the circle of radius $r$ with the center at $c$. That is, $L(j \omega)$ is in the circle such that

$$
\begin{equation*}
|L(j \omega)-c|^{2}<r^{2} \tag{22}
\end{equation*}
$$

which is equivalent to (18) with

$$
\Pi=\left[\begin{array}{cc}
1 & -c^{*}  \tag{23}\\
-c & |c|^{2}-r^{2}
\end{array}\right] \in \mathbf{H}_{2} .
$$

This requirement is designed to guarantee robustness in a high frequency range. Under these preparations, the generalized KYP lemma [9] is expressed as follows.

Lemma 1. Let $L(s)$ be $C(s I-A)^{-1} B+D . \Lambda(\Phi, \Psi)$ in (13) and $\Pi$ in (18) are given. Assume that $\operatorname{det}(s I-A) \neq 0$ for all $s \in \Lambda$. Then the finite frequency condition

$$
\left[\begin{array}{ll}
L(s) & I
\end{array}\right] \Pi\left[\begin{array}{c}
L^{*}(s)  \tag{24}\\
I
\end{array}\right] \prec 0, \quad \forall s \in \Lambda(\Phi, \Psi)
$$

holds if and only if there exist Hermitian matrices $P$ and $Q$ such that the matrix inequality condition

$$
\left[\begin{array}{cc}
\Gamma  \tag{25}\\
\Pi_{11}\left[\begin{array}{c}
B \\
D
\end{array}\right]^{*} & {\left[\begin{array}{c}
B \\
D
\end{array}\right] \Pi_{11}} \\
-\Pi_{11}
\end{array}\right] \prec 0
$$

is satisfied where

$$
\begin{align*}
\Gamma:= & {\left[\begin{array}{ll}
A & I \\
C & 0
\end{array}\right]\left(\Phi^{\top} \otimes P+\Psi^{\top} \otimes Q\right)\left[\begin{array}{ll}
A & I \\
C & 0
\end{array}\right]^{\top} } \\
& +\left[\begin{array}{cc}
0 & B \Pi_{12} \\
\Pi_{12}^{*} B^{\top} & D \Pi_{12}+\Pi_{12}^{*} D^{\top}+\Pi_{22}
\end{array}\right] . \tag{26}
\end{align*}
$$

Equation (25) is affine with respect to $B, D, P, Q$, and $\Pi_{22}$. In the case where $B$ and $D$ have affine design parameters, (25) is an LMI.

## 4. Control System Design

4.1. Filter Design. Considering the control performance and the noise level, we set the cut-off frequency in (9) as $\mu=$ $20([\mathrm{rad} / \mathrm{s}])$. We examine the effectiveness of this filter in


Figure 5: Time history and FFT analysis result of measurement signal with low-pass filter.
the same experimental setup as given in Section 2.2. Figure 5 shows the spectral analysis results of the measurement and filtered signals, which are represented by the dotted and solid curves, respectively. From these results, it can be seen that the noise at frequencies over $20([\mathrm{rad} / \mathrm{s}])$ has been reduced.
4.2. State-Space Realization of Open-Loop Transfer Function. To obtain an LMI based on the GKYP lemma, a state-space realization of $L(s)$ is required to be affine with respect to a set of the design parameters $\rho=\left(K_{p}, K_{i}, K_{d}\right)$. If we fix $\tau$ in $K(s)$ at $1.0 \times 10^{-2}$, the design parameters appear affinely in the numerator of $K(s)$. Indeed, the controllable canonical form of $K(s)$ is written as

$$
\left[\begin{array}{c|c}
A_{k} & B_{k}(\rho)  \tag{27}\\
\hline C_{k} & D_{k}(\rho)
\end{array}\right]=\left[\begin{array}{cc|c}
0 & 0 & \frac{K_{i}}{\tau} \\
1 & -\frac{1}{\tau} & K_{i}-\frac{K_{d}}{\tau^{2}} \\
\hline 0 & 1 & K_{p}+\frac{K_{d}}{\tau}
\end{array}\right]
$$

Realizations of $P(s)$ and $F(s)$ are also written as

$$
\begin{gather*}
{\left[\begin{array}{c|c}
A_{p} & B_{p} \\
\hline C_{p} & D_{p}
\end{array}\right]=\left[\begin{array}{cc|c}
0 & 1 & 0 \\
0 & 0 & K_{\text {bap }} \\
\hline 1 & 0 & 0
\end{array}\right],}  \tag{28}\\
{\left[\begin{array}{c|c}
A_{f} & B_{f} \\
\hline C_{f} & D_{f}
\end{array}\right]=\left[\begin{array}{c|c}
-\mu & 1 \\
\hline \mu & 0
\end{array}\right],} \tag{29}
\end{gather*}
$$

respectively. By combining these realizations (27)-(29), we obtain a realization of $L(s)$ as

$$
L(s)=\left[\begin{array}{c|c}
A & B(\rho)  \tag{30}\\
\hline C & D(\rho)
\end{array}\right]
$$

where

$$
\begin{align*}
& A=\left[\begin{array}{ccc}
A_{k} & 0 & 0 \\
C_{k} B_{p} & A_{p} & 0 \\
C_{k} D_{p} B_{f} & C_{p} B_{f} & A_{f}
\end{array}\right], \quad B=\left[\begin{array}{c}
B_{k}(\rho) \\
D_{k}(\rho) B_{p} \\
D_{k}(\rho) D_{p} B_{f}
\end{array}\right], \\
& C=\left[\begin{array}{llll}
C_{k} D_{p} B_{f} & C_{p} B_{f} & C_{f}
\end{array}\right], \quad D=D_{k}(\rho) D_{p} D_{f} . \tag{31}
\end{align*}
$$

Consequently, the state-space realization of $L(s)$ is affine with respect to $\rho$.
4.3. Specifications. To shape the Nyquist plot of $L(s)$, we require the following FDI specifications:

$$
\begin{gather*}
-2 \mathfrak{R}[L(j \omega)]+\mathfrak{J}[L(j \omega)]>\gamma_{\ell}, \quad{ }^{\forall} \omega \leq 0.8,  \tag{32}\\
\mathfrak{J}[L(j \omega)]<\gamma_{m}, \quad 2.5 \leq{ }^{\forall} \omega \leq 2.8,  \tag{33}\\
|L(j \omega)|<\gamma_{h}, \quad{ }^{\forall} \omega \geq 10 . \tag{34}
\end{gather*}
$$

Specification (32) with a large $\gamma_{\ell}(>0)$ ensures sensitivity reduction in the low frequency range by making the gain of $L(s)$ high. Specification (33) requires the Nyquist plot to be outside a circle with its center at the point $-1+j 0$ so that a certain stability margin is guaranteed. Specification (34) with a small $\gamma_{h}$ ensures robustness against the unmodeled dynamics that typically exists in the high frequency range.

In addition to the above basic specifications, we also require the following FDIs that improve the property of trajectory tracking

$$
\begin{array}{lc}
4 \Re[L(j \omega)]+\mathfrak{I}[L(j \omega)]<\gamma_{1}, & { }^{\forall} \omega \geq 0.3 \\
4 \Re[L(j \omega)]+\mathfrak{F}[L(j \omega)]>\gamma_{2}, & { }^{\forall} \omega \geq 0.3 . \tag{36}
\end{array}
$$

Since the integral action alone works on the error between the output and the reference signals, the property of trajectory tracking depends directly on the integrator. Here we focus on the corner angular frequency $\omega_{I}$ by the integral action in the I-PD controller which is given by $\omega_{I}=1 / T_{i}$, where $T_{i}=K_{p} / K_{i}$. We have a strong integral action, and the error is corrected quickly when the corner angular frequency is high, while too high a corner angular frequency causes overshoot and hunting. Thus we impose restrictions for the phase of the I-PD controller. It should be noted that the phase at lower frequencies is about -90 ([deg]) while the phase at the corner angular frequency is about 0 [deg]. If we find the frequency at a specific phase from -90 ([deg]) to 0 ([deg]), the corner angular frequency is greater than that frequency. Specifications (35) and (36) restrict the phase of the I-PD controller as well as the open-loop transfer function so that the corner angular frequency is greater than the frequency at the lowest point in the frequency range $0.3([\mathrm{rad} / \mathrm{s}])$.
4.4. I-PD Controller Design. We design an I-PD controller by maximizing $\gamma_{\ell}$ subject to Specifications (32)-(36) where

$$
\begin{equation*}
\left(\gamma_{m}, \gamma_{h}, \gamma_{1}, \gamma_{2}\right)=(-1,0.5,1,-100) \tag{37}
\end{equation*}
$$



Figure 6: Nyquist plot.

That is, the optimization problem is

$$
\max _{K_{i}, K_{p}, K_{d}, \gamma_{e}} \gamma_{\ell}
$$

subject to (19)-(36) and (37).
Each of the design Specifications (32)-(36) is reduced to an LMI condition through Lemma 1 with the realization (30). The Specification (32) is modified to $\varepsilon \leq{ }^{\forall} \omega \leq 0.8$, where $\varepsilon=$ $1.0 \times 10^{-4}$ because $L(s)$ includes the origin poles that prevent us from taking $\omega=0$. Then the LMI optimization problem is to maximize $\gamma_{\ell}$ subject to all these LMI conditions where $K_{i}, K_{p}$, and $K_{d}$ are the common decision variables, while $P_{1}, \ldots, P_{5}$ and $Q_{1}, \ldots, Q_{5}$ appear in the LMIs as independent decision variables. It should be noted that $\gamma_{\ell}$ appears alone in $\Pi_{22}$ in the LMI condition (25) corresponding to (36). In this sense, $\gamma_{l}$ is also an independent decision variable. It should also be noted that $P$ and $Q$ in (25) appear in each of the LMI conditions as the independent decision variables.

To solve this LMI optimization problem, we use YALMIP R20120806 [11], an LMI parser, and SPDT3 version 4.0 [12], an LMI solver, on MATLAB R2011b. The resulting optimal parameters in the I-PD controller and $\gamma_{\ell}$ are

$$
\begin{equation*}
\left(K_{p}, K_{i}, K_{d}, \gamma_{l}\right)=(9.2859,10.7806,4.138,11.9464) \tag{39}
\end{equation*}
$$

The Nyquist plots are shown in Figures 6 and 7 where $L(j \omega)$ satisfies the design specifications given in Section 4.3. Since $\gamma_{\ell}$ is maximized, sensitivity is reduced in the frequency range. The Bode plots of $F(s) P(s), L(s)$, and $K(s)$ are shown in Figure 8 where the corner angular frequency $\omega_{I}$ of the I-PD controller is larger than $0.3([\mathrm{rad} / \mathrm{s}])$.

## 5. Simulation Results

The I-PD controller whose design is described in the previous section was evaluated by a simulation of the step response. To compare the response with that of a standard PID controller,


Figure 7: Nyquist plot (magnification).


Figure 8: Bode plot.
we used the PID controller whose gain parameters are the same as those of the I-PD controller. Since both feedback systems have the same open-loop transfer function, their feedback properties must be the same, provided that each input signal does not saturate. The simulation results of the step response are shown in Figure 9 where the upper and lower parts are the output and input signals, respectively. The solid curves represent the responses given by the I-PD controller while the dashed curves represent those by the PID controller. One can see that the input signal given by the PID controller is saturated, while that by the I-PD controller is not saturated and satisfies the limitation (3). The output signal given by the I-PD controller settles down to the desired value without any overshoot.

It should be noted that the gain parameters in the designed controller are not tuned with the I-PD structure.


Figure 9: Step response simulation results.


Figure 10: Step response experimental results.

In our experience, it is difficult not to saturate the input limitation for the PID structure using any design method.

## 6. Experimental Results

This section evaluates the I-PD controller whose design is given in Section 4 through an experiment of the step response. The PID controller used in Section 4 was also evaluated for comparison. The results of trajectory tracking control by the I-PD controller were also evaluated.
6.1. Step Response Experiment. The results of the step response experiment are shown in Figure 10 where the description of the figure is the same as that of Figure 9. In this experiment, the influence of the time delay appeared and the input signals were slightly larger than those obtained in the simulations. The rise and settling time results are, however, almost the same as those obtained in the simulation.
6.2. Trajectory Tracking Experiments. We tested two kinds of trajectories for tracking control, a square and a circular


Figure 11: Experimental results of tracking control.


Figure 12: Experimental results of square response.


Figure 13: Experimental results of circle response.
trajectory. The side of the square trajectory was $0.1([\mathrm{~m}])$, and the radius of the circular trajectory was $0.05([\mathrm{~m}])$. The results are shown in Figure 11 where the left part shows a trajectory of the square trajectory tracking control, and the right part shows a trajectory of the circular trajectory tracking control. The time histories of the ball and input angles are shown in Figures 12 and 13. In the square trajectory tracking control experiment, the responses were similar to those in the step response experiment except for a slight vibration. Such vibration phenomena are noticeable in the responses of circular trajectory tracking control, in particular, the case when the input signal is relatively small. The reason for these phenomena could be the friction of the ball against the plate or a backlash of the gear system.

## 7. Conclusions

This paper applied the GKYP lemma to an open-loop transfer function including an I-PD controller and a noise reduction filter for the ball and plate system. The multiple FDI specifications for the finite frequency ranges were satisfied by a solution of the LMI optimization problem. The solution includes the optimal parameters in the I-PD controller. The first-order low-pass filter reduced the noise in the high frequency range and improved the steady-state response. Both simulations and experiments evaluated the effectiveness of the design method by comparing the standard PID controller.

The PI-D (proportional integral-derivative) control system, which moved the derivative controller to the inner feedback loop, also has the same open-loop transfer function as the standard PID controller. Thus the approach in this paper can also be applied to the PI-D controller.

## References

[1] M. Moarref, M. Saadat, and G. Vossoughi, "Mechatronic design and position control of a novel ball and plate system," in Proceedings of the Mediterranean Conference on Control and Automation (MED' 08), pp. 1071-1076, June 2008.
[2] P. Kokotovic, "The joy of feedback: nonlinear and adaptive," IEEE Control Systems Magazine, vol. 12, no. 3, pp. 7-17, 1992.
[3] F. Zheng, X. Li, X. Qian, and S. Wang, "Modeling and PID neural network research for the ball and plate system," in Proceedings of the International Conference on Electronics, Communications and Control (ICECC '11), pp. 331-334, September 2011.
[4] X. Fan, N. Zhang, and S. Teng, "Trajectory planning and tracking of ball and plate system using hierarchical fuzzy control scheme," Fuzzy Sets and Systems, vol. 144, no. 2, pp. 297-312, 2004.
[5] M. Bai, Y. Tian, and Y. Wang, "Decoupled fuzzy sliding mode control to ball and plate system," in Proceedings of the 2nd International Conference on Intelligent Control and Information Processing (ICICIP '11), pp. 685-690, July 2011.
[6] F. Borrelli, Constrained Optimal Control of Linear and Hybrid Systems, vol. 290 of Lecture Notes in Control and Information Sciences, Springer, 2003.
[7] K. J. Åstrőm and T. Hägglund, Advanced PID Control, Instrumentation Systems, 2005.
[8] Q. Li and Z. Kemin, Introduction to Feedback Control, Prentice Hall, 2009.
[9] T. Iwasaki and S. Hara, "Generalized KYP lemma: unified frequency domain inequalities with design applications," IEEE Transactions on Automatic Control, vol. 50, no. 1, pp. 41-59, 2005.
[10] A. Rantzer, "On the Kalman-Yakubovich-Popov lemma," Systems \& Control Letters, vol. 28, no. 1, pp. 7-10, 1996.
[11] J. Lőfberg, "YALMIP: a toolbox for modeling and optimization in MATLAB," in Proceedings of the IEEE International Symposium on Computer-Aided Control System Design (CACSD '11), Taipei, Taiwan, 2004.
[12] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3-a MATLAB software package for semidefinite programming, version 1.3", Optimization Methods and Software, vol. 11, no. 1, pp. 545-581, 1999.

# Robust Optimal Sliding-Mode Tracking Control for a Class of Uncertain Nonlinear MIMO Systems 

Haiping Pang ${ }^{1}$ and Xiuqin Yang ${ }^{2}$<br>${ }^{1}$ College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266042, China<br>${ }^{2}$ Naval Aviation Engineering Institute of Qingdao Branch, Qingdao 266041, China<br>Correspondence should be addressed to Haiping Pang; panghp123@163.com

Received 27 April 2013; Revised 19 September 2013; Accepted 23 September 2013
Academic Editor: Weihai Zhang
Copyright © 2013 H. Pang and X. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper addresses the problem of tracking a reference trajectory asymptotically given by a linear time-varying exosystem for a class of uncertain nonlinear MIMO systems based on the robust optimal sliding-mode control. The nonlinear MIMO system is transformed into a linear one by the input-output linearization technique, and at the same time the input-output decoupling is realized. Thus, the tracking error equation is established in a linear form, and the original nonlinear tracking problem is transformed into an optimal linear quadratic regulator (LQR) tracking problem. A LQR tracking controller (LQRTC) is designed for the corresponding nominal system, and the integral sliding-mode strategy is used to robustify the LQRTC. As a result, the original system exhibits global robustness to the uncertainties, and the tracking dynamics is the same as that of LQRTC for the nominal system. So a robust optimal sliding-mode tracking controller (ROSMTC) is realized. The proposed controller is applied to a two-link robot system, and simulation results show its effectiveness and superiority.


## 1. Introduction

Trajectory tracking control for multiple-input multipleoutput (MIMO) nonlinear systems has attached much attention during the past decades [1, 2]. Compared with singleinput single-output (SISO) systems, the optimal tracking control for MIMO systems is much more difficult and complex because the output variables are more than one and usually coupled. Many real nonlinear plants have MIMO structures, such as robots, electric motors, and aerocrafts. The key to solve MIMO problem is to introduce the decoupling technology, and several control schemes for decoupling have been quite mature, such as the cascade decoupling based on classical control theory [3], the linear state feedback decoupling based on modern control theory [4], the linear output feedback decoupling [5], the stable-state feedback decoupling, and the dynamic precompensate [6, 7]. In recent years, adaptive decoupling theory, fuzzy decoupling theory and neural network decoupling theory have also made great achievements [8-11]. But there are some difficulties when these schemes are applied in practical applications; for example, the decoupling control system based on classical control
theory often leads to a physically unrealizable problem, while the decoupling control system based on the modern control theory often leads to complex calculations and its realization is very difficult. As a branch of exact linearization, the input-output linearization is an effective way to decouple MIMO systems [12]. It could be achieved by exact inputoutput transformation and feedback, and any higher-order nonlinear terms are not neglected. Additionally, it could be employed to stabilize systems in a large scale. What's more, it could avoid complicated calculations in dealing with the tracking problem for nonlinear MIMO systems, and it is easy to achieve.

As is well known, optimal control is one of the most important branches in modern control theory and LQRTC has been used and developed well in linear MIMO systems. However, there would be several problems in applying LQRTC to uncertain nonlinear systems. The optimal LQRTC for nonlinear systems often leads to a nonlinear two-point boundary-value (TPBV) problem and the analytical solution generally does not exist except in some simple cases [13]. Additionally, the optimal controller design is usually based on the precise mathematical models. If the system is subject to
some uncertainties, such as parameter variations, unmodeled dynamics, and external disturbances, the performance criterion which is optimized based on the nominal system would deviate from the optimal value or even the system becomes unstable.

Sliding-mode control (SMC) is an effective robust control approach for uncertain nonlinear systems [14, 15]. Its outstanding advantage is that the sliding motion exhibits complete robustness to system uncertainties [16, 17]. However, during the reaching phase, the SMC system is sensitive to uncertainties. Therefore, various methods have been suggested by minimizing or even removing the reaching phase, such as time-varying sliding mode and integral sliding mode (ISM). Time-varying sliding-mode surfaces that can remove the reaching phase were studied in [18] for the SISO system. And for a class of uncertain MIMO nonlinear systems, three types of time-varying sliding-mode control were proposed in [19]. Another effective method to remove the reaching phase and obtain a global robustness is the ISM, which was proposed by Lee [20] for linear systems. Recently, the ISM control research has obtained many results, for example, the optimal and robust control for linear state-delay systems was proposed in [21], the LMI-based ISM control of mismatched uncertain systems was considered in [22]. Compared with the time-varying sliding mode, the ISM is simpler and easier to implement, especially for MIMO systems. How to make an optimal controller or tracking controller have the global robustness of ISM is a valuable subject. For the optimal control problem, [20] studied the problem for linear systems. Reference [23] proposed a higher order sliding-mode control methodology based on ISM for a class of nonlinear SISO systems. Reference [24] presented an ISM surface for a class of nonlinear uncertain systems based on the exact linearization. Reference [25] studied the global robust optimal slidingmode control based on ISM for class of MIMO nonlinear systems, with the nonlinear LQR problem solved by the sensitivity approach. But with system-order increasing, the complexity for calculating optimal solution increases rapidly. For the optimal tracking control problem, [26] studied the optimal sliding-mode output tracking control for linear uncertain systems with the reference signal given by a linear time-varying exosystem. Reference [27] proposed an optimal output tracking controller for nonlinear systems based on successive approximation approach, without uncertainties considered. Reference [28] studied the optimal sliding-mode control by combining ISM with optimal control and applied it to quaternion-based spacecraft attitude tracking maneuvres with external disturbances and an uncertainty inertia matrix. The control Lyapunov function (CLF) approach and Lyapunov optimizing control (LOC) methods were used to solve the nonlinear optimal control problems, respectively, and the desired reference was given by some known trajectories.

The optimal tracking problem for nonlinear MIMO systems with reference signals generated by a time-varying exosystem is more challenging because of the complexity of nonlinear, the difficulty of the optimal solution, the inevitability of uncertainties, the coupling problem, and so on.

In this paper, the input-output linearization is employed to linearize and decouple the original MIMO system. Based on the decoupled system and the exosystem, an error equation is constructed. Therefore, the optimal tracking problem of original system is transferred into an optimal state regulation problem about the linear error system, and the TPBV problem is avoided. Based on optimal control law, an ISM surface is constructed, which can remove the reaching phase of SMC and guarantee the global sliding mode. To reduce chattering, the reaching law is used to design the optimal sliding-mode tracking control law. As a result, not only the optimal performance can be obtained but also the global robustness to uncertainties is guaranteed. The proposed algorithm is applied to a two-link robot system, and simulation results show its effectiveness.

## 2. Problem Formulation

Consider a class of uncertain affine nonlinear MIMO systems as follows:

$$
\begin{gather*}
\dot{x}=f(x)+\Delta f(x)+G(x) u+d(x, t),  \tag{1}\\
y=h(x)
\end{gather*}
$$

where $x \in R^{n}$ is the system state vector, $u \in R^{m}$ is the control input vector, and $y \in R^{m}$ is the system output vector. $G(x)=\left[g_{1}(x), \ldots, g_{m}(x)\right], f(x)$, and $g_{i}(x)$ are sufficiently smooth vector fields on a domain $D \subset R^{n} . h(x)$ is a measured sufficiently smooth output function vector and $h(x)=$ $\left[h_{1}, \ldots, h_{m}\right]^{\mathrm{T}} . \Delta f(x)$ and $d(x, t)$ are unknown function vectors representing the system uncertainties, including system parameter variations, unmodeled dynamics, and external disturbances.

The reference signal $\tilde{y}(t)$ is given by the following exosystem:

$$
\begin{gather*}
\dot{z}(t)=F(t) z(t), \quad z\left(t_{0}\right)=z_{0} \\
\tilde{y}(t)=H(t) z(t) \tag{2}
\end{gather*}
$$

where $z \in R^{p}$ is the state vector and $\tilde{y} \in R^{m}$ is the output vector. $F(t)$ and $H(t)$ are time-varying matrices with appropriate dimensions. Suppose that the pair $[F(t), H(t)]$ is observable.

Our objective is to design a ROSMTC so that the output $y$ of system (1) can track the exosystem's output $\tilde{y}$ asymptotically, some given performance criterion is minimized and the system can exhibit global robustness to uncertainties.

Ignoring the uncertainties, the nominal system of uncertain affine nonlinear system (1) is

$$
\begin{gather*}
\dot{x}=f(x)+G(x) u,  \tag{3}\\
y=h(x) .
\end{gather*}
$$

Assumption 1. Equation (3) has the relative degree vector $\left\{r_{1}, \ldots, r_{m}\right\}$ and $r=r_{1}+\cdots+r_{m}=n$.

Assumption 2. The reference trajectory $\widetilde{y}(t)$ and its derivations $\widetilde{y}^{(i)}(t)(i=1, \ldots, n)$ can be obtained online, and they are bounded to all $t \geq 0$.

The nominal system (3) is a coupling nonlinear MIMO system. In the next part, exact linearization will be employed to transform the nonlinear system into a linear one and achieve decoupling between the inputs and outputs. Furthermore, the original tracking problem will be transformed into a robust optimal regulation problem for linear system.

## 3. Input-Output Linearization and Problem Transformation

Considering system (3) and differentiating $y=h(x)$, we have

$$
\begin{gather*}
y_{i}^{(k)}=L_{f}^{k} h_{i}(x), \quad 0 \leq k \leq r_{i}-1, \\
y_{i}^{\left(r_{i}\right)}=L_{f}^{r_{i}} h_{i}(x)+\sum_{j=1}^{m} L_{g_{j}} L_{f}^{r_{i}-1} h_{i}(x) u_{j} . \tag{4}
\end{gather*}
$$

Define

$$
M(x)=\left[\begin{array}{ccc}
L_{g_{1}} L_{f}^{r_{1}-1} h_{1}(x) & \cdots & L_{g_{m}} L_{f}^{r_{1}-1} h_{1}(x)  \tag{5}\\
\vdots & \cdots & \vdots \\
L_{g_{1}} L_{f}^{r_{m}-1} h_{m}(x) & \cdots & L_{g_{m}} L_{f}^{r_{m}-1} h_{m}(x)
\end{array}\right] .
$$

And $M(x)$ is nonsingular in some domain for all $x \in X_{0}$. According to Assumption 1, (4) can be written as

$$
\left[\begin{array}{c}
y_{1}^{\left(r_{1}\right)}  \tag{6}\\
\vdots \\
y_{m}^{\left(r_{m}\right)}
\end{array}\right]=\left[\begin{array}{c}
L_{f}^{r_{1}} h_{1}(x) \\
\vdots \\
L_{f}^{r_{m}} h_{m}(x)
\end{array}\right]+M(x)\left[\begin{array}{c}
u_{1} \\
\vdots \\
u_{m}
\end{array}\right] .
$$

Choose the control law in the form of

$$
u=-M^{-1}(x)\left[\begin{array}{c}
L_{f}^{r_{1}} h_{1}(x)  \tag{7}\\
\vdots \\
L_{f}^{r_{m}} h_{m}(x)
\end{array}\right]+M^{-1}(x) v ;
$$

then the input-output dynamic equation can be described as

$$
\left[\begin{array}{c}
y_{1}^{\left(r_{1}\right)}  \tag{8}\\
\vdots \\
y_{m}^{\left(r_{m}\right)}
\end{array}\right]=\left[\begin{array}{c}
v_{1} \\
\vdots \\
v_{m}
\end{array}\right]
$$

As can be seen, the output $y_{i}=h_{i}(x)$ is only related to the input $v_{i}$, which means the input-output decoupling has been realized. Noting that the relative degree of system (1) is $r=r_{1}+\cdots+r_{m}=n$, so the decoupling process is equivalent to the input-output linearization process. In the following part, the results above will be applied to uncertain affine nonlinear system (1) to structure a tracking error equation.

Considering system (1) and differentiating $y=h(x)$, we have

$$
\begin{align*}
y_{i}^{(k)} & =L_{f}^{k} h_{i}(x), \quad 0 \leq k \leq r_{i}-1, \\
y_{i}^{\left(r_{i}\right)}= & L_{f}^{r_{i}} h_{i}(x)+L_{\Delta f} L_{f}^{r_{i}-1} h_{i}(x)  \tag{9}\\
& +\sum_{j=1}^{m} L_{g_{j}} L_{f}^{r_{i}-1} h_{i}(x) u_{j}+L_{d} L_{f}^{r_{i}-1} h_{i} .
\end{align*}
$$

Define $\eta_{i}^{j}=L_{f}^{j} h_{i}(x), i=1, \ldots, m$, and $j=0,1, \ldots, \gamma_{i-1}$ and choose the following nonlinear state transformation:

$$
\begin{equation*}
\xi=\left[\eta_{1}^{0}, \ldots, \eta_{1}^{r_{1}-1}, \ldots, \eta_{m}^{0}, \ldots, \eta_{m}^{r_{m}-1}\right]^{\mathrm{T}} \tag{10}
\end{equation*}
$$

Define the tracking error as

$$
\begin{gather*}
e=\left[h_{1}(x)-\tilde{y}_{1}, \ldots, L_{f}^{\left(r_{1}-1\right)} h_{1}(x)-\tilde{y}_{1}^{\left(r_{1}-1\right)}, \ldots,\right. \\
= \\
\left.h_{m}(x)-\tilde{y}_{m}, \ldots, L_{f}^{\left(r_{m}-1\right)} h_{m}(x)-\tilde{y}_{m}^{\left(r_{m}-1\right)}\right]^{\mathrm{T}}  \tag{11}\\
= \\
=\left[e_{1}, \ldots, L_{f}^{\left(r_{1}-1\right)} h_{1}(x)-\tilde{y}_{1}^{\left(r_{1}-1\right)}, \ldots, e_{1}^{\left(r_{1}-1\right)}, \ldots, e_{m}, \ldots, e_{m}^{\left(r_{m}-1\right)}\right]^{\mathrm{T}} .
\end{gather*}
$$

Define $\tilde{Y}=\left[\tilde{y}_{1}^{r_{1}}, \ldots, \tilde{y}_{m}^{r_{m}}\right]^{\mathrm{T}} \in R^{m}$ and choose the control law in the form of

$$
\begin{equation*}
u=M^{-1}(x)[-b(x)+\tilde{Y}+v] \tag{12}
\end{equation*}
$$

where $b(x)=\left[\begin{array}{lll}L_{f}^{\gamma_{1}} h_{1} & \cdots & L_{f}^{\gamma_{m}} h_{m}\end{array}\right]^{\mathrm{T}}$.
So the tracking error equation can be written as

$$
\begin{equation*}
\dot{e}=\bar{A} e+\Delta \bar{A}+\bar{B} v+\Delta \bar{B} \tag{13}
\end{equation*}
$$

where $e \in R^{n}$ is the system tracking error vector and $v \in R^{m}$ is a new control input of the transformed system. $\bar{A}, \Delta \bar{A}$, $\bar{B}$, and $\Delta \bar{B}$ are corresponding constant matrices and defined, respectively, as follows:

$$
\begin{gather*}
\bar{A}=\operatorname{diag}\left(\bar{A}_{1}, \ldots, \bar{A}_{m}\right), \\
\bar{B}=\operatorname{diag}\left(\bar{B}_{1}, \ldots, \bar{B}_{m}\right), \\
\Delta \bar{A}=\left[\Delta \bar{A}_{1}, \ldots, \Delta \bar{A}_{m}\right]^{\mathrm{T}}, \\
\Delta \bar{B}=\left[\Delta \bar{B}_{1}, \ldots, \Delta \bar{B}_{m}\right]^{\mathrm{T}} \\
\bar{A}_{i}=\left[\begin{array}{ccccc}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & 0 & 0
\end{array}\right]_{r_{i} \times r_{i}}, \bar{B}_{i}=\left[\begin{array}{c}
0 \\
0 \\
\vdots \\
0 \\
1
\end{array}\right]_{r_{i} \times 1}  \tag{14}\\
\Delta \bar{A}_{i}=\left[\begin{array}{c}
0 \\
0 \\
0 \\
L_{\Delta f} L_{f}^{r_{i}-1} h_{i}(x)
\end{array}\right]_{r_{i} \times 1}, \\
\Delta \bar{B}_{i}=\left[\begin{array}{c}
0 \\
0 \\
\vdots \\
0 \\
L_{d} L_{f}^{r_{i}-1} h_{i}(x)
\end{array}\right]_{r_{i} \times 1},
\end{gather*}
$$

where $\Delta \bar{A}$ and $\Delta \bar{B}$ represent uncertainties of the transformed system. Obviously, $\Delta \bar{A}$ and $\Delta \bar{B}$ satisfy the matching conditions; that is, there exist unknown continuous function vectors $\Delta \widetilde{A}\left(\in R^{m}\right)$ and $\Delta \widetilde{B}\left(\in R^{m}\right)$ which satisfy

$$
\begin{equation*}
\Delta \bar{A}=\bar{B} \Delta \widetilde{A}, \quad \Delta \bar{B}=\bar{B} \Delta \widetilde{B} \tag{15}
\end{equation*}
$$

Assumption 3. There exist known constants $a_{m}$ and $b_{m}$ such that

$$
\begin{equation*}
\|\Delta \widetilde{A}\|_{1} \leq a_{m}, \quad\|\Delta \widetilde{B}\|_{1} \leq d_{m} \tag{16}
\end{equation*}
$$

where $\|\cdot\|_{1}$ denotes the 1-norm.
After exact linearization and decoupling, the optimal tracking problem of original system (1) is transferred into a robust optimal regulation problem about the error system (13). In the next part, the ROSMTC will be designed for system (13).

## 4. Design of Robust Optimal Sliding-Mode Tracking Controller (ROSMTC)

4.1. Optimal Tracking Control of Nominal System. Ignoring the uncertainties of system (13), the corresponding nominal system is

$$
\begin{equation*}
\dot{e}(t)=\bar{A} e(t)+\bar{B} v(t) \tag{17}
\end{equation*}
$$

For (17), let $v=v_{0}$ and $v_{0}$ can minimize the quadratic performance index as follows:

$$
\begin{equation*}
J_{N}=\frac{1}{2} \int_{0}^{\infty}\left[e^{\mathrm{T}}(t) Q e(t)+v_{0}^{\mathrm{T}}(t) R v_{0}(t)\right] \mathrm{d} t \tag{18}
\end{equation*}
$$

where $Q \in R^{n \times n}$ is a symmetric semipositive definite matrix and $R \in R^{m \times m}$ is a positive definite matrix.

According to optimal control theory, the optimal feedback control law can be described as

$$
\begin{equation*}
v_{0}(t)=-R^{-1} \bar{B}^{\mathrm{T}} P e(t), \tag{19}
\end{equation*}
$$

where $P$ is the solution of matrix Riccati equation as follows:

$$
\begin{equation*}
P \bar{A}+\bar{A}^{\mathrm{T}} P-P \bar{B} R^{-1} \bar{B}^{\mathrm{T}} P+Q=0 \tag{20}
\end{equation*}
$$

So the closed-loop system dynamics is

$$
\begin{equation*}
\dot{e}(t)=\left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(t) \tag{21}
\end{equation*}
$$

According to optimal control theory, the closed-loop system is asymptotically stable. However, if the control law (19) is applied to uncertain system (13), the state trajectory will deviate from the optimal trajectory and even the system will become unstable. Next we will adopt ISM control technique to robustify the optimal control law; to achieve the goal that the state trajectory of uncertain system (13) is the same as that of the optimal trajectory of the nominal system (17).
4.2. The Robust Optimal Sliding Surface. Considering the uncertain system (13), we define an integral sliding surface in the form of

$$
\begin{equation*}
s(e, t)=\bar{G} e(t)-\bar{G} \int_{0}^{t}\left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(\tau) \mathrm{d} \tau-\bar{G} e(0) \tag{22}
\end{equation*}
$$

where $\bar{G} \in R^{m \times n}$ which satisfies that $\bar{G} \bar{B}$ is nonsingular and $e(0)$ is the initial state vector. Differentiating (22) with respect to $t$ and considering (13), we obtain

$$
\begin{align*}
\dot{s}(e, t)= & \bar{G} \dot{e}(t)-\bar{G}\left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(t) \\
= & \bar{G}[\bar{A} e(t)+\Delta \bar{A}+\bar{B} v(t)+\Delta \bar{B}]  \tag{23}\\
& -\bar{G}\left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(t) \\
= & \bar{G} \bar{B} v(t)+\bar{G} \Delta \bar{A}+\bar{G} \Delta \bar{B}+\bar{G} \bar{B} R^{-1} \bar{B}^{\mathrm{T}} P e(t) .
\end{align*}
$$

Let $\dot{s}(t)=0$; then the equivalent control becomes

$$
\begin{equation*}
v_{\mathrm{eq}}(t)=-(\bar{G} \bar{B})^{-1}\left[\bar{G} \Delta A+\bar{G} \Delta \bar{B}+\bar{G} B R^{-1} B^{\mathrm{T}} P e(t)\right] \tag{24}
\end{equation*}
$$

Substituting (24) into (13) and considering (15), the slidingmode dynamics is

$$
\begin{align*}
\dot{e}(t)= & \bar{A} e(t)+\Delta \bar{A}+\bar{B} v_{\mathrm{eq}}(t)+\Delta \bar{B} \\
= & \bar{A} e(t)+\Delta \bar{A}-\bar{B}(\bar{G} \bar{B})^{-1} \\
& \times\left[\bar{G} \Delta \bar{A}+\bar{G} \Delta \bar{B}+\bar{G} \bar{B} R^{-1} \bar{B}^{\mathrm{T}} P e(t)\right]+\Delta \bar{B}  \tag{25}\\
= & \bar{A} e(t)+\Delta \bar{A}-\Delta \bar{A}-\Delta \bar{B} \\
& -\bar{G}^{-1} \bar{G} \bar{B} R^{-1} \bar{B}^{\mathrm{T}} P e(t)+\Delta \bar{B} \\
= & \left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(t) .
\end{align*}
$$

Comparing (25) with (21), we can see that the sliding mode of uncertain linear system (13) is the same as optimal dynamics of (17); therefore, the sliding mode is also asymptotically stable, and the sliding motion guarantees the controlled system global robustness to the uncertainties which satisfy the matching condition. We call (22) a global robust optimal sliding surface.
4.3. The Control Law. For uncertain system (13), we propose the control law as follows:

$$
\begin{gather*}
v(t)=v_{c}(t)+v_{d}(t), \\
v_{c}(t)=-R^{-1} \bar{B}^{\mathrm{T}} P e(t),  \tag{26}\\
v_{d}(t)=-(\bar{G} \bar{B})^{-1}[k s+\varepsilon \operatorname{sgn}(s)],
\end{gather*}
$$

where $\operatorname{sgn}(s)=\left[\operatorname{sgn}\left(s_{1}\right), \ldots, \operatorname{sgn}\left(s_{m}\right)\right]^{\mathrm{T}}$ and $k$ and $\varepsilon$ are appropriate positive constants, respectively. $v_{c}(t)$, used to stabilize
and optimize the nominal system, is the continuous part of the control law. $v_{d}(t)$ is the discontinuous part, which provides complete compensation for uncertainties of system (13).

Theorem 4. Consider uncertain system (13) with Assumption 3. Let the input $v(t)$ and the sliding surface be given by (26) and (22), respectively. The control law can force the system trajectories to reach the sliding surface in finite time and maintain it thereafter if $\varepsilon \geq\left(a_{m}+d_{m}\right)\|\bar{G} \bar{B}\|_{1}$.

Proof. Utilizing $V=(1 / 2) s^{T} s$ as a Lyapunov function candidate and considering Assumption 3, we obtain

$$
\begin{align*}
\dot{V}= & s^{\mathrm{T}} \dot{s} \\
= & s^{\mathrm{T}}\left[\overline{\mathrm{G}} \dot{e}(t)-\bar{G}\left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(t)\right] \\
= & s^{\mathrm{T}}\{\bar{G}[\bar{A} e(t)+\Delta \bar{A}+\bar{B} v(t)+\Delta \bar{B}] \\
& \left.\quad-\bar{G}\left(\bar{A}-\bar{B} R^{-1} \bar{B}^{\mathrm{T}} P\right) e(t)\right\} \\
= & s^{\mathrm{T}}\left\{\bar{G} \Delta \bar{A}+\bar{G} \bar{B}\left[-R^{-1} \bar{B}^{\mathrm{T}} P e(t)-(\overline{\mathrm{G}} \bar{B})^{-1}\right.\right.  \tag{27}\\
& \quad \times(k s+\varepsilon \operatorname{sgn}(s))+\overline{\mathrm{G}} \Delta \bar{B}] \\
= & \quad s^{\mathrm{T}}\left\{\bar{G} \Delta \bar{A}+\bar{G} \Delta R^{-1} \bar{B}^{\mathrm{T}} P e(t)\right\} \\
= & -k\|s\|_{2}^{2}-\varepsilon\|s\|_{1}+s^{\mathrm{T}}(\bar{G} \Delta \bar{A}+\bar{G} \Delta \bar{B}),
\end{align*}
$$

where $\|\cdot\|_{2}$ denotes the 2 -norm. As $s^{T}(\bar{G} \Delta \bar{A}+\bar{G} \Delta \bar{B})$ is a scalar quantity considering (15) and Assumption 3, we get

$$
\begin{align*}
s^{\mathrm{T}}(\bar{G} \Delta \bar{A}+\bar{G} \Delta \bar{B}) & \leq\left\|s^{\mathrm{T}}(\bar{G} \Delta \bar{A}+\bar{G} \Delta \bar{B})\right\|_{1} \\
& =\left\|s^{\mathrm{T}}(\bar{G} \bar{B} \Delta \widetilde{A}+\bar{G} \bar{B} \Delta \widetilde{B})\right\|_{1} \\
& \leq\|s\|_{1} \cdot\|\bar{G} \bar{B}\|_{1} \cdot\left(\|\Delta \widetilde{A}\|_{1}+\|\Delta \widetilde{B}\|_{1}\right)  \tag{28}\\
& \leq\|s\|_{1} \cdot\|\bar{G} \bar{B}\|_{1} \cdot\left(a_{m}+d_{m}\right)
\end{align*}
$$

Thus,

$$
\begin{align*}
\dot{V}= & s^{\mathrm{T}} \dot{s} \leq-k\|s\|_{2}^{2}-\varepsilon\|s\|_{1} \\
& +\left(a_{m}+d_{m}\right) \cdot\|\bar{G} \bar{B}\|_{1} \cdot\|s\|_{1} . \tag{29}
\end{align*}
$$

So, if

$$
\begin{equation*}
\varepsilon \geq\left(a_{m}+d_{m}\right)\|\bar{G} \bar{B}\|_{1}, \tag{30}
\end{equation*}
$$

then

$$
\begin{align*}
\dot{V}= & s^{\mathrm{T}} \dot{s} \leq-k\|s\|_{2}^{2} \\
& -\left[\varepsilon-\left(a_{m}+d_{m}\right)\|\bar{G} \bar{B}\|_{1}\right]\|s\|_{1}<0 . \tag{31}
\end{align*}
$$



Figure 1: The structure of two-link robot manipulator.

This implies that the trajectories of uncertain system (13) will be globally driven onto the specified sliding surface $s=$ 0 in finite time despite of the uncertainties. The proof is completed.

From (22), we have $s(0)=0$; that is, the initial condition is on the sliding surface. According to Theorem 4, uncertain system (13) achieves global sliding mode with the integral sliding surface (22) and the control law (26). So the system designed is globally robust and optimal.

## 5. Application to a Two-Link Robot Manipulator

Trajectory tracking of multilink robot manipulator has received a great deal of attention in recent years. But it is rather difficult to perform excellent tracking because multijoint robot manipulator is a complex system with high nonlinearity, coupling, and time-varying dynamic behavior. To verify the effectiveness and superiority of the proposed method, we apply it to a two-link robot manipulator in comparison with conventional LQRTC.

Consider a two-link robot manipulator shown in Figure 1.
In this figure, $L_{1}$ and $L_{2}$ denote the machine arms, $\tau_{1}$ and $\tau_{2}$ denote the driving torque, $q_{1}$ and $q_{2}$ present the angular displacement of the two joints, respectively, and $A-B$ is the tracking trajectory described by $\left(x_{d}, y_{d}\right)$. The dynamic equation is given by [29]

$$
\begin{equation*}
M(q) \ddot{q}+C(q, \dot{q}) \dot{q}+G(q)+\mathrm{d}(t)=\tau, \tag{32}
\end{equation*}
$$

where $q=\left[\begin{array}{ll}q_{1} & q_{2}\end{array}\right]^{\mathrm{T}}$ is the joint-displacement vector, $\tau=$ $\left[\tau_{1}, \tau_{2}\right]^{\mathrm{T}}$ is the applied joint-torque vector, and $\mathrm{d}(t)$ represents system uncertainties. $M(q), C(q, \dot{q}), G(q)$, and $g$ are defined as follows:

$$
\begin{gathered}
M(q)=\left[\begin{array}{cc}
0.1+0.01 \cos \left(q_{2}\right) & 0.01 \sin \left(q_{2}\right) \\
0.01 \sin \left(q_{2}\right) & 0.1
\end{array}\right], \\
C(q, \dot{q})=\left[\begin{array}{cc}
-0.005 \sin \left(q_{2}\right) \dot{q}_{2} & 0.005 \cos \left(q_{2}\right) \dot{q}_{2} \\
0.005 \cos \left(q_{2}\right) \dot{q}_{2} & 0
\end{array}\right],
\end{gathered}
$$

$$
\begin{gather*}
G(q)=\left[\begin{array}{ll}
0.01 g \cos \left(q_{1}+q_{2}\right) \\
0.01 g \cos \left(q_{1}+q_{2}\right)
\end{array}\right], \\
\mathrm{d}(t)= \begin{cases}{\left[\begin{array}{l}
0 \\
0
\end{array}\right],} & t<5 \mathrm{~s}, \\
{\left[\begin{array}{c}
\sin (0.5 \pi t) \\
10 \sin (3 \pi t)
\end{array}\right],} & t \geq 5 \mathrm{~s} .\end{cases} \tag{33}
\end{gather*}
$$

Suppose the reference signal is given by the following exosystem:

$$
\begin{gather*}
\dot{z}(t)=\left[\begin{array}{cc}
-1 & 4 \\
-5 & -1
\end{array}\right] z(t),  \tag{34}\\
\tilde{y}(t)=\left[\begin{array}{ll}
1 & 1 \\
5 & 5
\end{array}\right] z(t) .
\end{gather*}
$$

Our objective is to design an robust optimal tracking controller, such that the $q_{1}, q_{2}, \dot{q}_{1}$, and $\dot{q}_{2}$ can track $\tilde{y}_{1}, \tilde{y}_{2}, \dot{\tilde{y}}_{1}$, and $\dot{\tilde{y}}_{2}$, respectively. Therefore, a certain given performance criterion can be minimized and the system can exhibit robustness to uncertainties.

Choose a state vector as follows:

$$
\xi=\left[\begin{array}{llll}
\xi_{1} & \xi_{2} & \xi_{3} & \xi_{4}
\end{array}\right]^{\mathrm{T}}=\left[\begin{array}{llll}
q_{1} & \dot{q}_{1} & q_{2} & \dot{q}_{2} \tag{35}
\end{array}\right]^{\mathrm{T}} .
$$

Define $e=\xi-\tilde{y}=\left[q_{1}-\tilde{y}_{1}, \dot{q}_{1}-\dot{\tilde{y}}_{1}, q_{2}-\tilde{y}_{2}, q_{2}-\dot{\tilde{y}}_{2}\right]^{\mathrm{T}}$ and let the control law

$$
\tau=M(q)\left[\begin{array}{l}
v_{1}+\ddot{\tilde{y}}_{1}  \tag{36}\\
v_{2}+\ddot{\dddot{y}}_{2}
\end{array}\right]+C(q, \dot{q}) \dot{q}+G(q) .
$$

So the error state dynamic of the robot manipulator can be written as

$$
\begin{align*}
\dot{e}= & {\left[\begin{array}{llll}
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0
\end{array}\right] e } \\
& +\left[\begin{array}{ll}
0 & 0 \\
1 & 0 \\
0 & 0 \\
0 & 1
\end{array}\right] v-\left[\begin{array}{ll}
0 & 0 \\
1 & 1 \\
0 & 1 \\
1 & 1
\end{array}\right] M^{-1}(q) \mathrm{d}(t) . \tag{37}
\end{align*}
$$

The quadratic performance index is chosen as (18) and the weighting matrices are

$$
\begin{gather*}
Q=\left[\begin{array}{cccc}
100 & 1.5 & 1.5 & 1.5 \\
1.5 & 1 & 1 & 1 \\
1.5 & 1 & 1 & 1 \\
1.5 & 1 & 1 & 1
\end{array}\right],  \tag{38}\\
R=\left[\begin{array}{ll}
0.02 & 0.01 \\
0.01 & 0.01
\end{array}\right] .
\end{gather*}
$$

In order to show the efficiency and the advantage of the proposed approach, a conventional optimal LQRTC for the nominal system and a ROSMTC for the uncertain system


Figure 2: The tracking error $e_{1}(t)$ of position $q_{1}$ for link 1 .


Figure 3: The tracking error $e_{2}(t)$ of velocity $\dot{q}_{1}$ for link 1.
are designed, respectively. For ROSMTC, the sliding-mode surface is chosen in the form of (22) and the control law is chosen in the form of (26) with the designed parameters as follows:

$$
\bar{G}=\left[\begin{array}{cccc}
0 & 1.5 & 0 & 0  \tag{39}\\
0 & 0 & 0 & 1.5
\end{array}\right], \quad k=5, \quad \varepsilon=1.6
$$

With the initial state vectors $\left[\begin{array}{llll}q_{10} & \dot{q}_{10} & q_{20} & \dot{q}_{20}\end{array}\right]^{\mathrm{T}}=$ $\left[\begin{array}{llll}0.5 & 0 & 0.5 & 0\end{array}\right]^{\mathrm{T}}$ and $\left[\begin{array}{ll}z_{10} & z_{20}\end{array}\right]^{\mathrm{T}}=\left[\begin{array}{ll}0.1 & 0.1\end{array}\right]^{\mathrm{T}}$, the simulation results are shown in Figures 2-9.


Figure 4: The tracking error $e_{3}(t)$ of position $q_{2}$ for link 2.


Figure 5: The tracking error $e_{4}(t)$ of velocity $\dot{q}_{2}$ for link 2 .

Figures 2, 3, 4, and 5 show the system responses in the following three cases: LQRTC for system (32) with $\mathrm{d}(t)=0$, LQRTC for system (32) with the given $\mathrm{d}(t)$, and ROSMTC for system (32) with the given $\mathrm{d}(t)$. It can be seen from Figure 2 that when the system is subject to uncertainties, the response of the system with LQRTC deviates from the optimal trajectory, however, the response of the system with ROSMTC is almost the same as that of the nominal system with LQRTC.

The output tracking curves are shown in Figures 6, 7, 8 , and 9. It can be seen that, without external disturbance, the controlled system could track the exosystem output by


Figure 6: The tracking response curve of the position $q_{1}$ for link 1 .


Figure 7: The tracking response curve of the velocity $\dot{q}_{1}$ for link 1.
both controllers at about $t=0.7 \mathrm{~s}$. However, when the external disturbance influences the system at $t=5 \mathrm{~s}$, the output trajectory of LQRTC deviates from the desired trajectory while the tracking performance of ROSMTC is almost not affected. Thus, the ROSMTC provides better features than conventional LQRTC in terms of robustness to system uncertainties.

## 6. Conclusions

A robust optimal tracking control for a class of affine nonlinear MIMO systems with the reference signal given by an


Figure 8: The tracking response curve of the position $q_{2}$ for link 2.


Figure 9: The tracking response curve of the velocity $\dot{q}_{2}$ for link 2.
exosystem has been studied. A linear tracking error equation, with the input and output decoupled, has been established based on the input-output linearization technique. And the nonlinear optimal tracking problem was transformed into a linear LQRTC problem. Moreover, SMC has been used to robustify the LQRTC and a global ROSMTC was realized. That is, the tracking dynamics exhibits global robustness to the uncertainties and the given quadratic performance index can be minimized. The proposed controller was applied to a two-link robot system and simulation results show that good tracking performance can be achieved and global robustness to the uncertainties can be achieved.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant nos. 60940018 and 61104004), and the Shandong Natural Science Foundation (Grant no. ZR2011FQ006).

## References

[1] Y.-J. Pan, H. J. Marquez, and T. Chen, "Robust tracking control for a class of MIMO nonlinear systems with measurable output feedback," International Journal of Robust and Nonlinear Control, vol. 18, no. 1, pp. 69-87, 2008.
[2] Y.-C. Chang, "Robust tracking control for nonlinear MIMO systems via fuzzy approaches," Automatica, vol. 36, no. 10, pp. 1535-1545, 2000.
[3] T.-C. Lee and Z.-P. Jiang, "New cascade approach for global $\kappa$-exponential tracking of underactuated ships," IEEE Transactions on Automatic Control, vol. 49, no. 12, pp. 2297-2303, 2004.
[4] M. Shen and H. Guang, " $H_{2}$ state feedback controller design for continuous Markov jump linear systems with partly known information," International Journal of Systems Science. Principles and Applications of Systems and Integration, vol. 43, no. 4, pp. 786-796, 2012.
[5] F. N. Koumboulis and M. G. Skarpetis, "Output feedback decoupling of linear systems with nonlinear uncertain structure," Journal of the Franklin Institute B, vol. 333, no. 4, pp. 625-629, 1996.
[6] J. L. Neves, B. Heitmann, and N. Khaneja, "Heteronuclear decoupling by optimal tracking," Journal of Magnetic Resonance, vol. 201, no. 11, pp. 7-17, 2009.
[7] Y. S. Zhou and X. R. Zhu, "Decoupling control of the rotor crosscoupling voltage for doubly-fed induction generator based on multiple Degree-Freedom Internal Model Control," in Proceedings of Control and Decision Conference (CCDC '10), pp. 31583163, Yantai, China, 2010.
[8] W. X. Shi, M. Zhang, W. C. Guo, and L. Guo, "Stable adaptive fuzzy control for MIMO nonlinear systems," Computers \& Mathematics with Applications, vol. 62, no. 7, pp. 2843-2853, 2011.
[9] Y.-C. Chang, "An adaptive $H^{\infty}$ tracking control for a class of nonlinear multiple-input-multiple-output (MIMO) systems," IEEE Transactions on Automatic Control, vol. 46, no. 9, pp. 14321437, 2001.
[10] L. Salim and B. M. Seghir, "Indirect fuzzy adaptive control of a class of SISO nonlinear systems," The Arabian Journal for Science and Engineering B, vol. 31, no. 1, pp. 61-74, 2006.
[11] Q. Yang, Z. Yang, and Y. Sun, "Universal neural network control of MIMO uncertain nonlinear systems," IEEE Transactions on Neural Networks, vol. 23, pp. 1162-1169, 2012.
[12] D. X. Gao and Z. Q. Sun, "Fuzzy tracking control design for hypersonic vehicles via T-S model," Science China. Information Sciences, vol. 54, no. 3, pp. 521-528, 2011.
[13] G.-Y. Tang, "Suboptimal control for nonlinear systems: a successive approximation approach," Systems \& Control Letters, vol. 54, no. 5, pp. 429-434, 2005.
[14] J. F. Guo and C. C. Gao, "Variable structure control of linear time-delay uncertain singular system," International Journal of Systems, Control and Communications, vol. 1, no. 3, pp. 364-374, 2009.
[15] S. Bououden, M. Chadli, and H. R. Karimi, "Fuzzy sliding mode controller design using Takagi-Sugeno modelled nonlinear systems," Mathematical Problems in Engineering, vol. 2013, Article ID 734094, 7 pages, 2013.
[16] Y. Q. Xia, J. Chen, G.-P. Liu, L. Wang, and D. Rees, "Robust adaptive sliding mode control for uncertain time-delay systems," International Journal of Adaptive Control and Signal Processing, vol. 23, no. 9, pp. 863-881, 2009.
[17] S.-B. Choi, D.-W. Park, and S. Jayasuriya, "A time-varying sliding surface for fast and robust tracking control of secondorder uncertain systems," Automatica, vol. 30, no. 5, pp. 899904, 1994.
[18] A. Bartoszewicz, "A comment on a time-varying sliding surface for fast and robust tracking control of second-order uncertain systems," Automatica, vol. 31, no. 12, pp. 1893-1895, 1995.
[19] Y. Q. Jin, X. D. Liu, Q. Wei, and C. Hou, "Time-varying sliding mode control for a class of uncertain MIMO nonlinear system subject to control input constraint," Science China. Information Sciences, vol. 53, no. 1, pp. 89-100, 2010.
[20] J.-H. Lee, "Highly robust position control of BLDDSM using an improved integral variable structure system," Automatica, vol. 42, no. 6, pp. 929-935, 2006.
[21] M. Basin, J. Rodriguez-Gonzalez, and L. Fridman, "Optimal and robust control for linear state-delay systems," Journal of the Franklin Institute. Engineering and Applied Mathematics, vol. 344, no. 6, pp. 830-845, 2007.
[22] H. H. Choi, "LMI-based sliding surface design for integral sliding mode control of mismatched uncertain systems," IEEE Transactions on Automatic Control, vol. 52, no. 4, pp. 736-742, 2007.
[23] S. Laghrouche, F. Plestan, and A. Glumineau, "Higher-order sliding mode control based on integral sliding mode," Automatica, vol. 43, no. 3, pp. 531-537, 2007.
[24] H. P. Pang and X. Chen, "Global robust optimal sliding mode control for uncertain affine nonlinear systems," Journal of Systems Engineering and Electronics, vol. 20, no. 4, pp. 838-843, 2009.
[25] H. P. Pang and G. Y. Tang, "Global robust optimal sliding mode control for a class of nonlinear systems with uncertainties," ICIC Express Letters B, vol. 4, no. 6, pp. 2501-2508, 2010.
[26] H. P. Pang, P. Zhang, and N. N. Sun, "Optimal sliding mode output tracking control for linear systems with uncertainties based on observers," Proceedings of the 24th Chinese Control and Decision Conference, pp. 2432-2436, 2012.
[27] G. Y. Tang and M. Q. Fan, "Series-based approximate approach of optimal tracking control for nonlinear systems with timedelay," Progress in Natural Science, vol. 18, no. 12, pp. 1571-1576, 2008.
[28] C. Pukdeboon and A. S. I. Zinober, "Control Lyapunov function optimal sliding mode controllers for attitude tracking of spacecraft," Journal of the Franklin Institute. Engineering and Applied Mathematics, vol. 349, no. 2, pp. 456-475, 2012.
[29] L. Salim and B. M. Seghir, "Global sliding-mode control with generalized sliding dynamics," Asian Journal of Control, vol. 11, no. 4, pp. 449-456, 2009.

## Research Article

# Distributed Filter with Consensus Strategies for Sensor Networks 

Xie Li, Huang Caimou, and Hu Haoji<br>Faculty of Information, Zhejiang University, Hangzhou 310027, China<br>Correspondence should be addressed to Hu Haoji; haoji_hu@zju.edu.cn

Received 12 August 2013; Accepted 28 August 2013
Academic Editor: Baocang Ding
Copyright © 2013 Xie Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Consensus algorithm for networked dynamic systems is an important research problem for data fusion in sensor networks. In this paper, the distributed filter with consensus strategies known as Kalman consensus filter and information consensus filter is investigated for state estimation of distributed sensor networks. Firstly, an in-depth comparison analysis between Kalman consensus filter and information consensus filter is given, and the result shows that the information consensus filter performs better than the Kalman consensus filter. Secondly, a novel optimization process to update the consensus weights is proposed based on the information consensus filter. Finally, some numerical simulations are given, and the experiment results show that the proposed method achieves better performance than the existing consensus filter strategies.


## 1. Introduction

In recent years, there has been a surge of interests in the area of distributed sensor networks. The advantages of distributed sensor networks lie in their low processing power, cheap memory, scalable sensing features, and fault tolerance capabilities.

One of the most basic problems for distributed sensor networks is to develop distributed algorithms [1] for the state estimation of a process of interest. When a process is observed by a group of sensors organized in a network, the goal of each sensor node is to obtain the accurate state estimation for the process. Kalman filtering has been proved to be an effective algorithm for state estimation of dynamic processes [2, 3]. Because of this, most papers focusing on distributed estimation propose different mechanisms by combining the Kalman filter with a consensus filter in order to ensure that the estimates asymptotically converge to the same value, schemes which will be henceforth called consensus based distributed filtering algorithms. Based on the idea mentioned above, a scheme for distributed Kalman filtering (DKF) was proposed in [4] based on reaching an average-consensus [5, 6], and in [7] Olfati-Saber proposed a scalable and distributed Kalman filtering algorithm based on reaching a dynamic average consensus [8]. Olfati-Saber's algorithm [7] has been
further developed by other researchers [9] with similar algorithms. However, methods based on such kind of algorithms produce relatively weak performances. According to [10], the performance is compared with the collective estimation error of $n$ noncooperative local Kalman filters, which is a trivial base performance level for distributed estimation in sensor networks. To solve this, Olfati-Saber developed the Kalman consensus filter (KCF) in [11], where a consensus filter runs directly on the estimator state space variables. In addition, a formal derivation followed by optimality and stability analysis of KCF in discrete-time has been elaborated in [10]. However, in distributed implementations, there is a correlation between local estimates [12] in KCF. In general distributed networks, it is not possible to exactly determine this correlation [13] and it results in nonoptimal local estimates. Other techniques to accomplish distributed estimation for dynamic systems that rely on the inverse covariance filter or information filter have been around for many years [14, 15]. An information consensus filter (ICF) is presented in [16] that applies consensus filters to an information filter. This method does not exactly solve the problem of correlation between local estimates but it gives insight into the statistical effects of the correlation and is working much well in distributed sensor networks. Based on the ICF, we focus on designing the consensus weights to improve its performance.

In this paper, we firstly describe the existing distributed filter with consensus strategies. Then we make an in-depth comparison between the KCF and the ICF. Based on the ICF, we propose the consensus weights optimization for better performance of the system and refer this method as weights optimized information consensus filter (WO-ICF). We show experimentally that the proposed method achieves the best performance and it is closest to the optimal centralized performance.

The structure of this paper is organized as follows. In Section 2, the background knowledge on Kalman filter and the centralized information filter are provided. In Section 3, the consensus strategies are discussed. In Section 4, Kalman consensus filter is presented. In Section 5 we discuss the information consensus filter, and an ICF based weights optimization method is proposed. Simulation results and performance comparisons between the KCF and ICF algorithm are provided in Section 6. In Section 7, we make a brief summary.

## 2. Kalman Filter: Information Form

2.1. Kalman Filter. Consider a dynamic process with the linear time-varying model as follows:

$$
\begin{gather*}
x(k+1)=A(k) x(k)+B(k) w(k) ; \\
x(0) \in N\left(\bar{x}(0), P_{0}\right) \tag{1}
\end{gather*}
$$

where $x(k) \in R^{n}$ and $w(k) \in R^{m}$ are the state and input noise of the process at time $k \in\{0,1,2 \ldots\}$, respectively; $x(0)$ is the initial state with a Gaussian distribution; $A(k)$ is the model matrix, $B(k)$ is the state noise matrix. We are interested in tracking the state of this target by the use of a sensor network with $n$ sensors and the communication topology $G=(V ; E)$.

The observations at sensor $i$ and time $k$ are

$$
\begin{equation*}
z_{i}(k)=H_{i}(k) x(k)+v_{i}(k), \tag{2}
\end{equation*}
$$

where $z_{i}(k) \in R^{p_{i}}$ with $\sum_{i=1}^{n} p_{i}=p, H_{i}(k) \in R^{p_{i} \times n}$ is the local observation matrix for sensor $i$, and $v_{i}(k)$ is the local observation noise. We refer to $z_{i}(k)$ as sensor data. Assume that $w(k)$ and $v_{i}(k)$ are zero mean white Gaussian noise with the following statistics:

$$
\begin{align*}
& E\left[w(k) w(l)^{T}\right]=Q(k) \delta_{k l} \\
& E\left[v_{i}(k) v_{i}(l)^{T}\right]=R_{i}(k) \delta_{k l} \tag{3}
\end{align*}
$$

where $\delta_{r s}=1$ if $r=s$ and $\delta_{r s}=0$, otherwise. We stack the observations at all $n$ sensors in the sensor network to get the global observation model as follows.

Let the global observation vector $\mathbf{z}(k) \in R^{p}$, the global observation matrix $H(k) \in R^{p \times n}$, and the global observation noise vector $v(k) \in R^{n}$ be

$$
\begin{align*}
\mathbf{z}(k) & =\left[z_{1}(k), z_{2}(k), \ldots, z_{n}(k)\right]^{T}, \\
H(k) & =\left[H_{1}(k), H_{2}(k), \ldots, H_{n}(k)\right]^{T}  \tag{4}\\
v(k) & =\left[v_{1}(k), v_{2}(k), \ldots, v_{n}(k)\right]^{T} .
\end{align*}
$$

Then the global observation model is given by

$$
\begin{equation*}
\mathbf{z}(k)=H(k) x(k)+v(k) \tag{5}
\end{equation*}
$$

Since observation noises of different sensors are mutually independent, we can combine $R_{i}(k)$ into one global observation noise covariance matrix $R(k)$ as

$$
\begin{equation*}
R(k)=\operatorname{block} \operatorname{diag}\left[R_{i}(k), \ldots, R_{n}(k)\right] \tag{6}
\end{equation*}
$$

Given the collective information $Z(k)=\{\mathbf{z}(0), \mathbf{z}(1), \ldots$, $\mathbf{z}(k)\}$, the estimation of the state of the process can be expressed as

$$
\begin{align*}
\hat{x}(k):=\widehat{x}(k \mid Z(k)) & =E[x(k) \mid Z(k)], \\
\bar{x}(k):=\widehat{x}(k \mid Z(k-1)) & =E[x(k) \mid Z(k-1)] . \tag{7}
\end{align*}
$$

We refer to $\widehat{x}(k)$ and $\bar{x}(k)$ as estimate and prior estimate (or prediction) of the state $x(k)$, respectively. Then, the error covariance matrices associated with the estimates $\widehat{x}(k)$ and $\bar{x}(k)$ are given by

$$
\begin{gather*}
M(k):=E\left[(\widehat{x}(k)-x(k))(\widehat{x}(k)-x(k))^{T}\right]=E\left[\eta(k) \eta(k)^{T}\right], \\
P(k):=E\left[(\bar{x}(k)-x(k))(\bar{x}(k)-x(k))^{T}\right]=E\left[\bar{\eta}(k) \bar{\eta}(k)^{T}\right], \tag{8}
\end{gather*}
$$

where $\eta(k)=\widehat{x}(k)-x(k)$ and $\bar{\eta}(k)=\bar{x}(k)-x(k)$ denote the estimate errors and $P(0)=P_{0}$. Then, the Kalman filter is a linear estimator in the form

$$
\begin{equation*}
\widehat{x}(k)=\bar{x}(k)+K(k)(\mathbf{z}(k)-H(k) \bar{x}(k)) \tag{9}
\end{equation*}
$$

with the Kalman gain $K(k)$.
Remark 1. Throughout this paper, due to the importance of the node indices, we adopt a notation that is free of the timeindex $k$ and call it an index-free notation to represent all estimators [10]. The index-free form of the above estimator can be written as

$$
\begin{equation*}
\widehat{x}=\bar{x}+K(\mathbf{z}-H \bar{x}) \tag{10}
\end{equation*}
$$

We also use the update operation $\left\{.^{+}\right\}$defined in [10] to rewrite the sensing model of node $i$ of the sensor network as

$$
\begin{equation*}
x^{+}=A x+B w, \quad z_{i}=H_{i} x+v_{i} \tag{11}
\end{equation*}
$$

Then, we get the index-free recursive equations of a centralized Kalman filter (CKF) for system (11):

$$
\begin{gather*}
\hat{x}=\bar{x}+K(\mathbf{z}-H \bar{x}), \\
K=P H^{T}\left(R+H P H^{T}\right)^{-1}, \\
M=P-P H^{T}\left(R+H P H^{T}\right)^{-1} H P  \tag{12}\\
P^{+}=A M A^{T}+B Q B^{T} \\
\bar{x}^{+}=A \widehat{x}
\end{gather*}
$$

2.2. Information Form: Centralized Information Filter. Using the matrix inversion lemma

$$
\begin{align*}
\left(I+A B C^{-1} B^{T}\right)^{-1} A & =\left(A^{-1}+B C^{-1} B^{T}\right)^{-1} \\
& =A-A B\left(B^{T} A B+C\right)^{-1} B^{T} A \tag{13}
\end{align*}
$$

By use of the identity $B^{T} A B C^{-1}=\left[\left(B^{T} A B+C\right) C^{-1}-I\right]$, we have

$$
\begin{align*}
\left(I+A B C^{-1} B^{T}\right)^{-1} A B C^{-1} & =\left(A^{-1}+B C^{-1} B^{T}\right)^{-1} B C^{-1}  \tag{14}\\
& =A B\left(B^{T} A B+C\right)^{-1}
\end{align*}
$$

From (14), we have

$$
\begin{equation*}
K=P H^{T}\left(R+H P H^{T}\right)^{-1}=\left(P^{-1}+H^{T} R^{-1} H\right)^{-1} H^{T} R^{-1} \tag{15}
\end{equation*}
$$

Using the matrix inversion lemma

$$
\begin{equation*}
(A+B C D)^{-1}=A^{-1}-A^{-1} B\left(C^{-1}+D A^{-1} B\right)^{-1} D A^{-1} \tag{16}
\end{equation*}
$$

we have

$$
\begin{equation*}
\left(P^{-1}+H^{T} R^{-1} H\right)^{-1}=P-P H^{T}\left(R+H P H^{T}\right)^{-1} H P=M \tag{17}
\end{equation*}
$$

Then we can rewrite (15) as

$$
\begin{equation*}
K=M H^{T} R^{-1} \tag{18}
\end{equation*}
$$

Based on the above derivation, the recursive equations of the Kalman filter can be rewritten as

$$
\begin{gather*}
\widehat{x}=\bar{x}+K(\mathbf{z}-H \bar{x}), \\
K=M H^{T} R^{-1}, \\
M=\left(P^{-1}+H^{T} R^{-1} H\right)^{-1},  \tag{19}\\
P^{+}=A M A^{T}+B Q B^{T}, \\
\bar{x}^{+}=A \widehat{x} .
\end{gather*}
$$

From (19), the estimate $\hat{x}$ can be expressed as

$$
\begin{align*}
\widehat{x} & =\bar{x}+K(\mathbf{z}-H \bar{x})=\bar{x}+M H^{T} R^{-1}(\mathbf{z}-H \bar{x})  \tag{20}\\
& =\bar{x}+M\left(H^{T} R^{-1} \mathbf{z}-H^{T} R^{-1} H \bar{x}\right)
\end{align*}
$$

Now we define the $n$-dimensional global observation variables as

$$
\begin{equation*}
y=H^{T} R^{-1} \mathbf{z}, \quad S=H^{T} R^{-1} H \tag{21}
\end{equation*}
$$

and the $n$-dimensional local observation variables at sensor $i$ as

$$
\begin{equation*}
y_{i}=H_{i}^{T} R_{i}^{-1} z_{i}, \quad S_{i}=H_{i}^{T} R_{i}^{-1} H_{i} \tag{22}
\end{equation*}
$$

When the observations are distributed among the sensors, the KF can be implemented by collecting all the sensor observations at a central location, or with observation fusion by realizing that the global observation variables in (21), it can be written as

$$
\begin{equation*}
\mathbf{y}=H^{T} R^{-1} \mathbf{z}=H_{1}^{T} R_{1}^{-1} \mathbf{z}_{1}+\cdots+H_{N}^{T} R_{N}^{-1} \mathbf{z}_{N}=\sum_{i=1}^{N} y_{i} \tag{23}
\end{equation*}
$$

Similarly,

$$
\begin{equation*}
S=H^{T} R^{-1} H=H_{1}^{T} R_{1}^{-1} H_{1}+\cdots+H_{N}^{T} R_{N}^{-1} H_{N}=\sum_{i=1}^{N} S_{i} . \tag{24}
\end{equation*}
$$

Recall (20) where the estimate $\hat{x}$ could be written as

$$
\begin{equation*}
\widehat{x}=\bar{x}+M(\mathbf{y}-S \bar{x}) . \tag{25}
\end{equation*}
$$

Multiplication on the left by $M^{-1}$ yields a variation of (25) as

$$
\begin{equation*}
M^{-1} \widehat{x}=M^{-1} \bar{x}+\mathbf{y}-S \bar{x}=\left(P^{-1}+S\right) \bar{x}+\mathbf{y}-S \bar{x}=P^{-1} \bar{x}+\mathbf{y} \tag{26}
\end{equation*}
$$

Let the inverse of $M$ and $P$ be the information matrices, $\hat{I}$ and $\bar{I}$. Let $\hat{i}$ and $\bar{i}$ be the information vectors. We have the following relations:

$$
\begin{array}{cl}
\hat{I}=M^{-1}, & \bar{I}=P^{-1} \\
\hat{i}=M^{-1} \widehat{x}, & \bar{i}=P^{-1} \bar{x} \tag{27}
\end{array}
$$

Then (26) can be rewritten as

$$
\begin{equation*}
\widehat{i}=\bar{i}+\mathbf{y} \tag{28}
\end{equation*}
$$

and the update of predicted estimate $\bar{x}$ can be expressed as

$$
\begin{equation*}
\bar{x}^{+}=A \widehat{x}=A(M \widehat{I}) \widehat{x}=A M(\hat{I} \hat{x})=A \widehat{I}^{-1} \hat{i} \tag{29}
\end{equation*}
$$

and we have

$$
\begin{equation*}
\bar{i}^{+}=\left(P^{-1} \bar{x}\right)^{+}=\bar{I}^{+} A \widehat{I}^{-1} \widehat{i} \tag{30}
\end{equation*}
$$

Now we get the following simpler form of the filter in (19) and call it centralized information filter (CIF):

$$
\begin{gather*}
\hat{I}=\bar{I}+S, \quad \hat{i}=\bar{i}+\mathbf{y} \\
\bar{I}^{+}=\left(A \widehat{I}^{-1} A^{T}+B Q B^{T}\right)^{-1}, \quad \bar{i}^{+}=\bar{I}^{+} A \widehat{I}^{-1} \widehat{i} \tag{31}
\end{gather*}
$$

where (31) is the filter step and prediction step (or update step) of the CIF, respectively.

## 3. Consensus Strategy

Consensus strategy defines a set of rules for a team of agents to agree on specific consensus states. With these rules each agent exchanges information with its neighboring agents and finally reaches an agreement (or consensus) concerning the consensus state over time [17, 18]. Furthermore, average
consensus occurs when the final consensus state is the average of the initial values.

Consider a team of $n$ agents to agree on specific consensus states, and at any discrete-time instant $\tau$, the communication topology between $n$ agents can be described by the graph $G[\tau]=(V, E[\tau])$, the graph $G$ is undirected, $V=\{1,2, \ldots, n\}$ is the vertex set, and $E[\tau] \subseteq V \times V$ is the edge set. In the consensus algorithm, each agent in the network maintains a local copy of the consensus state $\zeta_{i} \in R^{n}$ and updates this value using its neighbors' consensus states according to the rule:

$$
\begin{equation*}
\zeta_{i}[\tau+1]=\zeta_{i}[\tau]+\sum_{j=1}^{n} \beta_{i j}[\tau]\left(\zeta_{j}[\tau]-\zeta_{i}[\tau]\right) \tag{32}
\end{equation*}
$$

where $\tau$ indicates the consensus filter iteration step. To choose the weights $\beta_{i j}[\tau]$, we can use the Maximum-degree weights or the Metropolis weights [19]. Here we use the latter which preserves the averaging in consensus filters and can be computed by

$$
\beta_{i j}[\tau]= \begin{cases}\left(1+\max \left\{d_{i}[\tau], d_{j}[\tau]\right\}\right)^{-1} & \text { if }(i, j) \in E[\tau]  \tag{33}\\ 1-\sum_{(i, l) \in E[\tau]} \beta_{i l}[\tau] & \text { if } i=j \\ 0 & \text { otherwise }\end{cases}
$$

where $d_{i}[\tau]$ is the degree of agent $i$ in the graph $G[\tau]$. Arrange the local consensus states into the vector $\zeta[\tau]=$ $\left[\zeta_{1}^{T}[\tau], \ldots, \zeta_{n}^{T}[\tau]\right]^{T}$, and define the matrix $(B[\tau])_{i j}=\beta_{i j}[\tau]$ for $i \neq j$; otherwise $(B[\tau])_{i i}=1-\sum_{(i, l) \in E[\tau]} \beta_{i l}[\tau]$, and we can rewrite the update in (32) as

$$
\begin{equation*}
\zeta[\tau+1]=(B[\tau] \otimes I) \zeta[\tau] \tag{34}
\end{equation*}
$$

where $I$ is the appropriate size identity matrix and $\otimes$ denotes the matrix Kronecker product.

The $i j$ th element of $B[\tau]$ in (34) satisfies the following four conditions: (1) $(B[\tau])_{i j} \geq 0$, (2) $\sum_{i}(B[\tau])_{i j}=1$, (3) $\sum_{j}(B[\tau])_{i j}=1,(4)$ and each nonzero entry is both uniformly upper and lower bounded. Based on these conditions, we have the following results [20] for average consensus.

Lemma 2. Under switching interaction topologies, if there exists a finite $T \geq 0$ that for every interval $[\tau, \tau+T]$ the union of the interaction graph across interval is strongly connected, then consensus protocol (34) achieves average consensus asymptotically; that is, $\zeta_{i}[\tau] \rightarrow(1 / n) \sum_{i=1}^{n} \zeta_{i}[0]$ as $\tau \rightarrow \infty$.

Remark 3. In order to calculate the metropolis weights in (33), we assume undirected communication throughout this paper. Therefore, if the graph $G[\tau]$ is connected, the matrix $B[\tau]$ is a doubly stochastic matrix, and the four conditions on $B[\tau]$ are satisfied. This implies that average consensus is achieved asymptotically as long as every graph is connected [21].

## 4. Distributed Kalman Filter: Consensus on Estimate

In this section, we discuss an alternative approach to distribute the Kalman filtering that relies on communicating state estimates between neighboring nodes and refer to it as Kalman consensus filter (KCF). Before presenting the KCF algorithm, we first need to discuss a more primitive DKF algorithm called local Kalman filter (LKF) which forms the basis of the KCF.
4.1. Local Kalman Filter. In local Kalman filtering, let $N_{i}=$ $\{j:(i, j) \in E\}$ be the set of neighbors of node $i$ on graph $G$. Each node $i$ of the sensor network communicates its measurement $z_{i}$, covariance information $R_{i}$, and observation matrix $H_{i}$ with its neighbors $N_{i}$. For node $i$, we assume that the information flow from nonneighboring nodes to node $i$ is prohibited if there is no nodes except for its neighbors $N_{i}$ exist. Therefore, node $i$ can use a central Kalman filter that only utilizes the observation vectors and observation matrices of the nodes in $J_{i}=N_{i} \cup\{i\}$ [11]. This leads to the following primitive DKF algorithm with no consensus on state estimation.

LKF Iterations. Assume that node $i$ only receives information from its neighbors. Then, we have the iterations of node $i$ in local Kalman filtering as

$$
\begin{gather*}
y^{i}=\sum_{j \in J_{i}} H_{j}^{T} R_{j}^{-1} z_{j}=\sum_{j \in J_{i}} y_{j}, \\
S^{i}=\sum_{j \in J_{i}} H_{j}^{T} R_{j}^{-1} H_{j}=\sum_{j \in J_{i}} S_{j}, \\
\widehat{x}_{i}=\bar{x}_{i}+M_{i}\left(y^{i}-S^{i} \bar{x}\right),  \tag{35}\\
M_{i}=\left(P_{i}^{-1}+S^{i}\right)^{-1}, \\
P_{i}^{+}=A M_{i} A^{T}+B Q B^{T}, \\
\bar{x}_{i}^{+}=A \widehat{x}_{i},
\end{gather*}
$$

where $y^{i}$ and $S^{i}$ are local aggregate information vector and matrix, respectively and node $i$ locally computes both $y^{i}$ and $S^{i}$.
4.2. Kalman Consensus Filter. We now present the Kalman consensus filter (KCF). The KCF uses consensus strategy (32) on the state estimate in a distributed Kalman filter, where each node maintains a local Kalman filter. Corresponding to (32), let $\zeta_{i}[\tau]=\bar{x}_{i}$ be the prior estimate at time $\tau$ and $\zeta_{i}[\tau+1]=\bar{x}_{i}{ }^{c}$ be fused prior estimate; each node fuses the prior estimates from its neighbors according to the rule:

$$
\begin{equation*}
\bar{x}_{i}^{c}=\bar{x}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{x}_{j}-\bar{x}_{i}\right) . \tag{36}
\end{equation*}
$$

Initialization (for node $i$ ):

$$
\begin{array}{ll}
\bar{x}_{i}=x(0) & P_{i}=P_{0} \\
\tau=1 & \tau_{p}=\tau+T_{p}
\end{array}
$$

Loop \{Local iteration on node $i$ \}
(1) Consensus update

$$
\begin{aligned}
\bar{x}_{i}^{c} & =\bar{x}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{x}_{j}-\bar{x}_{i}\right) \\
y^{i} & =\sum_{j \in J_{i}} H_{j}^{T} R_{j}^{-1} z_{j} \quad S^{i}=\sum_{j \in J_{i}} H_{j}{ }^{T} R_{j}^{-1} H_{j} \quad \tau \leftarrow \tau+1
\end{aligned}
$$

(2) If new observations are taken then the Kalman consensus state estimate are computed

$$
\begin{gathered}
\widehat{x}_{i}=\bar{x}_{i}^{c}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}^{c}\right)=\bar{x}_{i}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}\right)+\left(I-M_{i} S^{i}\right) \sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{x}_{j}-\bar{x}_{i}\right) \\
M_{i}=\left(P_{i}^{-1}+S^{i}\right)^{-1}
\end{gathered}
$$

(3) If time for a predication step (i.e., $\tau=\tau_{p}$ ) then prediction step

$$
\begin{aligned}
& P_{i}^{+}=A M_{i} A^{T}+B Q B^{T} \\
& \bar{x}_{i}^{+}=A \widehat{x}_{i} \\
& \tau_{p}=\tau+T_{p}
\end{aligned}
$$

Algorithm 1: Kalman consensus filter.

Using the fused prior estimate $\bar{x}_{i}^{c}$, the filter estimate at node $i$ could be implemented by

$$
\begin{align*}
\widehat{x}_{i} & =\bar{x}_{i}^{c}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}^{c}\right) \\
& =\bar{x}_{i}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}\right)+\left(I-M_{i} S^{i}\right) \sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{x}_{j}-\bar{x}_{i}\right) . \tag{37}
\end{align*}
$$

The local KCF is summarized in Algorithm 1, where $\tau$ is the time index for the consensus strategy and $T_{p} \in Z^{+}$is the time interval between prediction updates. One-time step $k-1 \rightarrow k$ is equivalent to $T_{p}$ time steps of the consensus time index $\tau \rightarrow \tau+1$; that is, for each node, the information exchanges between neighboring nodes occurred faster than the prediction update step. The three steps in KCF prediction, local filter estimate, and consensus update are not necessarily sequential.

The last term in (37) is the correction of filter estimate $\widehat{x}_{i}$ compared to the standard Kalman estimator. Intuitively, adding the consensus term in (37) will force local estimators to reach a consensus regarding state estimates. The structure of node $i$ in the KCF algorithm is shown in Figure 1.

In [11], the author proposed the following Kalman consensus estimator which is in the form of

$$
\begin{equation*}
\widehat{x}_{i}=\bar{x}_{i}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}\right)+C_{i} \sum_{j \in N_{i}}\left(\bar{x}_{j}-\bar{x}_{i}\right), \tag{38}
\end{equation*}
$$

where $C_{i}$ is named consensus gain of node $i$. The choice of the consensus gain $C_{i}$ is free. A poor choice of $C_{i}$ leads to either the lack of consensus on estimates (e.g., setting $C_{i}=0$, for all $i$ ) or the lack of stability of the error dynamics of the filter. One possible choice is to let

$$
\begin{equation*}
C_{i}=\gamma P_{i}=\varepsilon \frac{P_{i}}{1+\left\|P_{i}\right\|_{F}} \tag{39}
\end{equation*}
$$

where $\varepsilon>0$ is a relative small constant and $\|\cdot\|_{F}$ denotes the Frobenius norm of a matrix. The derivation of the optimal


Figure 1: The algorithm structure of node $i$ in the KCF.

Kalman consensus filter can be found in [10], where we can also find that the computational complexity of updating the error covariance $P_{i}^{+}=A M_{i} A^{T}+B Q B^{T}$ of the optimal Kalman consensus filter is not scalable in $n$. To obtain a suboptimal approximation of the KCF which is distributed and scalable in $n$, we make an assumption that the consensus gains $C_{i}=O(\varepsilon)$ are of the order of $\varepsilon$. Then we get the stable suboptimal KCF summarized in Algorithm 2.

## 5. Distributed Filter: Consensus on Information Matrix

The KCF discussed previously applies consensus strategy on the prior estimate to the Kalman filter and improves the state estimate of KF. However, the error covariance matrix $M_{i}$ is not improved because each node in KCF only fuses the prior estimates from its neighbors but neglects the helpful information about the error covariance matrix. In the next section, we adopt an information matrix weighted consensus strategy to improve the consensus based distributed Kalman filter algorithm in the estimation fusion of sensor networks. We refer to this method as information consensus filter (ICF) [16]. Before presenting the ICF algorithm, we need to discuss a more primitive DKF algorithm called local information filter (LIF) which forms the basis of the ICF.

## Initialization:

$$
\bar{x}_{i}=x(0), P_{i}=P_{0}, \text { and message } m_{j}=\left\{y_{j}, S_{j}, \bar{x}_{j}\right\}
$$

While new data exists do
(1) Compute local observation vector and matrix of node $i$ :

$$
y_{i}=H_{i}^{T} R_{i}^{-1} z_{i} \quad S_{i}=H_{i}^{T} R_{i}^{-1} H_{i}
$$

(2) Broadcast message $m_{i}=\left\{y_{i}, S_{i}, \bar{x}_{i}\right\}$ to neighbors.
(3) Receive messages from all neighbors.
(4) Compute the local aggregate information vector and matrix:

$$
y^{i}=\sum_{j \in J_{i}}^{\infty} y_{j} \quad S^{i}==\sum_{j \in J_{i}} S_{j}
$$

(5) Compute the Kalman consensus state estimate

$$
\begin{aligned}
& \widehat{x}_{i}=\bar{x}_{i}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}\right)+\gamma P_{i} \sum_{j \in N_{i}}\left(\bar{x}_{j}-\bar{x}_{i}\right) \\
& M_{i}=\left(P_{i}^{-1}+S^{i}\right)^{-1} \\
& \gamma=\frac{\varepsilon}{1+\left\|P_{i}\right\|}, \quad\|X\|=\operatorname{tr}\left(X^{T} X\right)^{1 / 2}
\end{aligned}
$$

(6) Update the state of the Kalman consensus filter:

$$
\begin{aligned}
& P_{i}^{+}=A M_{i} A^{T}+B Q B^{T} \\
& \bar{x}_{i}^{+}=A \widehat{x}_{i}
\end{aligned}
$$

## End While

Algorithm 2: Suboptimal Kalman consensus filter: DKF Algorithm with an estimator that has a rigorously derived consensus term (message passing during one time cycle for node).
5.1. Local Information Filter. To distribute the estimation of the global state vector, $x$ in CIF, we implement local information filter (LIF) at each sensor $i$, which is based on the sensor model (11) and can be derived from local Kalman filter (LKF) in (35) after we use its information form. Each LIF computes local objects (matrices and vectors) which are then fused (if required) by exchanging information among the neighbors. In LIF, there is no centralized knowledge of the estimation of the global state that exists in CIF; however, it can be obtain by fusing the local state vector.

Let the inverses of $M_{i}$ and $P_{i}$ be the local information matrices, $\widehat{I}_{i}$ and $\bar{I}_{i}$. Let $\hat{i}_{i}$ and $\bar{i}_{i}$ be the local information vector. We have the following relations:

$$
\begin{array}{cl}
\widehat{I}_{i}=M_{i}^{-1}, & \bar{I}_{i}=P_{i}^{-1}  \tag{40}\\
\hat{i}_{i}=M_{i}^{-1} \widehat{x}_{i}, & \bar{i}_{i}=P_{i}^{-1} \bar{x}_{i} .
\end{array}
$$

Then we get the following simpler form of the filter in (35).

LIF Iterations. The local information filtering iterations for node $i$ are in the form

$$
\begin{gather*}
\widehat{I}_{i}=\bar{I}_{i}+S^{i}, \quad \hat{i}_{i}=\bar{i}_{i}+y^{i}, \\
\bar{I}_{i}^{+}=\left(A \widehat{I}_{i}^{-1} A^{T}+B Q B^{T}\right)^{-1}, \quad \bar{i}_{i}^{+}=\bar{I}_{i}^{+} A \widehat{I}_{i}^{-1} \widehat{i}_{i} . \tag{41}
\end{gather*}
$$

5.2. Information Consensus Filter. We now present the information consensus filter (ICF). The ICF uses consensus strategy (32) on both the information state and the information matrix in a distributed Kalman filter, where each node maintains a local information filter. Recalling (32), let $\bar{i}_{i}^{c}$ be the fused local information vector and $\bar{I}_{i}^{c}$ the fused local
information matrix; each node fuses the local information from its neighbors according to the rule:

$$
\begin{align*}
& \bar{i}_{i}^{c}=\bar{i}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{i}_{j}-\bar{i}_{i}\right), \\
& \bar{I}_{i}^{c}=\bar{I}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{I}_{j}-\bar{I}_{i}\right) . \tag{42}
\end{align*}
$$

Using the fused local information vector and matrix, $\bar{i}_{i}^{c}$ and $\bar{I}_{i}{ }^{c}$, the local ICF is summarized in Algorithm 3.
5.3. Local Information Filter. Now we make a comparison between ICF and KCF based on the state estimate $\widehat{x}_{i}$ and the error covariance matrix $M_{i}$. Let $\beta_{i i}[\tau]=1-\sum_{j \in N_{i}} \beta_{i j}[\tau]$; we can rewrite (42) as

$$
\begin{align*}
& \bar{i}_{i}^{c}=\beta_{i i}[\tau] \bar{i}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau] \bar{i}_{j} \\
& \bar{I}_{i}^{c}=\beta_{i i}[\tau] \bar{I}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau] \bar{I}_{j} \tag{43}
\end{align*}
$$

Then we have

$$
\begin{align*}
\bar{x}_{i}^{c} & =\left(\bar{I}_{i}^{c}\right)^{-1} \bar{i}_{i}^{c}=\left(\bar{I}_{i}^{c}\right)^{-1}\left[\beta_{i i}[\tau] \bar{i}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau] \bar{i}_{j}\right]  \tag{44}\\
& =W_{i i} \bar{x}_{i}+\sum_{j \in N_{i}} W_{i j} \bar{x}_{j} .
\end{align*}
$$

Here we use $\bar{i}_{i}=\bar{I}_{i} \bar{x}_{i}, i \in J_{i}$ and $W_{i l}=\beta_{i l}[\tau]\left(\bar{I}_{i}^{c}\right)^{-1} \bar{I}_{l}, l \in J_{i}$. Then we finally get

$$
\begin{aligned}
\widehat{x}_{i} & =\bar{x}_{i}^{c}+M_{i}\left(y^{i}-S^{i} \bar{x}_{i}^{c}\right) \\
& =W_{i i} \bar{x}_{i}+M_{i}\left(y^{i}-S^{i} W_{i i} \bar{x}_{i}\right)+\left(I-M_{i} S^{i}\right) \sum_{j \in N_{i}} W_{i j} \bar{x}_{j}
\end{aligned}
$$

Initialization (for node $i$ ):

$$
\begin{array}{ll}
\bar{i}_{i}=\bar{i}(0) & \\
\bar{I} & \bar{I}_{i}=\bar{I}(0) \\
\tau=1 & \\
\tau_{p}=\tau+T_{p}
\end{array}
$$

Loop \{Local iteration on node $i\}$
(1) Consensus update

$$
\begin{aligned}
& \left.\left.\bar{i}_{i}^{c}=\bar{i}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\right] \bar{i}_{j}-\bar{i}_{i}\right) \\
& \bar{I}_{i}^{c}=\bar{I}_{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\overline{I_{j}}-\bar{I}_{i}\right) \\
& y^{i}=\sum_{j \in H_{i}} H_{j}{ }^{T} R_{j}{ }^{-1} z_{j} \quad S^{i}=\sum_{j \in \epsilon_{i}} H_{j}{ }^{T} R_{j}{ }^{-1} H_{j} \quad \tau \leftarrow \tau+1
\end{aligned}
$$

(2) If new observations are taken then the information consensus estimate are computed

$$
\begin{aligned}
& \hat{i}_{i}=\bar{i}_{i}^{c}+y^{i}=\bar{i}_{i}+y^{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{i}_{j}-\bar{i}_{i}\right) \\
& \widehat{I}_{i}=\bar{I}_{i}^{c}+S^{i}=\bar{I}_{i}+S^{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{I}_{j}-\bar{I}_{i}\right)
\end{aligned}
$$

(3) If time for a predication step (i.e., $\tau=\tau_{p}$ ) then prediction step

$$
\begin{aligned}
& \bar{I}_{i}^{+}=\left(A \widehat{I}_{i}^{-1} A^{T}+B Q B^{T}\right)^{-1} \\
& \bar{i}_{i}^{+}=\bar{I}_{i}^{+} A \widehat{I}_{i}^{-1} \hat{i}_{i} \\
& \tau_{p}=\tau+T_{p}
\end{aligned}
$$

End Loop

Algorithm 3: Information consensus filter.

$$
\begin{align*}
= & \left(\left(\bar{I}_{i}^{c}\right)^{-1} \bar{I}_{i}\right) \bar{x}_{i}+M_{i}\left[y^{i}-S^{i}\left(\left(\bar{I}_{i}^{c}\right)^{-1} \bar{I}_{i}\right) \bar{x}_{i}\right] \\
& +\left(I-M_{i} S^{i}\right) \sum_{j \in N_{i}} \beta_{i j}[\tau]\left[\left(\bar{I}_{i}^{c}\right)^{-1} \bar{I}_{j} \bar{x}_{j}-\left(\bar{I}_{i}^{c}\right)^{-1} \bar{I}_{i} \bar{x}_{i}\right],  \tag{47}\\
M_{i}= & \left(\widehat{I}_{i}\right)^{-1}=\left[\bar{I}_{i}+S^{i}+\sum_{j \in N_{i}} \beta_{i j}[\tau]\left(\bar{I}_{j}-\bar{I}_{i}\right)\right]^{-1},
\end{align*}
$$

$$
\bar{i}_{i, a v}=\frac{1}{1+d_{i}}\left(\bar{i}_{i}+\sum_{j \in N_{i}} \bar{i}_{j}\right)
$$

$\alpha_{i}^{1}, \alpha_{i}^{2}$ are the weight coefficients, and $\alpha_{i}^{1}+\alpha_{i}^{2}=1\left(0<\alpha_{i}^{1}<\right.$ $1,0<\alpha_{i}^{2}<1$ ).

The first term in the objective function $F_{i}$ is used to assess the prior estimate error covariance of node $i$ after fusing the local information of its neighbors, and the second term is used to assess the consensus of the fused local information vector in node $i$ and the prior estimates of its neighbors. Base on the objective function $F_{i}$, the consensus weights optimization problem can be described as

$$
\begin{align*}
\boldsymbol{\beta}_{i}^{*} & =\underset{\boldsymbol{\beta}_{i}}{\arg \min } F_{i} \\
\text { s.t. } \beta_{i j} & \geq 0, \quad(i, j) \in E[\tau]  \tag{48}\\
\beta_{i j} & =0, \quad(i, j) \notin E[\tau] \\
\left\|\boldsymbol{\beta}_{i}\right\|_{1} & =1,
\end{align*}
$$

where $\boldsymbol{\beta}_{i}=\left[\beta_{i 1}, \beta_{i 2}, \ldots, \beta_{i n}\right]$. To solve the optimization problem (48), we only need the local information of node $i$ and that of its neighbors. We refer to this method as weights optimized information consensus filter (WO-ICF).

## 6. Numerical Simulations

Let the linear system under consideration be represented by a second-order discrete time-varying model:

$$
\begin{equation*}
x(k+1)=A x(k)+B w, \tag{49}
\end{equation*}
$$



Figure 2: A sensor network with 20 nodes and 51 links.
where $A=I_{2}+\delta A_{0}+\left(\delta^{2} / 2\right) A_{0}^{2}+\left(\delta^{3} / 6\right) A_{0}^{3}$ with $A_{0}=2\left[\begin{array}{ll}0 & 1 \\ 1 & 0\end{array}\right]$, $\delta=0.015$ and $B=\delta B_{0}$ with $B_{0}=I_{2}, Q=25 I_{2}$. The initial conditions are $x_{0}=[15,-10], P_{0}=20 I_{2}$. A sensor network with 20 randomly located nodes is used in this experiment (see Figure 2). The local observation matrix for sensor $i$ is $H_{i}=\left[\begin{array}{ll}0 & 1 \\ 1 & 0\end{array}\right]$, and the local observation noise covariance is $R_{i}=100 I_{2}$ for $i \leq 10$ otherwise $R_{i}=3000 I_{2}$.

Define the averaged estimation error $E(k)$ and the averaged consistency estimation error $D(k)$ as the algorithm performance metrics, which can be computed as follows:

$$
\begin{align*}
& E(k)=\sqrt{\frac{1}{n} \sum_{i=1}^{n}\left(\widehat{x}_{i}(k)-x(k)\right)^{T}\left(\widehat{x}_{i}(k)-x(k)\right)},  \tag{50}\\
& D(k)=\sqrt{\frac{1}{n} \sum_{i=1}^{n}\left(\widehat{x}_{i}(k)-\widehat{x}_{a v}(k)\right)^{T}\left(\widehat{x}_{i}(k)-\widehat{x}_{a v}(k)\right)},
\end{align*}
$$

where $\widehat{x}_{a v}(k)=(1 / n) \sum_{i=1}^{n} \widehat{x}_{i}(k)$ is the averaged estimation of state.

Figure 3 demonstrates the averaged estimation error using different algorithms. We can see that the ICF and the WO-ICF behave in a similar manner (with comparable performances), and the averaged estimation accuracy in ICF and WO-ICF is improved compared to KCF. After 50 iterations, their performances are very close to CKF; this is because the average consensus is achieved after constantly information exchanging, fusing, and filtering.

Figure 4 shows that our WO-ICF performs significantly better than both KCF and ICF, it was the fastest converged, and the consistency of estimates between different nodes in WO-ICF was improved by optimizing the consensus weights.

Figure 5 demonstrates the comparisons of different algorithms on the traces of the averaged estimation covariance matrices $\operatorname{tr}\left((1 / n) \sum_{i=1}^{n} M_{i}(k)\right)$. A quick look at Figure 5 reveals that both ICF and our WO-ICF perform significantly better than KCF, of which the reason is that the information matrix weighted consensus strategy is adopted. Furthermore, by optimizing the consensus weights, the error covariance matrix $M_{i}$ is improved significantly compared to ICF.


Figure 3: The averaged estimation errors of different algorithms.


Figure 4: The averaged consistency estimation errors of different algorithms.

## 7. Conclusions

In this paper, a description about the existing distributed Filter with consensus strategies is presented, including Kalman consensus filter (KCF) and information consensus filter (ICF). In addition, an in-depth comparison between the KCF and the ICF is made. Based on the ICF, the weights optimized information consensus filter (WO-ICF) is proposed to optimize the consensus weights. Simulation shows that both ICF and WO-ICF perform better than KCF; they improve not only the state estimate but also the error covariance matrix, and the proposed WO-ICF achieves better consistency estimation performance than ICF. Compared with the existing consensus filter, our WO-ICF achieves the best performance and is closest to the optimal centralized performance.


Figure 5: The traces of the averaged estimation covariance matrices of different algorithms.

## Acknowledgments

This work is supported in part by NSFC (Natural Science Foundation of China) Projects 61202400, Natural Science Foundation of Zhejiang Q12F020078, Shaoxing Project of Science and Technology 2011A22013, and Wenzhou Project of Science and Technology H20090054, S20100029, and H20100095.

## References

[1] N. A. Lynch, Distributed Algorithms, Morgan Kaufmann, 1997.
[2] B. D. O. Anderson and J. B. Moore, Optimal Filtering, Prentice Hall, Englewood Cliffs, NJ, USA, 1979.
[3] A. E. Bryson and Y. C. Ho, Applied Optimal Control: Optimization, Estimation, and Control, Hemisphere, New York, NY, USA, 1975.
[4] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Approximate distributed kalman filtering in sensor networks with quantifiable performance," in Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN '05), pp. 133-139, April 2005.
[5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," IEEE Transactions on Automatic Control, vol. 49, no. 9, pp. 15201533, 2004.
[6] R. Olfati Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in Proceedings of the American Control Conference, vol. 2, pp. 951-956, June 2003.
[7] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference (CDC-ECC '05), pp. 8179-8184, December 2005.
[8] D. Spanos, R. Olfati-Saber, and R. M. Murray, "Dynamic consensus on mobile networks," in Proceedings of the 16th International Federation of Automatic Control World Congress, Prague, Czech, 2005.
[9] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," IEEE Journal on Selected Areas in Communications, vol. 26, no. 4, pp. 622633, 2008.
[10] R. Olfati-Saber, "Kalman-Consensus filter: optimality, stability, and performance," in Proceedings of the 48th IEEE Conference on Decision and Control held jointly with 28th Chinese Control Conference (CDC/CCC '09), pp. 7036-7042, Shanghai, China, December 2009.
[11] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in Proceedings of the 46th IEEE Conference on Decision and Control (CDC '07), pp. 5492-5498, December 2007.
[12] Y. Bar-Shalom, "On the track-to-track correlation problem," IEEE Transactions on Automatic Control, vol. 26, no. 2, pp. 571572, 1981.
[13] S. Utete and H. F. Durrant-Whyte, "Reliability in decentralized data fusion networks," in Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI '94), pp. 215-221, October 1994.
[14] S. Grime and H. F. Durrant-Whyte, "Data fusion in decentralized sensor networks," Control Engineering Practice, vol. 2, no. 5, pp. 849-863, 1994.
[15] A. G. O. Mutambara, Decentralized Estimation and Control For Multisensor Systems, CRC Press, Boca Raton, Fla, USA, 1998.
[16] D. W. Casbeer and R. Beard, "Distributed information filtering using consensus filters," in Proceedings of the American Control Conference (ACC '09), pp. 1882-1887, Hyatt Regency Riverfront, St. Louis, Mo, USA, June 2009.
[17] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control," IEEE Control Systems Magazine, vol. 27, no. 2, pp. 71-82, 2007.
[18] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," Proceedings of the IEEE, vol. 95, no. 1, pp. 215-233, 2007.
[19] L. Xiao, S. Boyd, and S. Lall, "A space-time diffusion scheme for peer-to-peer least-squares estimation," in Proceedings of the 5th International Conference on Information Processing in Sensor Networks (IPSN '06), pp. 168-176, April 2006.
[20] D. B. Kingston and R. W. Beard, "Discrete-time average-consensus under switching network topologies," in Proceedings of the American Control Conference, pp. 3551-3556, June 2006.
[21] D. W. Casbeer and R. W. Beard, "Multi-static radar target tracking using information consensus filters," in Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, Chicago, Ill, USA, August 2009.

## Research Article

# Sliding-Mode Control of a Dc/Dc Postfilter for Ripple Reduction and Efficiency Improvement in POL Applications 

S. I. Serna-Garcés, ${ }^{1}$ R. E. Jiménez, ${ }^{2}$ and C. A. Ramos-Paja ${ }^{2}$<br>${ }^{1}$ Department of Electronics and Telecommunications, Instituto Tecnológico Metropolitano (ITM), Calle 54A No. 30-01, Medellín, Colombia<br>${ }^{2}$ Departamento de Energía Eléctrica y Automática, Universidad Nacional de Colombia, Carrera 80 No. 65-223, Medellín, Colombia

Correspondence should be addressed to S. I. Serna-Garcés; sergioserna@itm.edu.co
Received 28 June 2013; Revised 2 September 2013; Accepted 5 September 2013
Academic Editor: Dewei Li
Copyright © 2013 S. I. Serna-Garcés et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper proposes an active postfilter based on two Buck converters, connected in parallel, operating in complementary interleaving. In such a configuration the ripple in the load current could be virtually eliminated to improve the power quality in comparison with classical Point-Of-Load (POL) regulators based on a single Buck converter. The postfilter is designed to isolate the load from the main Buck regulator, leading to the proposed three-converter structure named BuckPS. The correct operation of the postfilter is ensured by means of a sliding-mode controller. Finally, the proposed solution significantly reduces the current harmonics injected into the load, and at the same time, it improves the overall electrical efficiency. Such characteristics are demonstrated by means of analytical results and illustrated using numerical results.


## 1. Introduction

The power supplies designed for computers and communications systems must provide sharp requirements: low voltage, high current, and load voltage ripples. Such conditions are imposed to ensure a high performance of the microprocessors, DSP, ASIC, or memory devices [1, 2]. Since the Buck converter provides output voltages lower than its input voltage [3], it is widely adopted in power architectures designed for this kind of electronic equipment [4], named Point-of-Load (POL) regulators. However, the quality of the current and voltage signals generated by a Buck converter is affected by the load, source, or parameters variation, which changes the ripple magnitudes among other problems $[5,6]$.

Several solutions have been proposed in the literature to improve the quality of the current and voltage generated by a Buck-based POL regulator. In [7], the ripple of a POL converter is reduced by means of a L-C output filter with two stages, which is classically regulated using a controller with two feedback points: the first point sensing the capacitor voltage of the first L-C stage and second point sensing the
capacitor voltage of the second L-C stage, that is, the load voltage. In such a solution, the authors demonstrate that a controller with a single feedback point could be used to stabilize the POL converter, but it requires adjusting one L-C filter to cancel out the zeros of the other L-C filter. In any case, the ripple magnitudes depend on the load impedance, which could change depending on the application conditions.

In [8] the performance of Buck-based POL with different current controllers is analyzed, taking into account the bandwidth of the voltage loop and changes in the input voltage. But such a solution does not analyze the ripple behavior with load variations.

A different approach was presented in [9], where a digital controller is used to reduce the load current ripple in a nonisolated POL converter. Such a solution is based on peak current and average current controllers implemented in an FPGA. This controller scheme requires an A/D converter, which increases the system cost in comparison with analog implementations. In the same way, [10] proposes a selfoscillating digital modulator to change instantaneously the dutycycle of the PWM signal driving the converter switch.


Figure 1: Typical structure of a Buck converter (BuckS).

Thus, it is possible to achieve a sampling frequency of the output voltage, required in the control loop, higher than the switching frequency of the power converter. In this way, a short time response is achieved in the compensation of load variations. But such a solution does not introduce current or voltage ripples analyses.

In [11] a method to design the output filter of a low-voltage/high-current synchronous Buck converter using performance boundary curves is proposed. Such curves constrain the regions in the space of parameters to ensure an acceptable output voltage ripple. But, similar to the previous solutions, the load and source changes that affect the ripple magnitudes are not analyzed.

The previous solutions address the current and voltage ripple limitations by means of passive filters, which are impossible to be modified in operation time. Hence, such solutions are sensible to changes in the load impedance, source voltage, and tolerances of the electronic component parameters. Therefore, this paper proposes a POL based on a synchronous Buck converter operating in cascade with an active postfilter, providing an almost ripple-free current to the load. The postfilter is composed of two parallel-connected Buck converters operating in complementary interleaving [12], and it is regulated using a sliding-mode controller to ensure its correct operation. The proposed POL compensates changes in the load impedance, source voltage, and electronic component parameters. Moreover, the proposed solution improves the electrical efficiency of classical Buck-based POL.

The paper structure is as follows: Section 2 analyzes the Buck converter, introduces the postfilter, and analyzes the proposed POL solution. Then, Section 3 presents the slidingmode current controller designed to reduce the current and voltage ripples injected to the load. Section 4 illustrates the solution benefits in a realistic scenario by means of numerical results. Finally, conclusions close the work.

## 2. POL Regulator Based on a Postfilter

The proposed step-down POL regulator is based on the classical Buck topology, named BuckS, shown in Figure 1. In such a topology the output voltage ripple directly depends on the output capacitance, which is typically implemented using a large capacitor [13]. Using the classical approach given in [3], the output voltage ripple $\Delta V_{o}$ in the Buck converter is given in (1), which depends on the inductor peak current ripple $\Delta I_{L}$. In (1), $T$ represents the switching period, $C$ the output capacitance, $R_{\text {Loss }}$ the aggregated parasitic
resistances of the inductor $L$ and the MOSFETs [14], and $D$ the converter duty cycle, while $I_{o}$ and $V_{o}$ represent the steadystate load current and voltage, respectively. Since the POL converters must provide reduced voltage ripples to the load $[1,2,4], \Delta V_{o}$ can be reduced by increasing $C$ and $L$ or by reducing $T$ (increasing the switching frequency $F_{\text {sw }}$ ) and the parasitic losses $R_{\text {Loss }}$. Instead, this paper proposes to use a $\mathrm{dc} / \mathrm{dc}$ converter to reduce the effective $\Delta I_{L}$ that reaches the capacitor $C$ to achieve the required small $\Delta V_{o}$ condition

$$
\begin{gather*}
\Delta V_{o}=\frac{\Delta I_{L} \cdot T}{8 \cdot C} \\
\Delta I_{L}=\frac{\left(V_{o}+R_{\mathrm{Loss}} \cdot I_{o}\right) D^{\prime} \cdot T}{2 \cdot L} \tag{1}
\end{gather*}
$$

Moreover, from the small ripple approximation and voltsecond and change balances [3], the steady-state induction current $I_{L}$, which is equal to the steady-state load current $I_{o}$, and the voltage conversion ratio are given in (2). In such equations $R$ represents the load impedance at the desired operation condition and $V_{g}$ represents the power source voltage

$$
\begin{gather*}
I_{L}=\frac{V_{o}}{R} \\
\frac{V_{o}}{V_{g}}=\frac{D}{1+R_{\text {Loss }} / R} \tag{2}
\end{gather*}
$$

Finally, the efficiency $\eta$ of the BuckS POL regulator is given in (3). Such efficiency is reduced when the load current increases since the impedance $R$ is reduced. Therefore, the solution proposed in this paper is also intended to improve the overall electrical efficiency. Consider

$$
\begin{equation*}
\eta=\frac{1}{1+R_{\mathrm{Loss}} / R} \tag{3}
\end{equation*}
$$

In the following subsections the proposed postfilter and POL regulator are introduced, contrasting their performance with the classical BuckS solution.
2.1. Postfilter Based on Parallel Buck Converters. Figure 2 presents the proposed postfilter consisting of two Buck converters, operating in complementary interleaving, where the output capacitor is common for both Buck branches. The postfilter main MOSFETs ( $S_{1 U}$ and $S_{1 L}$ ) are complementary activated to generate complementarily inductor current waveforms on $L_{1}$ and $L_{2}$. Such a condition produces the cancelation of the inductor current ripples to provide an almost ripple-free current to the output capacitor $C$, and based on (1), a small voltage ripple is imposed on the load. It must be pointed out that the secondary MOSFETs ( $S_{2 U}$ and $S_{2 L}$ ) are also complementarily activated with respect to the main MOSFET of each branch.

To ensure the cancelation of the inductor current ripples, both Buck branches must operate in continuous conduction mode (CCM); otherwise if a branch current is zero (in discontinuous conduction mode (DCM)), the other branch


Figure 2: Buck converter in interleaved topology.
ripple is propagated to the output capacitor. From (1), and considering the scheme of Figure 2, where $R_{\text {Loss1 }}$ and $R_{\text {Loss } 2}$ represent the parasitic resistances in each branch, the CCM on both branches is achieved when the condition given in (4) is fulfilled. Such an expression takes into account the complementary activation of $S_{1 U}$ and $S_{1 L}$; therefore the first branch has a duty cycle $D$ while the second branch has a complementary duty cycle $D^{\prime}=1-D$. Consider

$$
\begin{equation*}
\frac{V_{o}}{V_{g}}=\frac{D}{1+R_{\mathrm{Loss} 1} / R}=\frac{D^{\prime}}{1+R_{\mathrm{Loss} 2} / R} \tag{4}
\end{equation*}
$$

Solving (4) for $D$, relation (5) is obtained

$$
\begin{equation*}
D=\frac{R+R_{\text {Loss } 1}}{R_{\text {Loss } 1}+2 R+R_{\text {Loss } 2}} \tag{5}
\end{equation*}
$$

Therefore, the symmetrical interleaved Buck converter must be operated at the duty cycle $D$ given in (5) to achieve the desired reduction of the output voltage ripple. Moreover, such a duty cycle imposes the voltage conversion ratio given in

$$
\begin{equation*}
\frac{V_{o}}{V_{g}}=\frac{D R_{\text {Loss } 2}+R_{\mathrm{Loss} 1}(1-D)}{R_{\mathrm{Loss} 2}+\left(R_{\mathrm{Loss} 1} R_{\mathrm{Loss} 2} / R\right)+R_{\mathrm{Loss} 1}} \tag{6}
\end{equation*}
$$

In a practical implementation the inductors $L_{1}$ and $L_{2}$ and the MOSFETs could be selected to have similar values and construction characteristics: $L_{1}=L_{2}=L_{f}$ and $R_{\text {Loss1 }}=R_{\text {Loss2 }}=R_{L f}$. Such a condition is useful to simplify the postfilter design and control since both branches must process the same power. On the basis of such a practical consideration, the required duty cycle given in (5) becomes $D=0.5$, while the voltage conversion ratio provided by the postfilter becomes

$$
\begin{equation*}
\frac{V_{o}}{V_{g}}=\frac{1}{2+\left(R_{L f} / R\right)} \tag{7}
\end{equation*}
$$

Moreover, the steady-state currents in each inductor are equal as given in (8), while the steady-state load current is the sum of such currents, that is, the double of a branch current:

$$
\begin{equation*}
I_{L 1}=I_{L 2}=\frac{V_{g}}{2\left(2 R+R_{L f}\right)} \tag{8}
\end{equation*}
$$



Figure 3: Inductors currents waveform.

In addition, the current ripples in both braches have the same magnitude, while the current ripple injected into the load is near to zero since the postfilter branches operate in complementary mode, which generates opposite slopes for both inductors current as illustrated in Figure 3, where $\Delta I_{L 1}$ and $\Delta I_{L 2}$ represent the ripple magnitudes for each branch. Therefore, since the current ripple reaching the postfilter output capacitor ( $C$ in Figure 2) is near to zero, the output voltage ripple is also near to zero.

To ensure a correct ripple cancelation in the postfilter it is required that both branches exhibit the same average current with opposite instantaneous slopes, and at the same time, it is required that both branches operate in CCM. Such conditions must be ensured despite the load current magnitude or the aging of the components. But, from (5) and (7), it is noted that in all cases the duty cycle and voltage conversion ratio are fixed. Therefore, classical control paradigms based on fixed-frequency drivers, such as the PWM, are not suitable to regulate the postfilter: classical controllers, such as PI, PID, or leadlag, change the duty cycle (using a PWM [3, 15]) to compensate the system perturbations, but because the postfilter requires a fixed duty cycle, it is not possible to regulate it using such type of controllers. For this reason it is necessary to adopt another control paradigm that provides an additional freedom degree. In such a way, this paper proposes to regulate the postfilter using the sliding-mode technique to dynamically change the switching frequency, which according to (1) is given by $f=\left(V_{o}+R_{\text {Loss }} \cdot I_{o}\right) D^{\prime} /(2$. $L \cdot \Delta I_{L}$ ), to ensure that both branches operate with a fixed maximum difference between their currents, which ensures that both branches exhibit the same duty cycle, average current, and ripple magnitude for any system condition.


Figure 4: Synchronized Buck converter with postfilter in interleaved operation (BuckPS).

Finally, since the postfilter was designed using synchronous Buck converters, the CCM operation is granted.
2.2. POL Converter. The main drawback of the postfilter is evident from (5), (6), and (7): the voltage conversion ratio is constant. Therefore, an additional Buck converter is used to regulate the load voltage. Figure 4 presents the proposed POL topology, named BuckPS, obtained from the cascade connection of a Buck converter (interacting with the source) with the postfilter (interacting with the load), where the Buck converter must be independently controlled to regulate the load voltage. Hence, such a Buck converter can be controller using classical approaches based on PWM drivers and PI or PID controllers.

To provide a design criterion, which also ensures a fair comparison with the classical POL based on a Buck converter (BuckS), the inductors of both the postfilter and the Buck converter are considered to be equal; thus $L=L_{1}=L_{2}=L_{f}$ and $R_{L}=R_{L 1}=R_{L 2}=R_{L f}$. Then, the voltage conversion ratio of BuckPS is given in (9), where $D_{\text {PS }}$ represents the duty cycle of the Buck converter. Such an equation puts in evidence that the Buck converter could be independently controlled using a classical PWM-based technique:

$$
\begin{equation*}
\frac{V_{o}}{V_{g}}=\frac{2 D_{\mathrm{PS}} R}{4 R+3 R_{L}} . \tag{9}
\end{equation*}
$$

Moreover, the inductor current of the BuckPS first stage, that is, the Buck converter, and the output current are given in (10). Contrasting such results with the BuckS characteristics given in (2), it is recognized that the BuckPS requires three inductors instead of one, but such devices must support the half of the current imposed on the BuckS inductor:

$$
\begin{align*}
I_{L} & =\frac{D_{\mathrm{PS}} V_{g}}{4 R+3 R_{L}} \\
I_{o} & =\frac{2 D_{\mathrm{PS}} V_{g}}{4 R+3 R_{L}} \tag{10}
\end{align*}
$$

An additional condition of the BuckPS solution is extracted from (9): the voltage conversion ratio is always lower than 0.5 . This condition is illustrated in Figure 5 considering four cases for $R_{L} / R=\{0 \%, 10 \%, 25 \%, 35 \%\}$. In such an example, for $R_{L} / R=10 \%$, a $V_{o} / V_{g}=0.25$ is obtained by operating the BuckPS at $D_{\mathrm{PS}}=0.5375$, while in a BuckS


Figure 5: Conversion ratio of BuckPS and BuckS.
$D_{\mathrm{S}}=0.275$ is required to achieve the same voltage conversion ratio. In general, from (2) and (9) it is concluded that the BuckPS always provides lower output voltage than the BuckS for a given duty cycle. Such a condition is also verified in Figure 5. Hence, lower POL voltages can be achieved with the BuckPS by avoiding duty cycle saturations imposed by turnON and turn-OFF times of the MOSFETs, which limit the minimum operative duty cycle.

Then, the efficiency of the BuckPS is obtained from (9) and (10):

$$
\begin{equation*}
\eta_{\text {BuckPS }}=\frac{4 R}{\left(4 R+3 R_{L}\right)} \tag{11}
\end{equation*}
$$

Such an expression shows an efficiency improvement over BuckS (3). Such a condition is because the BuckPS generates currents in each of the three inductors equal to half of the inductor current in the BuckS. Therefore, since the power losses depend on the square of the current, the losses in the BuckPS are lower. To illustrate such an aspect, the resistance relation $k_{r}$ given in (12) and the efficiency factor $\alpha$ given in (13) have been defined. In particular, $\alpha>1$ implies an improved efficiency of BuckPS over BuckS, while $\alpha<1$ implies a reduced efficiency of BuckPS in comparison with BuckS. Consider

$$
\begin{gather*}
k_{r}=\frac{R_{L}}{R},  \tag{12}\\
\alpha=\frac{\eta_{\text {BuckPS }}}{\eta_{\text {BuckS }}}=\frac{4+4 k_{r}}{4+3 k_{r}} . \tag{13}
\end{gather*}
$$

Since $k_{r}>0$ for real values of $R$ and $R_{L}, \alpha>1$ is always granted in (13), which demonstrates that the proposed BuckPS solution is more efficient than the classical BuckS


Figure 6: Improvement efficiency factor.

TAble 1: Resistance relation and improvement efficiency factor with commercial elements.

| Load (A) | $R_{L}(\mathrm{~m} \Omega)$ | $k_{r}(\%)$ |  | $\alpha-1(\%)$ |  | Cost (\$US) |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 5 | $18+49$ | 10.15 | 30.45 | 2.36 | 6.20 | 3.80 |
| 20 | $8.0+9.5$ | 10.61 | 31.82 | 2.46 | 6.42 | 4.07 |
| 40 | $5.0+2.3$ | 8.85 | 26.55 | 2.07 | 5.53 | 5.68 |
| 60 | $3.5+3.0$ | 11.82 | 35.52 | 2.71 | 7.01 | 6.07 |

implementation. Figure 6 presents the efficiency improvement factor $\alpha-1$, which quantifies the relative efficiency improvement of BuckPS over BuckS, for different values of the resistance relation $k_{r}$. Such numerical results illustrate the improved efficiency of the BuckPS solution.

Table 1 shows values of the efficiency improvement factor considering commercial elements [16]. The calculations were made for load currents equal to $5 \mathrm{~A}, 20 \mathrm{~A}, 40 \mathrm{~A}$, and 60 A , with a ripple current of $10 \%$. Moreover, $R_{L}$ is calculated by adding the inductor resistance and the ON-resistance of the MOSFETs. Then, $k_{r}$ and $\alpha-1$ have two values: the left value corresponds to a load voltage $V_{o}=3.3 \mathrm{~V}$, while the right value corresponds to $V_{o}=1.1 \mathrm{~V}$. In the first case, the efficiency improvement is near to $2.4 \%$, while in the second case the efficiency improvement is between $5 \%$ and $7.5 \%$. Therefore, for modern microprocessors requiring very low operation voltages, the proposed BuckPS could provide a significant improvement in the electrical efficiency.

## 3. Sliding-Mode Current Control

The sliding-mode control technique has been extensively used in the literature to regulate power converters due to its robustness and speed [17]. Moreover, sliding-mode controllers have been also used to regulate active filters to improve power quality in AC environments [18]. In the same


Figure 7: Logic scheme of sliding-mode controller.
way, this paper proposes to design a sliding-mode controller to regulate the postfilter, this with aim of ensuring the correct behavior of the system in any operation condition.

The controller design requires a state-space model of the POL converter. In such a way, the state-space system that describes the BuckPS dynamic behavior, depending on $u_{B}$ (driving signal of the first Buck converter) and $u_{1 U}$ (driving signal of the postfilter), is given in (14). Such a system considers the states vector $x=\left[i_{L} i_{L 1} i_{L 2} v_{C 1} v_{C 2}\right]^{T}$ and follows the nomenclature defined in Figure 4. Consider

$$
\begin{gather*}
\dot{i_{L}}=-\frac{R_{L} i_{L}}{L}-\frac{v_{C 1}}{L}+\frac{V_{g} u_{B}}{L} \\
i_{L 1}=-\frac{R_{L 1} i_{L 1}}{L_{1}}+\frac{v_{C 1} u_{1 U}}{L_{1}}-\frac{v_{C 2}}{L_{1}} \\
i_{L 2}=-\frac{R_{L 2} i_{L 2}}{L_{2}}+\frac{v_{C 1} \overline{u_{1 U}}}{L_{2}}-\frac{v_{C 2}}{L_{2}}  \tag{14}\\
v_{C 1}^{\cdot}=\frac{i_{L}}{C_{1}}-\frac{i_{L 1} u_{1 U}}{C_{1}}-\frac{i_{L 2} \overline{u_{1 U}}}{C_{1}} \\
v_{C 2}=\frac{i_{L 1}}{C_{2}}+\frac{i_{L 2}}{C_{2}}-\frac{v_{C 2}}{\left(R C_{2}\right)}
\end{gather*}
$$

In (14) $\overline{u_{1 U}}=1-u_{1 U}$, where $u_{1 U}=1$ means that MOSFET $S_{1 U}$ is turned ON and MOSFET $S_{1 L}$ is turned OFF, while $u_{1 U}=0$ means that MOSFET $S_{1 U}$ is turned OFF and MOSFET $S_{1 L}$ is turned ON.

Following the same approach proposed in [19], a slidingmode controller was designed to regulate both postfilter inductor currents. The adopted sliding surface, given in (15), is intended to guarantee the same current in both postfilter branches:

$$
\begin{equation*}
S(x)=i_{L 1}-i_{L 2}=0 \tag{15}
\end{equation*}
$$

But to design a practical realization, the surface must be constrained into a hysteretic band $\pm H(t)$, where the MOSFET commutation is determined by (16): when the difference between the indictor currents is smaller than the lower boundary of the hysteretic band $-H(t), u_{1 U}$ must be turned ON (set to 1 ); while if the difference between the inductor currents is larger than the upper boundary of the hysteretic band $+H(t), u_{1 U}$ must be turned OFF (set to 0 ). Therefore, $H(t)$ defines the steady-state value of the currents ripple. Moreover such surface $S(x)=0$ imposes the same average value for both currents, which guarantee the correct
operation of the postfilter. Figure 7 presents the logic scheme for both the sliding surface and the hysteretic comparator:

$$
\begin{array}{ll}
i_{L 1}-i_{L 2}<-H(t), & u_{1 U} \text { set to } 1,  \tag{16}\\
i_{L 1}-i_{L 2}>+H(t), & u_{1 U} \text { set to } 0 .
\end{array}
$$

The necessary and sufficient conditions for surface reachability are given in [20]

$$
\begin{array}{ll}
\lim _{S \rightarrow 0^{-}} \frac{d S(x)}{d t}>0 & u_{1 U}=1, \\
\lim _{S \rightarrow 0^{+}} \frac{d S(x)}{d t}<0 & u_{1 U}=0 . \tag{17}
\end{array}
$$

The time derivative of the sliding surface, given in (18), is obtained from (15). Then, by introducing the relation (18) in (17) and replacing also the second and third rows of (14) in (17), the expressions for surface reachability given in (19) are obtained

$$
\begin{align*}
& \frac{d S(x)}{d t}=\frac{d i_{L 1}}{d t}-\frac{d i_{L 2}}{d t},  \tag{18}\\
\lim _{S \rightarrow 0^{-}} \frac{d S(x)}{d t}= & v_{C 2}\left(\frac{1}{L_{2}}-\frac{1}{L_{1}}\right)+\left(\frac{R_{L 2}}{L_{2}} i_{L 2}-\frac{R_{L 1}}{L_{1}} i_{L 1}\right) \\
& +\frac{v_{C 1}}{L_{1}}>0, \\
\lim _{S \rightarrow 0^{+}} \frac{d S(x)}{d t}= & v_{C 2}\left(\frac{1}{L_{2}}-\frac{1}{L_{1}}\right)+\left(\frac{R_{L 2}}{L_{2}} i_{L 2}-\frac{R_{L 1}}{L_{1}} i_{L 1}\right)  \tag{19}\\
& -\frac{v_{C 1}}{L_{1}}<0 .
\end{align*}
$$

Since for a practical implementation the postfilter inductors are selected equally, $L_{f}=L_{1}=L_{2}$ and $R_{L f}=R_{L 1}=R_{L 2}$, relation (19) is simplified as in (20). In such an expression it is evident that both inequalities are fulfilled, this is because inductors are always positive $\left(L_{f}>0\right)$ and Buck converters provide output voltages with the same polarity of the input voltage ( $v_{C 1}>0$ ). Therefore, the surface reachability of the postfilter controller is always granted

$$
\begin{align*}
\lim _{S \rightarrow 0^{-}} \frac{d S(x)}{d t} & =\frac{v_{C 1}}{L_{f}}>0 \\
\lim _{S \rightarrow 0^{+}} \frac{d S(x)}{d t} & =-\frac{v_{C 1}}{L_{f}}<0 \tag{20}
\end{align*}
$$

The other important aspect in terms of control concerns the local stability, which is verified by using the equivalent control condition given in (21) [19], where $u_{\text {eq }}$ represents an equivalent continuous control input that constrains the system evolution into the sliding surface

$$
\begin{equation*}
\frac{d S(x)}{d t}=0, \quad 0<u_{\mathrm{eq}}<1 \tag{21}
\end{equation*}
$$

From (18) and the second and third rows of (14), in which the control input $u_{1 U}$ has been replaced by the equivalent
continuous variable $u_{\text {eq }}$, the condition given in (21) can be rewritten as in

$$
\begin{equation*}
0<u_{\mathrm{eq}}=\frac{R_{L 1} L_{2} i_{L 1}-R_{L 2} L_{1} i_{L 2}+\left(L_{2}-L_{1}\right) v_{\mathrm{C} 2}+L_{1} v_{\mathrm{C} 1}}{\left(L_{1}+L_{2}\right) v_{\mathrm{C} 1}}<1 . \tag{22}
\end{equation*}
$$

Taking into account that the inductors are selected equally, then (22) becomes

$$
\begin{equation*}
0<R_{L f} L_{f}\left(i_{L 1}-i_{L 2}\right)+L_{f} v_{C 1}<2 L_{f} v_{C 1} \tag{23}
\end{equation*}
$$

Therefore, the difference between the inductor currents must satisfy (24) to guarantee local stability

$$
\begin{equation*}
-\frac{v_{C 1}}{R_{L f}}<i_{L 1}-i_{L 2}<\frac{v_{C 1}}{R_{L f}} \Longrightarrow\left|i_{L 1}-i_{L 2}\right|<\frac{v_{C 1}}{R_{L f}} \tag{24}
\end{equation*}
$$

To ensure that relation (24) is fulfilled in any condition, the maximum magnitude of the inductors current difference must be constrained as in

$$
\begin{equation*}
\max \left|i_{L 1}-i_{L 2}\right|=\Delta_{\max }<\frac{v_{C 1}}{R_{L f}} \tag{25}
\end{equation*}
$$

From the second and third rows of (14) with $u_{1 U}=1$ and $\bar{u}_{1 U}=0$, the ripple magnitudes of both postfilter currents, as defined in Figure 3, are given in (26). It is noted that the maximum difference between the inductor currents is constrained by the sum of such ripple magnitudes as in (27):

$$
\begin{gather*}
\Delta i_{L 1}=\frac{T}{4 L_{f}}\left(-R_{L f} i_{L 1}+v_{C 1}-v_{C 2}\right), \\
\Delta i_{L 2}=-\frac{T}{4 L_{f}}\left(-R_{L f} i_{L 2}-v_{C 2}\right),  \tag{26}\\
\Delta_{\max }=\Delta i_{L 1}+\Delta i_{L 2} \\
\Delta_{\max }=\frac{T}{4 L_{f}}\left(-R_{L f}\left(i_{L 1}-i_{L 2}\right)+v_{C 1}\right) . \tag{27}
\end{gather*}
$$

Since the maximum difference between the inductor currents is max $\left|i_{L 1}-i_{L 2}\right|=\Delta_{\text {max }}$, the second row of (27) must consider $i_{L 1}-i_{L 2}=\Delta_{\max }$. Therefore, the maximum difference between the inductor currents is given by

$$
\begin{equation*}
\Delta_{\max }=\frac{v_{\mathrm{C} 1}}{\left(4 L_{f} / T\right)+R_{L f}} \tag{28}
\end{equation*}
$$

The local stability condition of the sliding-mode controller given in (25) is rewritten as in

$$
\begin{equation*}
\frac{v_{C 1}}{\left(4 L_{f} / T\right)+R_{L f}}<\frac{v_{C 1}}{R_{L f}} . \tag{29}
\end{equation*}
$$

Such an inequality leads to the condition given in (30), which is fulfilled for any operating condition since both the inductance and period are positive quantities. Hence, relation (30) confirms the local stability of the proposed sliding-mode controller:

$$
\begin{equation*}
\frac{4 L_{f}}{T}>0 \tag{30}
\end{equation*}
$$



Figure 8: Practical implementation of the proposed Core 2 Duo POL regulator.

Therefore, since surface reachability is granted by (20) and the local stability is granted by (30), the proposed slidingmode controller always drives the postfilter, from any initial condition, to operate within the space $\left|i_{L 1}-i_{L 2}\right|<H$, which ensures the same average current for both branches and a maximum current difference constrained to $H$. Such characteristics ensure a correct operation of the postfilter.

## 4. Numerical Results

A realistic application was considered to illustrate the operation and advantages of the proposed POL structure by means of numerical results. The example considers a POL regulator designed to supply an Intel Core 2 Duo processor [21, 22], which requires a regulated 1.1 V with $1 \%$ voltage ripple and 60 A . Then, the POL converter was designed to provide a maximum voltage ripple equal to 11 mV with a constant current ripple equal to $10 \%$ of the maximum load current ( 6 A ). Moreover, the switching frequency was selected equal to 100 kHz for the single Buck converters and near to 100 kHz for the postfilter. Therefore, the inductors were calculated to ensure such current ripple and switching frequencies; hence all the inductors were selected equal to $1.5 \mu \mathrm{H}$. Similarly, the capacitors were calculated to fulfill the desired voltage ripple hence all the capacitors we selected equal to $280 \mu \mathrm{~F}$. Moreover, from the last row of Table 1 the parasitic resistance for the single Buck converter and each postfilter branches is extracted, which for all the inductors and MOSFETs are equal to $6.5 \mathrm{~m} \Omega$. Finally, the application considers a 12 V battery as the main power source.

Figure 8 shows the practical implementation of the proposed POL regulator to supply the Core 2 Duo processor. Such a scheme shows the two control systems required: the sliding-mode controller to regulate the postfilter, named SMC, and a PID controller acting on the Buck converter to regulate the load voltage.

Figure 9 shows the postfilter operation in two conditions: start-up and load transient. The former one considers the start-up of the POL converter, where the voltage and currents of all the capacitors and inductors are zero. The postfilter time simulation (top-left) shows the satisfactory current ripple
cancelation, where the output current Io is almost ripple free. It must be pointed out that in such a figure Io is presented divided by 2 to be in the same scale of the postfilter inductor currents. In addition, the figure also presents, in black traces, the maximum limits of the inductors current difference, which is in agreement with the current ripple condition imposed by the application (6 A). From such a behavior it is noted that, in the start-up condition, the slidingmode controller successfully guarantees the correct postfilter operation: both inductor currents have the same average current and the same current ripple, which produces a fixed duty cycle equal to 0.5 to ensure the ripple cancelation.

The postfilter phase plane for the start-up operation is also presented at the bottom-left figure, where it is confirmed that the system is into the sliding surface for any steady-state or transient condition.

The same behavior is achieved for a step-down load transient, in which time simulation is presented in the figure at top-right, where a $10 \%$ load perturbation was introduced. Similar to the start-up case, in this transient condition the postfilter provides an almost ripple-free load current, while the system is always within the sliding surface (depicted at the bottom-right). Therefore, the simulation in Figure 9 confirms the correct operation of the postfilter and the stability of the sliding-mode controller predicted in (20) and (30) for any operation condition.

Another component to design in the proposed POL solution concerns the load voltage regulator, named PID in Figure 8. To design such a controller, the postfilter is modeled to operate in closed loop with the sliding-mode controller, where both inductor currents are equal and the duty cycle of the prefilter is 0.5 . Therefore, the statespace (14) can be simplified as given in (31), where the single control variable is $u_{B}$ :

$$
\begin{gather*}
\dot{i_{L}}=-\frac{R_{L} i_{L}}{L}-\frac{v_{C 1}}{L}+\frac{V_{g} u_{B}}{L} \\
\dot{i}_{L f}=-\frac{R_{L f} i_{L f}}{L_{f}}+\frac{v_{C 1}}{\left(2 L_{f}\right)}-\frac{v_{C 2}}{L_{f}} \\
\dot{v_{C 1}}=\frac{i_{L}}{C_{1}}-\frac{i_{L f}}{C_{1}}  \tag{31}\\
v_{C 2}=\frac{2 i_{L f}}{C_{2}}-\frac{v_{C 2}}{\left(R C_{2}\right)}
\end{gather*}
$$

Then, using the PWM-based averaging technique described in $[3,23]$, the state-space system in (31) was linearized by replacing $u_{B}$ with the duty cycle of the Buck converter. Such a system was used to design the PID controller in agreement with the following criteria: closed loop bandwidth equal to 6 kHz , phase margin higher than $60^{\circ}$, and gain margin higher than 6 dB . The design of the controller was performed in SISOTOOL from MATLAB, obtaining the expression given in

$$
\begin{equation*}
\operatorname{PID}(s)=1200 \frac{\left(1+3.1 \times 10^{-5} s\right)^{2}}{s} \tag{32}
\end{equation*}
$$



Figure 9: Postfilter simulation: start-up and load transient conditions.

To illustrate the improvement of the proposed POL, a BuckS POL was also designed and simulated. Figure 10 compares the BuckS and BuckPS output voltage ripples, obtaining magnitudes of $3.1 \%$ and $0.032 \%$, respectively. Such results put in evidence the large reduction in the voltage ripple provided by the proposed solution, which avoids the requirement of electrolytic capacitances. Similarly, Figure 11 shows the power spectral density (PSD) of the output voltage harmonics for both the BuckS and BuckPS, where a large harmonic component at 100 kHz produced by the BuckS is observed, while the BuckPS exhibits a much attenuated component due to the complementary inductor currents of the postfilter. The simulation also shows that the BuckPS produces a different PSD due to the contribution of both inductor currents, which results in a new harmonic component at 143 kHz . In any case, those results confirm the improvement in the power quality provided to the load by the proposed solution.

To show the overall system performance, Figure 12 shows the dynamic behavior of the BuckPS under a load transient. In such a case, the PID controller must regulate the load voltage
while the sliding-mode controller regulates the postfilter. The simulation considers a load current perturbation equal to $10 \%$ of the steady-state value (from 60 A to 66 A ). The results show the satisfactory compensation of the load voltage provided by the PID controller given in (32). Similarly, Figure 12 also shows the satisfactory regulation of the postfilter inductor currents. Such a correct operation of the sliding-mode controller is also evident from the system evolution reported in the bottom figure, where the system is always constrained with the sliding surface $S(x)$ for any operation condition.

It must be point out that a more complex controller for the Buck converter, such a high-order lead-lag structure, could be used to improve the output voltage dynamics.

## 5. Conclusions

This paper has presented a POL converter based on the cascade connection of an interleaved postfilter with a Buck converter. This solution, named BuckPS, has the aim of


Figure 10: Output voltage ripple comparison between BuckS and BuckPS.


Figure 11: Power spectral density of BuckS and BuckPS.
improving the quality of the power provided to the load, by reducing the output voltage ripple. Moreover, the BuckPS provides an improved efficiency (between the $2.5 \%$ and $7.5 \%$ ) over a classical POL based on a single Buck converter, named BuckS. Similarly, since the BuckPS strongly reduces the output current ripple, its output capacitor could be significantly smaller in comparison with the classical BuckS implementation. This characteristic allows designing the BuckPS without using electrolytic capacitances, which improves the system reliability.


Figure 12: Dynamic behavior of the BuckPS output voltage.

Despite the advantages of the BuckPS structure, it requires more elements and its regulation strategy is more complex in comparison with the BuckS, which could lead to a more costly device. In any case, the elements required by the BuckPS have lower ratings, therefore lower cost, which is especially important for the output capacitor: in BuckS structures a large electrolytic capacitor is required, which increases the system size and cost. Therefore, a comparison between the cost and size of BuckS and BuckPS solutions depends on the specific application conditions.

To illustrate the benefits of the proposed solution, a practical application based on real load requirements was analyzed and simulated. The numerical results of such an example confirm the correctness of the POL converter and the stability of the sliding-mode controller. In the same way, the simulation also puts in evidence the improvement of the proposed BuckPS regulator over a classical BuckS solution.

Finally, this paper describes an analog implementation of the POL controllers. Therefore, a future research may be focused on the digital implementation of the POL control system to provide a more flexible and industrial oriented solution. In such a further work, one of the open problems concerns the fast acquisition of the postfilter currents since the sampling circuit could filter such high-frequency signals. Similarly, the time-delay effect generated by the acquisition and processing circuits could introduce errors in the slidingmode comparator, degrading the controller accuracy and stability.

## Acknowledgments

This work was supported by the Automatic, Electronic and Computer Science Research Group of the Instituto Tecnológico Metropolitano (ITM) and the GAUNAL Research

Group of the Universidad Nacional de Colombia under the Projects P10-102, INAE41, and VECTORIAL-MPPT.

## References

[1] R. V. White, "Emerging on-board power architectures," in Proceedings of the 18th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '03), pp. 799-804, February 2003.
[2] R. V. White, "Using on-board power systems," in Proceedings of the 26th Annual International Telecommunications Energy Conference (INTELEC '04), pp. 234-240, September 2004.
[3] R. W. Erickson and D. Maksimovic, Fundamentals of Power Electronics, Springer, Rocky Mountain National Park, Colo, USA, 2nd edition, 2001.
[4] M. Schlecht, "Intermediate bus architecture: is it for everyone?" IEE Power Engineer, vol. 17, no. 5, pp. 40-41, 2003.
[5] M. Salato, "Datacenter power architecture: IBA versus FPA," in Proceedings of the 33rd International Telecommunications Energy Conference (INTELEC '11), October 2011.
[6] S. Chander, P. Agarwal, and I. Gupta, "Design, modeling and simulation of point of load converter (niPOL)," in Proceedings of the 7th Annual International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON '10), pp. 1206-1210, May 2010.
[7] S. Nagar, F. Al-Zahara Said, M. Orabi, and A. Abou-Alfotouh, "Design of high performance point of load converters with ultra-low output voltage ripple," in Proceedings of the 2nd IEEE Energy Conversion Congress and Exposition (ECCE '10), pp. 4145-4150, September 2010.
[8] Y. Yan, F. C. Lee, and P. Mattavelli, "Dynamic performance comparison of current mode control schemes for Point-of-Load Buck converter application," in Proceedings of the 27th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '12), pp. 2484-2491, February 2012.
[9] K.-Y. Lee, C.-A. Yeh, and Y.-S. Lai, "Design and implementation of fully digital controller for non-isolated-point-of-load converter with high current slew rate," in Proceedings of the 32nd Annual Conference on IEEE Industrial Electronics (IECON '06), pp. 2605-2610, November 2006.
[10] L. T. Jakobsen and M. A. E. Andersen, "Digitally controlled point of load converter with very fast transient response," in Proceedings of the European Conference on Power Electronics and Applications (EPE '07), September 2007.
[11] A. De Nardo, N. Femia, G. Petrone, and G. Spagnuolo, "Optimal buck converter output filter design for point-of-load applications," IEEE Transactions on Industrial Electronics, vol. 57, no. 4, pp. 1330-1341, 2010.
[12] E. Arango, C. A. Ramos-Paja, J. Calvete, R. Giral, and S. Serna, "Asymmetrical interleaved DC/DC switching converters for photovoltaic and fuel cell applications-part 1: circuit generation, analysis and design," Energies, vol. 5, no. 11, pp. 4590-4623, 2012.
[13] R. K. Singh and S. Mishra, "Synthetic-ripple-based digital hysteretic modulator for point-of-load converters," IEEE Transactions on Industrial Electronics, vol. 60, no. 11, pp. 4996-5007, 2013.
[14] C. A. Ramos-Paja, E. Arango, R. Giral, A. J. Saavedra-Montes, and C. Carrejo, "DC/DC pre-regulator for input current ripple reduction and efficiency improvement," Electric Power Systems Research, vol. 81, no. 11, pp. 2048-2055, 2011.
[15] M. M. Peretz and S. Ben-Yaakov, "Time-domain design of digital compensators for PWM DC-DC converters," IEEE Transactions on Power Electronics, vol. 27, no. 1, pp. 284-293, 2012.
[16] Digi-Key Corporation, "Electronic Components DistributorDigiKey Corp.-US Home Page," http://www.digikey.com/ .
[17] E. Bianconi, J. Calvente, R. Giral et al., "A fast currentbased MPPT technique employing sliding mode control," IEEE Transactions on Industrial Electronics, vol. 60, no. 3, pp. 11681178, 2013.
[18] J. Fei and Z. Wang, "Adaptive neural sliding mode control of active power filter," Journal of Applied Mathematics, vol. 2013, Article ID 341831, 8 pages, 2013.
[19] E. Bianconi, J. Calvente, R. Giral et al., "Perturb and observe MPPT algorithm with a current controller based on the sliding mode," Electrical Power and Energy Systems, vol. 44, no. 1, pp. 346-356, 2013.
[20] V. I. Utkin, Sliding Modes in Control and Optimization, Springer, Berlin, Germany, 1992.
[21] Intel Corporation, Intel Core 2 Extreme Processor X6800 and Intel Core 2 Duo Desktop Processor E6000 and E4000 Sequences, 2007.
[22] Intel Corporation, Voltage Regulator-Down (VRD) 11.0 Processor Power Delivery Design Guidelines-for Desktop LGA775 Socket, 2006.
[23] M. A. Shrud, A. Kharaz, A. S. Ashur, M. Shater, and I. Benyoussef, "A study of modeling and simulation for interleaved Buck converter," in Proceedings of the 1st Power Electronic \& Drive Systems \& Technologies Conference (PEDSTC '10), 2010.

# Incomplete Phase Space Reconstruction Method Based on Subspace Adaptive Evolution Approximation 

Tai-fu Li, ${ }^{1}$ Wei Jia, ${ }^{\mathbf{2}}$ Wei Zhou, ${ }^{1}$ Ji-ke Ge, ${ }^{1}$ Yu-cheng Liu, ${ }^{1}$ and Li-zhong Yao ${ }^{1}$<br>${ }^{1}$ School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing 401331, China<br>${ }^{2}$ School of Electronic Engineering, Xian Shiyou University, Xian 710065, China

Correspondence should be addressed to Wei Jia; jiaweifire@163.com and Wei Zhou; zhouw1985@163.com
Received 19 July 2013; Accepted 19 September 2013
Academic Editor: Baocang Ding
Copyright © 2013 Tai-fu Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The chaotic time series can be expanded to the multidimensional space by phase space reconstruction, in order to reconstruct the dynamic characteristics of the original system. It is difficult to obtain complete phase space for chaotic time series, as a result of the inconsistency of phase space reconstruction. This paper presents an idea of subspace approximation. The chaotic time series prediction based on the phase space reconstruction can be considered as the subspace approximation problem in different neighborhood at different time. The common static neural network approximation is suitable for a trained neighborhood, but it cannot ensure its generalization performance in other untrained neighborhood. The subspace approximation of neural network based on the nonlinear extended Kalman filtering (EKF) is a dynamic evolution approximation from one neighborhood to another. Therefore, in view of incomplete phase space, due to the chaos phase space reconstruction, we put forward subspace adaptive evolution approximation method based on nonlinear Kalman filtering. This method is verified by multiple sets of wind speed prediction experiments in Wulong city, and the results demonstrate that it possesses higher chaotic prediction accuracy.


## 1. Introduction

In recent years, industrial disasters and accidents occurred frequently, the meteorological and hydrological conditions were complicated and changeable, and financial markets fluctuated drastically. These phenomena often contain chaotic characteristics [1, 2], and prediction [3] for these phenomena is imminent. For a long time, there was no scientific tool for handling this issue, because the changing mechanisms of characteristics in these phenomena were not understood very well. Hence, aiming at the chaotic characteristics, some scholars worked with structures and made a lot of new researches on the prediction of chaotic time series [4-8].

To study and deal with the measurement data of chaotic system, Kennel et al. presented the reconstruction method of phase space system. Two parameters, the embedding dimension $m$ and delay time $\tau$, needed to be determined before the phase space reconstruction [9, 10]. At present, time delay selection methods that are commonly used in the chaotic short-term prediction mainly include autocorrelation method [11], mutual information method [12], and
singular value fraction method [13]. Calculating methods of embedding dimension mainly include saturated correlation dimension [14], false nearest neighbors method [15], and Cao's method [16]. Hu and Chen put forward the C-C method [17], which can simultaneously estimate the delay time $\tau$ and embedding dimension $m$ with the correlation integral. Autocorrelation method extracts only linear correlation degree between time series, which is hard to be applied to high-dimensional chaos system and nonlinear dynamical system. Mutual information method, which can determine the optimal delay time by calculating the first minimal value of mutual information function, is a nonlinear analysis method, but it cannot avoid massive calculation and cannot satisfy the requirement of complicated space division. It is difficult to determine the threshold of the singular value, because the singular value fraction method is largely affected by noise. When the embedded dimension with the saturated correlation dimension is calculated, the main question is to choose the different neighborhood radius. The radius selection has certain randomness, and the result will be in large deviation with improper choice, because of the influence
of noise in the data and excessive concentration of the data. The determination of threshold has very strong subjectivity when we use false nearest neighbors method to determine the embedding dimension. There is no objective standard to determine the threshold value, especially for the experimental data, which may get a wrong result. Cao's method, an improved false nearest neighbors method, can effectively distinguish random signals and deterministic signals, and embedding dimension can be obtained through a less amount of data. C-C method is based on the statistical theory, so $m$ cannot be precisely determined.

Researches have showed that different phase space reconstruction methods get different $m$ and $\tau$. Moreover, the same chaotic time series with the same kind of method in different times may get different $m$ and $\tau$. There is no phase space reconstruction that can obtain complete and independent phase space. After phase space reconstruction, prediction model is often established through the functional approximation method.

The prediction model based on phase space reconstruction has been used to adopt the functional approximation method based on the neural network [18-21], which has strong nonlinear fitting capability and can approximate any complex nonlinear relationships. However, since neural network is only suitable for approximation of a deterministic system, it is difficult to guarantee the time-varying system performance and ensure its generalization performance in other untrained neighborhood. Meanwhile, the prediction effect of neural network is not good, because the chaotic time series is a complex nonlinear uncertain system.

In this study, we introduce Kalman filtering to neural network model [22], inspired by Kalman iteration and Bucy and Sunahara's nonlinear extended Kalman filtering theory [23]. The subspace approximation of neural network based on the nonlinear extended Kalman filtering (EKF) has a function which is dynamic evolution approximation from one neighborhood to another. Therefore, we can constitute a phase space by choosing a kind of phase space reconstruction method, and the space may be incomplete, not separate, and can be seen as a subspace of the ideal phase space. On this basis, we put forward adaptive neural network model based on nonlinear Kalman filtering and finally realize the subspace approximation of dynamic evolution system. In addition, we simulate wind speed series in Wulong city using the proposed method. By comparing with BP neural network prediction model, the results show that our method possesses higher prediction accuracy.

The paper is organized as follows. Section 2 discusses about the subspace approximation of phase space reconstruction. In Section 3, we describe the neural network model based on nonlinear Kalman filtering. Section 4 uses practical examples and series tests to verify the proposed method, while Section 5 contains the conclusions of the present work.

## 2. Subspace Approximation of Phase Space Reconstruction

Reconstructing phase space by chaos theory needs to identify the chaos of time series. Single variable time series can
be reconstructed into a phase space by Takens' embedding theorem in phase space reconstruction [24, 25]; that is, the original dynamical system can be restored in the sense of topological equivalence as long as the embedding dimension is sufficiently high. For the observed time series $x(1), x(2), \ldots, x(t)$, after time delay reconstruction by Takens embedding theorem, it will receive a set of space vector

$$
\begin{array}{r}
\mathbf{X}(t)=\{x(t), x(t+\tau), \ldots, x(t+(m-1) \tau)\} \\
t=1,2, \ldots M, M=N-(m-1) \tau . \tag{1}
\end{array}
$$

After phase space reconstruction, the data space is

$$
\left[\begin{array}{cccc}
x(1) & x(2) & \cdots & x(t)  \tag{2}\\
x(1+\tau) & x(2+\tau) & \cdots & x(t+\tau) \\
\vdots & \vdots & \ddots & \vdots \\
x(1+(m-1) \tau) & x(2+(m-1) \tau) & \cdots & x(t+(m-1) \tau)
\end{array}\right] .
$$

Accordingly, we acquire

$$
\begin{equation*}
f: \mathbf{R}^{m} \longrightarrow \mathbf{R} \tag{3}
\end{equation*}
$$

where $f$ is a single-valued function. Then, we have

$$
\begin{equation*}
x(t+m \tau)=f(x(t), x(t+\tau), \ldots, x(t+(m-1) \tau)) \tag{4}
\end{equation*}
$$

However, it cannot be really obtained as the data are often limited. Hence, $\widehat{f}: \mathbf{R}^{m} \rightarrow \mathbf{R}$ can only be constituted by limited measurement data, making $\widehat{f}$ sufficiently approximate to $f$, consequently we can get a nonlinear prediction model.

This paper employs the neural network to predict chaotic series. However, the neural network cannot readily handle the inconsistency of the phase space reconstruction because of uncertain nonlinear chaotic time series. Therefore, it is crucial to adaptively construct subspace to approximate chaotic series through the incomplete phase space. The feature of adaptive subspace approximation is that it can add new data in real time and forget old data in the process of training. Consequently, weights and thresholds of the neural network are continuously modified to realize the dynamic evolution modeling.

## 3. Neural Network Model Based on Nonlinear Kalman Filtering

Kalman filtering has good adaptability. It can dynamically update and forecast the system information in real time with limited data. However, it cannot be readily used for complicated nonlinear model. Meanwhile, the extended Kalman filtering (EKF) is a kind of effective method to handle nonlinear filtering.

The mathematical model of EKF is as follows:

$$
\begin{gather*}
\mathbf{X}_{k+1}=f\left(\mathbf{X}_{k}, k\right)+\Gamma\left(\mathbf{X}_{k}, k\right) \mathbf{W}_{k} \\
\mathbf{Z}_{k}=h\left(\mathbf{X}_{k}, k\right)+\mathbf{V}_{k}, \tag{5}
\end{gather*}
$$

where $\mathbf{W}_{k}$ and $\mathbf{V}_{k}$ are independent, zero mean, and Gaussian random processes with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively. The statistical properties are as follows:

$$
\begin{equation*}
p(w) \sim N(0, \mathbf{Q}), \quad p(v) \sim N(0, \mathbf{R}) . \tag{6}
\end{equation*}
$$

EKF spreads nonlinear functions $f(\cdot)$ and $h(\cdot)$ to Taylor series around filtering value $\widehat{\mathbf{X}}_{k}$ and predicted value $\widehat{\mathbf{X}}_{k}^{-}$, respectively, only retaining the first-order information. Hence, the linearization model of the nonlinear system is obtained, and then we can obtain the EKF formula in nonlinear system by basic equations of Kalman filtering.

Given a forward network with $N$ layers, the numbers of neurons in each layer are $S_{k}(k=1,2, \ldots, N)$. Suppose that input layer is the first layer and output layer is the $N$ th layer. The weights of the $k$ th layer neurons are $W_{i j}^{k}(i=$ $\left.1,2, \ldots, S_{k-1} ; j=1,2, \ldots, S_{k}\right)$. In order to convert the calculation of connection weights $W_{i j}^{k}$ in the above problem into filter recursive estimation form, we let all of the network weights constitute the state vector

$$
\begin{equation*}
\mathbf{W}=\left[W_{11}^{1} \cdots W_{S_{1} S_{2}}^{1} W_{11}^{2} \cdots W_{S_{2} S_{3}}^{2} \cdots W_{11}^{N-1} W_{S_{N-1} S_{N}}^{N-1}\right]^{T} \tag{7}
\end{equation*}
$$

where state vector $\mathbf{W}$ consists of all of the weights according to the linear array, and its dimension is as follows:

$$
\begin{equation*}
N_{W}=\sum_{i=1}^{N-1} S_{i} S_{i+1} \tag{8}
\end{equation*}
$$

Then the state equation and measurement equation of the system can be expressed as

$$
\begin{gather*}
\mathbf{W}_{k}=\mathbf{W}_{k-1},  \tag{9}\\
\mathbf{Y}_{e k}=h\left(\mathbf{W}_{k}, \mathbf{X}_{k}\right)+\mathbf{V}_{k}=\mathbf{Y}_{r k}+\mathbf{V}_{k}, \tag{10}
\end{gather*}
$$

where $\mathbf{Y}_{e k}$ is the expected output, $\mathbf{X}_{k}$ is the input vector, and $\mathbf{Y}_{r k}$ is the actual output.

The measurement noise $\mathbf{V}_{k}$ is assumed to be additive, white, and Gaussian, with zero mean and with covariance matrix defined by

$$
\begin{equation*}
E\left(\mathbf{V}_{k}\right)=0, \quad E\left(\mathbf{V}_{k} \mathbf{V}_{k}^{T}\right)=\mathbf{R}_{k} \tag{11}
\end{equation*}
$$

Suppose that the output of the $j$ th node for the $l$ th layer in the $k$ th iteration is

$$
\begin{equation*}
O_{j k}^{l}=F_{j}^{l}\left(W_{j k}^{l}, O_{k}^{l-1}\right) \tag{12}
\end{equation*}
$$

From (10) and (12), we have

$$
\begin{aligned}
\mathbf{Y}_{e k} & =h\left(\mathbf{W}_{k}, \mathbf{X}_{k}\right)+\mathbf{V}_{k} \\
& =F^{N}\left(W_{k}^{N}, F^{N-1}\left(W_{k}^{N-1} \cdots F^{2}\left(W_{k}^{2}, \mathbf{X}_{k}\right)\right)\right)+\mathbf{V}_{k}, \\
\mathbf{Y}_{e k} & =h\left(\widehat{\mathbf{W}}_{k}^{-}, \mathbf{X}_{k}\right)+\left.\frac{\partial h}{\partial W}\right|_{\mathbf{W}_{k}=\widehat{\mathbf{W}}_{k}^{-}}\left(\mathbf{W}_{k}-\widehat{\mathbf{W}}_{k}^{-}\right)+\mathbf{V}_{k} .
\end{aligned}
$$

Table 1: Extended Kalman filtering neural network algorithm.

| Extended Kalman filtering neural network |  |
| :--- | :---: |
| (1) Initialization | $\widehat{\boldsymbol{\theta}}_{0}=E\left(\boldsymbol{\theta}_{0}\right)=\left[\widehat{\mathbf{W}}_{0}, \widehat{\mathbf{b}}_{0}\right]^{T}$ |
|  | $\mathbf{P}_{0}=E\left[\left(\boldsymbol{\theta}_{0}-\widehat{\boldsymbol{\theta}}_{0}\right)\left(\boldsymbol{\theta}_{0}-\widehat{\boldsymbol{\theta}}_{0}\right)^{T}\right]$ |
| (2) Time update (forecast) | $\widehat{\boldsymbol{\theta}}_{k}^{-}=\widehat{\boldsymbol{\theta}}_{k-1}$ |
|  | $\mathbf{P}_{k}^{-}=\mathbf{P}_{k-1}+\mathbf{Q}_{k-1}$ |
| (3) Measurement update | $\mathbf{K}_{k}=\mathbf{P}_{k}^{-} \mathbf{H}_{k}^{T}\left(\mathbf{H}_{k} \mathbf{P}_{k}^{-} \mathbf{H}_{k}^{T}+\mathbf{R}_{k}\right)^{-1}$ |
| (correct) | $\widehat{\boldsymbol{\theta}}_{k}=\widehat{\boldsymbol{\theta}}_{k}^{-}+\mathbf{K}_{k}\left[\mathbf{Y}_{e k}-h\left(\widehat{\boldsymbol{\theta}}_{k}^{-}, \mathbf{X}_{k}\right)\right]$ |
|  | $\mathbf{P}_{k}=\left(\mathbf{I}-\mathbf{K}_{k} \mathbf{H}_{k}\right) \mathbf{P}_{k}^{-}$ |

$\mathbf{Q}_{k-1}$ and $\mathbf{R}_{k}$ are process noise covariance and measurement noise covariance, respectively, $\mathbf{H}_{k}$ is the Jacobian matrix of observable model, $\hat{\boldsymbol{\theta}}_{k}^{-}$is the optimal predictive value for step $k$ according to step $k-1$, and $\hat{\boldsymbol{\theta}}_{k}$ is the optimal filter estimate for step $k$.

Assume that

$$
\begin{equation*}
\left.\frac{\partial h}{\partial W}\right|_{\mathbf{W}_{k}=\widehat{\mathbf{W}}_{k}^{-}}=\mathbf{H}_{k}, \quad h\left(\widehat{\mathbf{W}}_{k}^{-}, \mathbf{X}_{k}\right)-\left.\frac{\partial h}{\partial W}\right|_{\mathbf{W}_{k}=\widehat{\mathbf{W}}_{k}^{-}} \widehat{\mathbf{W}}_{k}^{-}=\mathbf{C}_{k} . \tag{14}
\end{equation*}
$$

Accordingly, the measurement equation may also be expressed as

$$
\begin{equation*}
\mathbf{Y}_{e k}=\mathbf{H}_{k} \mathbf{W}_{k}+\mathbf{C}_{k}+\mathbf{V}_{k} \tag{15}
\end{equation*}
$$

The Jacobian matrix of the function $h(\cdot)$ is described by

$$
\mathbf{H}_{k}=\left[\begin{array}{cccc}
\frac{\partial h_{1}}{\partial w_{1}} & \frac{\partial h_{1}}{\partial w_{2}} & \cdots & \frac{\partial h_{1}}{\partial w_{n}}  \tag{16}\\
\frac{\partial h_{2}}{\partial w_{1}} & \frac{\partial h_{2}}{\partial w_{2}} & \cdots & \frac{\partial h_{2}}{\partial w_{n}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial h_{n}}{\partial w_{1}} & \frac{\partial h_{n}}{\partial w_{2}} & \cdots & \frac{\partial h_{n}}{\partial w_{n}}
\end{array}\right] .
$$

Similarly, all thresholds of the network constitute the state vector

$$
\begin{equation*}
\mathbf{b}=\left[b_{1}^{1} \cdots b_{S_{2}}^{1} b_{1}^{2} \cdots b_{S_{3}}^{2} \cdots b_{1}^{N-1} \cdots b_{S_{N}}^{N-1}\right]^{T} \tag{17}
\end{equation*}
$$

where the dimension is

$$
\begin{equation*}
N_{b}=\sum_{i=1}^{N-1} S_{i+1} \tag{18}
\end{equation*}
$$

Suppose that $\mathbf{W}$ and $\mathbf{b}$ are both state variable; that is, the state vector composed of weights and thresholds is described by

$$
\begin{align*}
\boldsymbol{\theta}= & {[\mathbf{W}, \mathbf{b}]^{T} } \\
= & {\left[W_{11}^{1} \cdots W_{S_{1} S_{2}}^{1} b_{1}^{1} \cdots b_{S_{2}}^{1} W_{11}^{2} \cdots W_{S_{2} S_{3}}^{2} b_{1}^{2}\right.}  \tag{19}\\
& \left.\cdots b_{S_{3}}^{2} \cdots W_{11}^{N-1} W_{S_{N-1} S_{N}}^{N-1} b_{1}^{N-1} \cdots b_{S_{N}}^{N-1}\right]^{T} .
\end{align*}
$$

Kalman filtering algorithm on training weights and thresholds of the neural network is as in Table 1.

TABLE 2: Comparison among phase space reconstruction methods.

| Parameter |  | Method |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Autocorrelation | Mutual information | False nearest neighbors | Cao | C-C |
| $\tau$ | 1 | 12 | - | - | 5 |
| $m$ | - | - | 4 | 3 | 5 |

"-" means nothing.
Table 3: Parameters of the same phase space reconstruction during different time periods.

|  |  | Interval |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Parameter | $T_{1}$ | $T_{2}$ | $T_{4}$ | $T_{5}$ |  |
|  | $(k=1,2, \ldots, 600)$ | $(k=601,602, \ldots, 1200)$ | $(k=1201, \ldots, 1800)$ | $(k=1801, \ldots, 2400)$ | $(k=2401, \ldots, 3000)$ |
| $\tau$ | 5 | 3 | 4 | 4 | 5 |
| $m$ | 5 | 5 | 4 | 9 | 10 |

TABLE 4: Comparison among phase space reconstruction methods.

| Parameter | Autocorrelation | Mutual information | Method | False nearest neighbors | Cao |
| :--- | :---: | :---: | :---: | :---: | :---: |

Table 5: Various combinations on two forecasting methods.

| Model | Combination | Parameter | Forecasting |
| :--- | :--- | :---: | :---: |
| a1 | Autocorrelation + false nearest neighbors | $\tau=1, m=3$ | BPNN |
| b1 | Autocorrelation + false nearest neighbors | $\tau=1, m=3$ | EKFNN |
| a2 | Mutual information + false nearest neighbors | $\tau=12, m=3$ | BPNN |
| b2 | Mutual information + false nearest neighbors | $\tau=12, m=3$ | EKFNN |
| a3 | Autocorrelation + Cao | $\tau=1, m=7$ | BPNN |
| b3 | Autocorrelation + Cao | $\tau=1, m=7$ | EKFNN |
| a4 | Mutual information + Cao | $\tau=12, m=7$ | BPNN |
| b4 | Mutual information + Cao | $\tau=12, m=7$ | EKFNN |
| a5 | C-C | $\tau=3, m=4$ | BPNN |
| b5 | C-C | $\tau=3, m=4$ | EKFNN |

## 4. Simulation Examples

4.1. Determining of Embedding Dimension and Delay Time. One of the most popular chaos logistic mapper is selected as the study object. Logistic equation is

$$
\begin{equation*}
x_{n+1}=\alpha x_{n}\left(1-x_{n}\right), \quad \alpha \in[0,4] . \tag{20}
\end{equation*}
$$

The related time series are produced according to (20). It is a chaotic system when $\alpha=4$. Assume that initial value of series is 0.1 , and 4000 points are calculated. The first 1000 points are eliminated as transition phenomenon, leaving the remaining 3000 points to reconstruct phase space. Before the phase space reconstruction, we determine the embedding dimension $m$ and delay time $\tau$. A comparison among several methods is present in Table 2.

Obviously, the optimal embedding dimension and delay time are generally different by different methods of phase space reconstruction.

In order to verify the fact that data at different time will obtain different embedding dimension $m$ and delay time $\tau$ with the same phase space reconstruction method, we have the following experiment.

The remaining 3000 points ( $k=1,2, \ldots, 3000$ ) are divided into five parts, with time intervals $T_{1}, T_{2}, T_{3}, T_{4}$, and $T_{5}$, respectively. Embedding dimension and delay time are present in Table 3 by C-C method.

Apparently, the data during different time periods will acquire different embedding dimension and delay time by using the same phase space reconstruction method.
4.2. Wind Speed Chaotic Series Forecasting Simulation. Analysis about the chaotic characteristics of wind speed in the process of wind power generation has been presented in a related article [26]. We record one of the wind speed data every 10 minutes, and 150 groups of wind speed data in Wulong city are used to simulate experiments in our study.


Figure 1: The effect comparison between al and b1.

(a) The predicted wind speed data


$$
\begin{aligned}
& -0 \text { a2 } \\
& \ldots+\cdots \mathrm{b} 2
\end{aligned}
$$

(b) Relative wind speed error

Figure 2: The effect comparison between a 2 and b 2 .


Figure 3: The effect comparison between a3 and b3.


Figure 4: The effect comparison between a 4 and b 4 .


Figure 5: The effect comparison between a5 and b5.

We obtain the corresponding $m$ and $\tau$ by different phase space reconstruction methods, as shown in Table 4.

Various combinations are present in Table 5.
Wind speed prediction $[27,28]$ of chaotic time series about neural network model usually extracts phase space reference points as the BP neural network training samples on the basis of phase space reconstruction. We establish the neural network model based on nonlinear Kalman filtering, including two parts: predict wind speed and constantly modify weights and thresholds of the neural network by Kalman recursion. In this paper, BPNN model in the same structure is employed to forecast wind speed time series, in order to illustrate the validity of EKFNN on predicting the chaotic time series. The same 150 groups of wind speed data are used to simulate experiments. The predicted curves and error curves are shown in Figures 1, 2, 3, 4, and 5.

Comparisons among several models in four indices are present in Table 6.

We list 12 groups, a total of 2 hours of wind speed forecasting results in two methods, under the same phase

Table 6: Different wind speed model index.

| Model |  | Error |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | MAE | MRE | MSE | SSE |
| a1 | 0.8482 | 0.1076 | 1.3786 | 206.7967 |
| b1 | 0.3365 | 0.0416 | 0.1919 | 28.7834 |
| a2 | 1.2279 | 0.1674 | 2.9614 | 444.2152 |
| b2 | 0.6190 | 0.0759 | 0.6209 | 93.1357 |
| a3 | 0.5106 | 0.0656 | 0.4089 | 61.3361 |
| b3 | 0.3783 | 0.0458 | 0.2654 | 39.8130 |
| a4 | 2.4005 | 0.3079 | 8.4766 | $1.2715 e+003$ |
| b4 | 0.0852 | 0.0110 | 0.0177 | 2.6584 |
| a5 | 0.8952 | 0.1173 | 1.7970 | 269.5495 |
| b5 | 0.2867 | 0.0357 | 0.1359 | 20.3850 |

MAE, MRE, MSE, and SSE are Mean Absolute Error, Mean Relative Error, Mean Square Error, and Sum of Squared Error, respectively.
space reconstruction. Compare the prediction performance in the next $10 \mathrm{~min}, 20 \mathrm{~min}, 30 \mathrm{~min}$, and up to, 120 min .

Table 7: Observed wind speed data and predicted data.

| Time (min) | Observed value ( $\mathrm{m} / \mathrm{s}$ ) | Predicted value ( $\mathrm{m} / \mathrm{s}$ ) | Relative error <br> (\%) | $\begin{aligned} & \text { Predicted value } \\ & (\mathrm{m} / \mathrm{s}) \end{aligned}$ | Relative error <br> (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | al |  | b1 |  |
| 10 | 10.4900 | 10.0801 | 3.9561 | 10.5101 | 0.1823 |
| 20 | 10.3700 | 9.9783 | 3.7762 | 10.4011 | 0.2630 |
| 30 | 10.6900 | 10.1024 | 5.4902 | 10.4621 | 2.1476 |
| 40 | 10.2100 | 10.8802 | 6.5374 | 10.6202 | 4.0235 |
| 50 | 10.3000 | 9.1424 | 11.2402 | 10.1531 | 1.4947 |
| 60 | 9.8000 | 10.3508 | 5.6401 | 10.2210 | 4.2446 |
| 70 | 9.3000 | 8.6881 | 6.5810 | 9.6032 | 3.2585 |
| 80 | 9.4900 | 8.1883 | 13.7201 | 9.3163 | 1.8295 |
| 90 | 10.1100 | 9.6013 | 5.0321 | 9.7631 | 3.4282 |
| 100 | 9.2100 | 10.4726 | 13.7200 | 10.0300 | 8.9132 |
| 110 | 8.5500 | 7.2082 | 15.6903 | 8.8621 | 3.6533 |
| 120 | 9.0000 | 7.0824 | 21.3101 | 8.5470 | 5.0347 |
|  |  | a2 |  | b2 |  |
| 10 | 10.4900 | 9.6930 | 7.5977 | 10.4785 | 0.1095 |
| 20 | 10.3700 | 10.4006 | 0.2955 | 10.4249 | 0.5292 |
| 30 | 10.6900 | 10.0489 | 5.9968 | 10.4992 | 1.7845 |
| 40 | 10.2100 | 10.5699 | 3.5250 | 10.3357 | 1.2311 |
| 50 | 10.3000 | 10.2559 | 0.4285 | 10.2518 | 0.4675 |
| 60 | 9.8000 | 10.1445 | 3.5151 | 10.1132 | 3.1960 |
| 70 | 9.3000 | 9.2583 | 0.4488 | 10.0703 | 8.2826 |
| 80 | 9.4900 | 9.8944 | 4.2609 | 9.9547 | 4.8963 |
| 90 | 10.1100 | 10.6937 | 5.7739 | 9.7503 | 3.5580 |
| 100 | 9.2100 | 11.4420 | 24.2351 | 9.5212 | 3.3787 |
| 110 | 8.5500 | 11.8315 | 38.3804 | 9.1823 | 7.3954 |
| 120 | 9.0000 | 11.2543 | 25.0475 | 9.3012 | 3.3463 |
|  |  | a3 |  | b3 |  |
| 10 | 10.4900 | 10.9775 | 4.6468 | 10.7194 | 2.1870 |
| 20 | 10.3700 | 10.4931 | 1.1872 | 10.4277 | 0.5567 |
| 30 | 10.6900 | 10.3878 | 2.8269 | 10.4168 | 2.5559 |
| 40 | 10.2100 | 13.7393 | 34.5672 | 10.1893 | 0.2032 |
| 50 | 10.3000 | 9.8085 | 4.7716 | 10.2918 | 0.0793 |
| 60 | 9.8000 | 12.1826 | 24.3126 | 10.0993 | 3.0540 |
| 70 | 9.3000 | 9.7924 | 5.2943 | 9.8370 | 5.7739 |
| 80 | 9.4900 | 9.3521 | 1.4529 | 9.6689 | 1.8855 |
| 90 | 10.1100 | 14.5803 | 44.2168 | 9.7772 | 3.2920 |
| 100 | 9.2100 | 14.6095 | 58.6264 | 9.5805 | 4.0233 |
| 110 | 8.5500 | 8.6425 | 0.0821 | 9.3442 | 9.2884 |
| 120 | 9.0000 | 9.7379 | 8.1988 | 9.3314 | 3.6817 |
|  |  | a4 |  | b4 |  |
| 10 | 10.4900 | 8.9295 | 14.8759 | 10.1137 | 3.5868 |
| 20 | 10.3700 | 9.2037 | 11.2466 | 9.7344 | 6.1296 |
| 30 | 10.6900 | 8.6494 | 19.0887 | 10.3308 | 3.3597 |
| 40 | 10.2100 | 8.5516 | 16.2426 | 9.9803 | 2.2501 |
| 50 | 10.3000 | 9.3038 | 9.6717 | 10.5685 | 2.6064 |
| 60 | 9.8000 | 8.7467 | 10.7479 | 10.1183 | 3.2477 |
| 70 | 9.3000 | 8.7126 | 6.3161 | 9.7779 | 5.1391 |

Table 7: Continued.
$\left.\left.\begin{array}{lcccc}\hline \text { Time (min) } & \begin{array}{c}\text { Observed value } \\ (\mathrm{m} / \mathrm{s})\end{array} & \begin{array}{c}\text { Predicted value } \\ (\mathrm{m} / \mathrm{s})\end{array} & \begin{array}{c}\text { Relative error } \\ (\%)\end{array} & \begin{array}{c}\text { Predicted value } \\ (\mathrm{m} / \mathrm{s})\end{array} \\ \hline 80 & 9.4900 & 8.6188 & 9.1802 & 9.8413 \\ 90 & 10.1100 & 9.4713 & 6.3176 & 10.9886 \\ 100 & 9.2100 & 9.3204 & 1.1992 & 8.7021 \\ 110 & 8.5500 & 9.5029 & 11.1452 & 8.8816\end{array}\right] \begin{array}{c}\text { Relative error }\end{array}\right)$

Comparisons among several prediction results in two methods are present in Table 7.

Figures 1-5 show that relative error of wind speed prediction by EKF neural network is much smaller than that by BP neural network, through observing the future wind speed prediction of 150 groups. As can be seen in Table 6, the prediction effects are largely different by different kinds of phase space reconstruction methods. Four performance indices, which are Mean Absolute Error (MAE), Mean Relative Error (MRE), Mean Square Error (MSE), and Sum of Squared Error (SSE), of EKF neural network, are also far less than those of corresponding general neural network.

Apparently, EKF neural network can solve the inconsistency problem of phase space reconstruction and approximate chaotic time series well through subspace. The neural network model based on EKF has outstanding adaptability, so it can predict the wind speed chaotic time series with higher precision, compared with BP neural network.

Furthermore, we can conclude that in Table 7, prediction accuracy of EKF neural network is higher than that of BP neural network, by comparing the prediction performance of wind speed in the next $10 \mathrm{~min}, 20 \mathrm{~min}, 30 \mathrm{~min}$, and up to, 120 min . It demonstrates that EKF neural network model, which has better dynamic adaptability, can better the prediction of wind speed time series with nonlinear chaotic characteristics. Therefore, the proposed phase space reconstruction method of the adaptive evolution approximation in this paper is an effective approach.

## 5. Conclusion and Further Work

The phase space reconstruction cannot meet characteristics of the completeness and independence, and the results with different reconstruction methods are obviously inconsistent. The reconstructed phase space is a subspace of the ideal space. If a subspace approximation can make the real-time dynamic evolution, then the initial constructed phase space, for which the evolution is adaptive subspace approximation, can finally approximate to the ideal phase space much better.

In this paper, neural network model based on nonlinear Kalman filter is established, by dynamic adaptivity of nonlinear Kalman filter. The model will add new samples in real time and gradually eliminate previous data, as a moving samples window, and the evolution of the training sample continually updates weights and thresholds of the neural network. As a result, adaptive subspace approximation is implemented by reconstructed incomplete phase space.

The optimized plan, which combines the nonlinear Kalman filter with neural network, sufficiently utilizes the nonlinear approximation capability of neural network and dynamic adaptive ability of real-time update correction of nonlinear Kalman filter. Consequently, it can realize subspace adaptive evolution approximation and solve the inconsistency problem of phase space reconstruction. Therefore, it is a nice direction in research into chaotic prediction. Future research can be performed in a number of areas. It provides a good technical support in studying problems of meteorology, hydrology, and finance fields.

## Acknowledgments

This work was supported by the National Science Foundation of China (no. 51075418), the National Science Foundation of China (no. 61174015), Chongqing CMEC Foundations of China (no. CSTC 2013jjB40007), and Chongqing Scientific Personnel Training Plan of China (no. CSTC 2013kjrcqnrc40008).

## References

[1] G. Kaddoum, F. Gagnon, P. Chargé, and D. Roviras, "A generalized BER prediction method for differential chaos shift keying system through different communication channels," Wireless Personal Communications, vol. 64, no. 2, pp. 425-437, 2012.
[2] B. Sivakumar, "Chaos theory in geophysics: past, present and future," Chaos, Solitons \& Fractals, vol. 19, no. 2, pp. 441-462, 2004.
[3] P. A. Mastorocostas, J. B. Theocharis, and A. G. Bakirtzis, "Fuzzy modeling for short term load forecasting using the orthogonal least squares method," IEEE Transactions on Power Systems, vol. 14, no. 1, pp. 29-36, 1999.
[4] H. Y. Yang, H. Ye, G. Wang, J. Khan, and T. Hu, "Fuzzy neural very-short-term load forecasting based on chaotic dynamics reconstruction," Chaos, Solitons \& Fractals, vol. 29, no. 2, pp. 462-469, 2006.
[5] B. Sivakumar, "A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers," Journal of Hydrology, vol. 258, no. 1-4, pp. 149-162, 2002.
[6] A. Porporato and L. Ridolfi, "Nonlinear analysis of river flow time sequences," Water Resources Research, vol. 33, no. 6, pp. 1353-1367, 1997.
[7] P. Zhao, L. Xing, and J. Yu, "Chaotic time series prediction: from one to another," Physics Letters A, vol. 373, no. 25, pp. 2174-2177, 2009.
[8] H. D. I. Abarbanel, Analysis of Observed Chaotic Data, Springer, New York, NY, USA, 1996.
[9] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," Physical Review A, vol. 45, no. 6, pp. 3403-3411, 1992.
[10] H. Ma and C. Han, "Selection of embedding dimension and delay time in phase space reconstruction," Frontiers of Electrical and Electronic Engineering in China, vol. 1, no. 1, pp. 111-114, 2006.
[11] B. Chen, G. Liu, J. Tang et al., "Research on chaotic sequence autocorrelation by phase space method," Journal of the University of Electronic Science and Technology of China, vol. 39, no. 6, pp. 859-863, 2010.
[12] A. Jiang, X. Huang, Z. Zhang, J. Li, Z.-Y. Zhang, and H.-X. Hua, "Mutual information algorithms," Mechanical Systems and Signal Processing, vol. 24, no. 8, pp. 2947-2960, 2010.
[13] N. Abu-Shikhah and F. Elkarmi, "Medium-term electric load forecasting using singular value decomposition," Energy, vol. 36, no. 7, pp. 4259-4271, 2011.
[14] C. Gao and X. Liu, "Chaotic identification of BF ironmaking process I: the calculation of saturated correlative dimension," Acta Metallurgica Sinica, vol. 40, no. 4, pp. 347-350, 2004.
[15] I. M. Carrión and E. A. Antúnez, "Thread-based implementations of the false nearest neighbors method," Parallel Computing, vol. 35, no. 10-11, pp. 523-534, 2009.
[16] J. Hite Jr., Learning in Chaos, Gulf Professional Publishing, Amsterdam, The Netherlands, 1999.
[17] Y. Hu and T. Chen, "Phase-space reconstruction technology of chaotic attractor based on C-C method," Journal of Electronic Measurement and Instrument, vol. 35, no. 10-11, pp. 425-430, 2012.
[18] D. A. Fadare, "The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria," Applied Energy, vol. 87, no. 3, pp. 934-942, 2010.
[19] S. Salcedo-Sanz, A. M. Ángel M. Pérez-Bellido, E. G. OrtizGarcía, A. Portilla-Figueras, L. Prieto, and D. Paredes, "Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction," Renewable Energy, vol. 34, no. 6, pp. 1451-1457, 2009.
[20] D. C. Dracopoulos, Evolutionary Learning Algorithms for Neural Adaptative Control, Springer, London, UK, 1997.
[21] J. Xue and Z. Shi, "Short-time traffic flow prediction based on chaos time series theory," Journal of Transportation Systems Engineering and Information Technology, vol. 8, no. 5, pp. 68-72, 2008.
[22] H. Yang, J. Li, and F. Ding, "A neural network learning algorithm of chemical process modeling based on the extended Kalman filter," Neurocomputing, vol. 70, no. 4-6, pp. 625-632, 2007.
[23] R. S. Bucy and K. D. Senne, "Digital synthesis of non-linear filters," Automatica, vol. 7, no. 3, pp. 287-298, 1971.
[24] J. Wang and Y. Xie, "Solar radiation prediction based on phase space reconstruction of wavelet neural network," Procedia Engineering, vol. 15, no. 1, pp. 4603-4607, 2011.
[25] H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, UK, 1997.
[26] T. E. Karakasidis and A. Charakopoulos, "Detection of lowdimensional chaos in wind time series," Chaos, Solitons \& Fractals, vol. 41, no. 4, pp. 1723-1732, 2009.
[27] P. Louka, G. Galanis, N. Siebert et al., "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," Journal of Wind Engineering and Industrial Aerodynamics, vol. 96, no. 12, pp. 2348-2362, 2008.
[28] M. Poncela, P. Poncela, and J. R. Perán, "Automatic tuning of Kalman filters by maximum likelihood methods for wind energy forecasting," Applied Energy, vol. 108, no. 12, pp. 349362, 2013.

## Research Article

# Two-Layer Predictive Control of a Continuous Biodiesel Transesterification Reactor 

Hongyan Shi, ${ }^{1,2}$ Dingding Wang, ${ }^{3}$ Decheng Yuan, ${ }^{2}$ and Tianran Wang ${ }^{1}$<br>${ }^{1}$ Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China<br>${ }^{2}$ College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China<br>${ }^{3}$ College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

Correspondence should be addressed to Hongyan Shi; shihongyan@sia.cn
Received 19 July 2013; Accepted 29 August 2013
Academic Editor: Baocang Ding
Copyright © 2013 Hongyan Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

A novel two-layer predictive control scheme for a continuous biodiesel transesterification reactor is presented. Based on a validated mechanistic model, the least squares (LS) algorithm is used to identify the finite step response (FSR) process model adapted in the controller. The two-layer predictive control method achieves the steady-state optimal setpoints and resolves the multivariable dynamic control problems synchronously. Simulation results show that the two-layer predictive control strategy leads to a significant improvement of control performance in terms of the optimal set-points tracking and disturbances rejection, as compared to conventional PID controller within a multiloop framework.


## 1. Introduction

With the depletion of fossil fuels and global environmental degradation, the development of alternative fuels from renewable resources has received considerable attention. Biodiesel has become the foremost alternative fuel to those refined from petroleum products. It can be produced from renewable sources, such as vegetable and animal oils, as well as from wastes, such as used cooking oil. Transesterification is the primary method of converting these oils to biodiesel [1-3]. A block diagram for a biodiesel production process by transesterification is shown in Figure 1.

A modern transesterification plant is continuous instead of batch. A continuous biodiesel production leads to better heat economization, better product purity from phase separation by removing only the portion of the layer furthest from the interface, better recovery of excess methanol in order to save on methanol cost and regulatory issues, minimal operator interference in adjusting plant parameters, and lower capital costs per unit of biodiesel produced. The same technology can also be applied to other biofuels production [4$6]$.

Biodiesel transesterification reactor is the most crucial operation unit to be controlled because any drift in standard operating condition may lead to significant changes in process variable and production quality specification [47]. These reactors have complicated dynamics and heat transfer characteristics. Moreover, they are inherently concerned with nonlinearity which arises from fluctuations of reactant concentration, reactant temperature, coolant temperature, and instrumentation noise or complex microbial interactions. The complicated nonlinear, multivariable, and coupling in nature are the fundamental control problems involved in biodiesel reactor [8, 9].

Recently, a number of reports have appeared on the controller design and dynamic optimization in continuous and batch biodiesel reactors. Mjalli et al. developed a rigorous mechanistic model of a continuous biodiesel reactor and proposed a multimodel adaptive control strategy which realized the set-point tracking and disturbance rejection [4]. Ho et al. further adopted adaptive generalized predictive control strategy to handle multivariable problems of a biodiesel reactor [8]. Wali et al. proposed an artificial intelligence technique to design online genetic-ANFIS temperature control based


Figure 1: Biodiesel production by transesterification.
on LabVIEW for a novel continuous microwave biodiesel reactor [10]. Benavides and Diwekar realized the optimal control of a batch biodiesel reactor involved optimization of the concentration based on maximum principle [11].

This work considers the advanced control strategy of biodiesel continuous transesterification reactor. Model predictive control (MPC) is one of the most popular advanced control strategies. It is a class of model-based control algorithm, which has become a complex standard process industry solving complicated constrained multivariable control problems, and widely used in the chemical and petrochemical processes [12]. The main technical characteristics of MPC, include using mathematical models and history input and output data to predict future output, combined with the established control objectives, to calculate the optimal feedback rate. Compared with the traditional multiloop PID controllers, MPC takes into account simultaneously the effects of all manipulated variables to all controlled variables. Usually successfully put into operation, MPC can significantly reduce the standard deviation of the controlled variable and then through the card edge operations, improve the overall efficiency of the control system.

In recent years, there has been an integrated steadystate optimization of the two-layer predictive control strategy in MPC industry technology [13-15]. Two-layer predictive control is divided into upper steady-state optimization (SSO) layer and lower dynamic control layer. SSO can achieve real
time optimization (RTO) objectives tracking asymptotically, independently complete local economic optimization of the corresponding MPC procedure. Specifically, the upper SSO uses steady-state gain of MPC dynamic mathematical model as the mathematical model and searches the optimum value within the constraints space of MPC. Part steady-state values of the operating or output variables will be in the position of "card edge". The calculation results of the SSO layer will be as the set- points to the lower MPC layer.

Although two-layer predictive control strategy has been widely used in many applications of chemical reactors, hardly any work was done on the biodiesel transesterification reactor. In this paper, a two-layer predictive control strategy is designed, tested, and simulated on a continuous biodiesel transesterification reactor. The scheme can amplify the advantages of both technologies in terms of process stability, and optimal and improved performances. Section 2 discusses the transesterification mechanism, which uses a validated mechanistic model of Mjalli et al. [4]. Then the twolayer predictive control strategy is developed in Section 3. Section 4 gives the control system design based on twolayer predictive control theory. Section 5 discusses model identification results and the performances of the control strategy.

## 2. Mathematical Models

The modeling of transesterification reactors starts with understanding the complex reaction kinetic mechanism. The stoichiometry of vegetable oil methanolysis reaction requires three mol of methanol (A) and one mol of triglyceride (TG) to give three mol of fatty acid methyl ester (E) and one mol of glycerol (G) [16]. The overall reaction scheme for this reaction is

$$
\begin{equation*}
\mathrm{TG}+3 \mathrm{~A} \longleftrightarrow 3 \mathrm{E}+\mathrm{G} \tag{1}
\end{equation*}
$$

The methanolysis, in turn, consists of three consecutive reversible reactions, where a mole of fatty acid methyl ester is released in each step, and monoglycerides (MG) and diglycerides (DG) are intermediate products. The stepwise reactions are



The stepwise reactions can be termed as pseudo-homogeneous catalyzed reactions, following second-order kinetics. The second-order kinetic model can be explained through the following set of differential equations [17]:

$$
\begin{gather*}
\frac{d C_{\mathrm{TG}}}{d t}=-k_{1}^{\prime} C_{\mathrm{TG}} C_{\mathrm{A}}+k_{2}^{\prime} C_{\mathrm{DG}} C_{\mathrm{E}}, \\
\frac{d C_{\mathrm{DG}}}{d t}=k_{1}^{\prime} C_{\mathrm{TG}} C_{\mathrm{A}}-k_{2}^{\prime} C_{\mathrm{DG}} C_{\mathrm{E}}-k_{3}^{\prime} C_{\mathrm{DG}} C_{\mathrm{A}}+k_{4}^{\prime} C_{\mathrm{MG}} C_{\mathrm{E}} \\
\frac{d C_{\mathrm{MG}}}{d t}=k_{3}^{\prime} C_{\mathrm{DG}} C_{\mathrm{A}}-k_{4}^{\prime} C_{\mathrm{MG}} C_{\mathrm{E}}-k_{5}^{\prime} C_{\mathrm{MG}} C_{\mathrm{A}}+k_{6}^{\prime} C_{\mathrm{GL}} C_{\mathrm{E}} \\
\frac{d C_{\mathrm{E}}}{d t}=k_{1}^{\prime} C_{\mathrm{TG}} C_{\mathrm{A}}-k_{2}^{\prime} C_{\mathrm{DG}} C_{\mathrm{E}}+k_{3}^{\prime} C_{\mathrm{DG}} C_{\mathrm{A}}-k_{4}^{\prime} C_{\mathrm{MG}} C_{\mathrm{E}} \\
+k_{5}^{\prime} C_{\mathrm{MG}} C_{\mathrm{A}}-k_{6}^{\prime} C_{\mathrm{GL}} C_{\mathrm{E}} \\
\frac{d C_{\mathrm{A}}}{d t}=-\frac{d C_{\mathrm{E}}}{d t}, \\
\frac{d C_{\mathrm{GL}}}{d t}=k_{5}^{\prime} C_{\mathrm{MG}} C_{\mathrm{A}}-k_{6}^{\prime} C_{\mathrm{GL}} C_{\mathrm{E}} \tag{3}
\end{gather*}
$$

where $C_{\mathrm{TG}}, C_{\mathrm{DG}}, C_{\mathrm{MG}}, C_{\mathrm{E}}, C_{\mathrm{A}}$, and $C_{\mathrm{GL}}$ are concentrations of triglyceride, diglyceride, monoglyceride, methyl ester, methanol, and glycerol, respectively. $k_{1}^{\prime}, k_{3}^{\prime}$, and $k_{5}^{\prime}$ are the effective rate constants for the forward reactions, and $k_{2}^{\prime}, k_{4}^{\prime}$, and $k_{6}^{\prime}$ are the effective rate constants for the reverse reactions.

The previously selected kinetic model can be formulated in terms of a general reaction equation

$$
\begin{equation*}
r_{j}=k_{j}^{\prime}\left[C_{i}\right]^{2} \tag{4}
\end{equation*}
$$

The catalyst concentration remained constant because the sidereactions that consume the catalyst were supposed to be negligible. Therefore, each effective rate constant includes the catalyst concentration $\left(C_{c a t}\right)$ and the corresponding rate constant for the catalyzed reaction [18]:

$$
\begin{equation*}
k_{j}^{\prime}=k_{j} C_{c a t} . \tag{5}
\end{equation*}
$$

The temperature influence on the reaction rate was studied from the Arrhenius equation (6) that shows the temperature dependency of the reaction rate constant

$$
\begin{equation*}
k_{j}=k_{0} e^{\left(-E_{a} / R T\right)} \tag{6}
\end{equation*}
$$

where $k_{0}$ is a constant called the preexponential factor, $E_{a}$ is the activation energy of the reaction, and $R$ is the gas constant.

In order to realize the optimization and control of continuous biodiesel production process, the model used in the paper on the basis of the second-order kinetic model jointing the material and energy balance equations as well as the dynamic equation of the coolant temperature. The material balance for each component is expressed as follows [4]:

$$
\begin{equation*}
V \frac{d C_{i}}{d t}=F_{i 0} C_{i 0}-F_{i} C_{i}-\sum_{j=1}^{n} r_{j} V . \tag{7}
\end{equation*}
$$



Figure 2: Framework of two-layer predictive control of industrial processes.

The reactor energy balance is expressed as

$$
\begin{align*}
V \sum_{i=1}^{n_{i}} C_{i} C_{p_{i}} \frac{d T}{d t}= & \left(F_{\mathrm{TG} 0} C_{\mathrm{TG} 0} C_{P_{\mathrm{TG}}}+F_{\mathrm{A} 0} C_{\mathrm{A} 0} C_{P_{\mathrm{A}}}\right)\left(T_{0}-T\right) \\
& -\left(V \sum_{i=j}^{n} r_{j} \Delta H_{j}\right)-\left(U A_{H} \Delta T\right) \tag{8}
\end{align*}
$$

The coolant fluid energy balance is expressed as

$$
\begin{equation*}
\frac{d T_{C}}{d t}=\frac{F_{C_{0}}}{V_{C}}\left(T_{C_{0}}-T_{C}\right)+\frac{U A_{H} \Delta T}{\rho_{C} V_{C} C_{P_{C}}} \tag{9}
\end{equation*}
$$

The function equation of heat transfer coefficient is approximately expressed as

$$
\begin{equation*}
U=\alpha F_{C}^{\beta} N^{\gamma}=735.5 F_{C}^{1.095} N^{0.405} \tag{10}
\end{equation*}
$$

## 3. Theory of Two-Layer Predictive Control

In modern process industries, the MPC controller is part of a multilevel hierarchy of optimization and control functions. Typically it is three-layer structure; that is, an RTO block is at the top layer, a MPC block is at the middle, and a PID block is at the bottom [19]. Therefore, under this multilevel hierarchy control system structure, the primary task of the MPC is to dynamic track the computational target calculated by the RTO. RTO layer should be optimized for the whole device.

Reference [20] proposed the framework of two-layer predictive control shown in Figure 2. SSO is added between RTO and MPC. Left branch, the SSO layer is used for recalculating the results of RTO layer, make the output steady-state target be located in the steady state gain matrix column space, so as to meet the compatibility and consistency
conditions of steady state solution. Right branch, the role of SSO is to conduct local optimization to further improve the MPC steady-state performance, which can effectively resolve the nonparty system setpoints in the given problem.

Mathematical description of the two-layer predictive control include establishing steady-state mathematical model, steady-state target calculation, and a dynamic controller design [21].
3.1. Establish Steady-State Mathematical Model. Assume an MIMO plant with $m$ control input and $p$ controlled output and the coefficients of the corresponding step response model between control input $u_{j}$ and output $y_{i}$ are given; the model vector is

$$
\begin{equation*}
a_{i j}(t)=\left[a_{i j}(1), \ldots, a_{i j}(N)\right]^{T}, \tag{11}
\end{equation*}
$$

where $i=1, \ldots, p ; j=1, \ldots, m$. $N$ in (11) denotes modeling horizon of step response model. Thus, a multistep predictive model can be obtained:

$$
\begin{equation*}
y(k+1)=y(k)+A_{1} \Delta u(k) \tag{12}
\end{equation*}
$$

where

$$
\begin{gather*}
y(k+1)=\left[\begin{array}{c}
y_{1}(k+1) \\
\vdots \\
y_{p}(k+1)
\end{array}\right] ; \quad y(k)=\left[\begin{array}{c}
y_{1}(k) \\
\vdots \\
y_{p}(k)
\end{array}\right] ; \\
\Delta u(k)=\left[\begin{array}{c}
\Delta u_{1}(k) \\
\vdots \\
\Delta u_{m}(k)
\end{array}\right] ; \quad A_{1}=\left[\begin{array}{ccc}
a_{11}(1) & \cdots & a_{1 m}(1) \\
\vdots & \vdots & \vdots \\
a_{p 1}(1) & \cdots & a_{p m}(1)
\end{array}\right] . \tag{13}
\end{gather*}
$$

Under the control increment $\Delta u(k), \ldots, \Delta u(k+M-1)$ action, the output predictive value of the system is

$$
\begin{gather*}
y(k+1)=y(k)+A_{1} \Delta u(k), \\
y(k+2)=y(k)+A_{2} \Delta u(k)+A_{1} \Delta u(k+1), \\
\vdots  \tag{14}\\
y(k+N)=y(k)+A_{N} \Delta u(k)+\cdots \\
+A_{N-M+1} \Delta u(k+M-1),
\end{gather*}
$$

abbreviated as

$$
\begin{equation*}
\partial y(k)=A \Delta u_{M}(k), \tag{15}
\end{equation*}
$$

where

$$
\partial y(k)=\left[\begin{array}{c}
y(k+1)-y(k) \\
\vdots \\
y(k+N)-y(k)
\end{array}\right]
$$

$$
\begin{gather*}
\Delta u_{M}(k)=\left[\begin{array}{c}
\Delta u(k) \\
\vdots \\
\Delta u(k+M-1)
\end{array}\right], \\
A=\left[\begin{array}{ccc}
A_{1} & & 0 \\
\vdots & \ddots & \\
A_{M} & \cdots & A_{1} \\
\vdots & & \vdots \\
A_{N} & \cdots & A_{N-M+1}
\end{array}\right] . \tag{16}
\end{gather*}
$$

The system can be written at the steady-state time

$$
\begin{equation*}
\Delta y(\infty)=A_{N} \Delta u(\infty) \tag{17}
\end{equation*}
$$

where $\Delta y(\infty)=\left[\Delta y_{1}(\infty), \Delta y_{2}(\infty), \ldots, \Delta y_{p}(\infty)\right]^{T}, \Delta u(\infty)=$ $\left[\Delta u_{1}(\infty), \Delta u_{2}(\infty), \ldots, \Delta u_{m}(\infty)\right]^{T}$ are the steady-state output increment and input increment, respectively, and $A_{N}$ is the steady-state step response coefficients matrix

$$
A_{N}=\left[\begin{array}{ccc}
a_{11}(N) & \cdots & a_{1 m}(N)  \tag{18}\\
\vdots & \vdots & \vdots \\
a_{p 1}(N) & \cdots & a_{p m}(N)
\end{array}\right] .
$$

To meet the requirements of steady-state target calculation, model (17) can also be written as

$$
\begin{equation*}
\Delta y_{\infty}(k)=A_{N} \Delta u_{\infty}(k) \tag{19}
\end{equation*}
$$

### 3.2. Steady-State Target Calculation

3.2.1. Basic Problem Description. Steady-state target calculation is to maximize economic benefits for the purpose of self-optimization under MPC existing configuration mode according to the process conditions. According to the production process characteristics and objectives, the basic problem of steady-state target calculation is the optimization process, which controlled input as cost variables, controlled output as steady-state variables. A common description of the objective function is as follows [21]:

$$
\begin{equation*}
\min _{\Delta u_{\infty}(k), \Delta y_{\infty}(k)} J=\alpha^{T} \Delta u_{\infty}(k)+\beta^{T} \Delta y_{\infty}(k) . \tag{20}
\end{equation*}
$$

Since $\Delta u_{\infty}$ and $\Delta y_{\infty}$ are linearly related, the input output variation of objective function can be unified to control the input change. The formula (20) can be unified as

$$
\begin{equation*}
\min _{\Delta u_{\infty}(k)} J=c^{T} \Delta u_{\infty}(k) \tag{21}
\end{equation*}
$$

where $c^{T}=\left[c_{1}, \ldots, c_{m}\right]$ is the cost coefficient vector, constructed by the normalized benefit, or cost of each input variable. $\Delta u_{\infty}(k)=\left[\Delta u_{\infty}^{1}, \ldots, \Delta u_{\infty}^{m}\right]^{T}$ is the steady-state change value of every input at time $k$.

Given the steady-state constraints of input and output variables, global-optimization problem of steady-state target
calculation can be described as the following linear program (LP) problem:

$$
\begin{array}{ll}
\min _{\Delta u_{\infty}(k)} & J=c^{T} \Delta u_{\infty}(k) \\
\text { s.t. } & \Delta y_{\infty}(k)=G_{u} \Delta u_{\infty}(k)+G_{f} \Delta f_{\infty}(k)+e,  \tag{22}\\
& u_{\min } \leq u_{\infty}(k)+\Delta u_{\infty}(k) \leq u_{\max } \\
& y_{\min } \leq y_{\infty}(k)+\Delta y_{\infty}(k) \leq y_{\max },
\end{array}
$$

where $G_{u}, G_{f}$ are the steady-state gain matrices of control input and disturbance variables; and $e$ is the model bias. $u_{\text {min }}, u_{\text {max }}$ are low limit and upper limit of steady-state input variables $y_{\min }, y_{\max }$ are low limit and upper limit of steady state output variables.

The global-optimization problem of steady-state target calculation can be described as the following quadratic program (QP) problem:

$$
\begin{array}{cl}
\min _{\Delta u_{\infty}(k)} & J=c^{T}\left(\Delta u_{\infty}(k)-\text { Maxprofit }\right)^{2} \\
\text { s.t. } & \Delta y_{\infty}(k)=G_{u} \Delta u_{\infty}(k)+G_{f} \Delta f_{\infty}(k)+e,  \tag{23}\\
& u_{\min } \leq u_{\infty}(k)+\Delta u_{\infty}(k) \leq u_{\max } \\
& y_{\min } \leq y_{\infty}(k)+\Delta y_{\infty}(k) \leq y_{\max }
\end{array}
$$

where Maxprofit is the potential maximum economic profit.

### 3.2.2. Feasibility Judgment and Soft Constraint Adjustment.

 Mathematically, optimization feasibility is the existence problem of the optimal solution. Feasibility of steady-state target calculation means that optimal steady state of input-output should meet their operating constraints; if feasible solution does not exist, the optimization calculation has no solution. The solving process is as follows: first, judge the existence of space domain formed by the constraints and if there is in it for optimization, if does not exist, then through the soft constraints adjustment to obtain the feasible space domain, and then to solve.Soft constraints adjustment is an effective way to solve infeasible optimization [22, 23]. By relaxing the output constraints within the hard constraints, increasing the optimization problem feasible region that feasible solution to be optimized. Hard constraints refer to unalterable constraints limited by the actual industrial process.

Engineering standards of the priority strategy of soft constraints adjustment are the following: give priority to meet the highly important operating constraints, and allow less
important operating constraints to be violated appropriately under the premise of satisfying the engineering constraints.

Considering the following constraints (24), constituted by steady-state model input constraints and output constraints containing slack variables, the priority rank is " $N$ ", where

$$
\begin{gather*}
\Delta y_{\infty}(k)=G_{u} \Delta u_{\infty}(k)+G_{f} \Delta f_{\infty}(k)+e \\
u_{L L} \leq u_{\infty}(k)+\Delta u_{\infty}(k) \leq u_{H L} \\
y_{L L}^{j}-\varepsilon_{2}^{j} \leq y_{\infty}(k)+\Delta y_{\infty}(k) \leq y_{H L}^{j}+\varepsilon_{1}^{j} \\
\varepsilon_{1}^{j} \geq 0, \quad \varepsilon_{2}^{j} \geq 0  \tag{24}\\
\varepsilon_{1}^{j} \leq y_{H H L}-y_{H L} \\
\varepsilon_{2}^{j} \leq y_{L L}-y_{L L L} \\
j=1, \ldots, N
\end{gather*}
$$

The algorithm steps of feasibility judgment and soft constraint adjustment based on the priority strategy are as follows.

Step 1. Initialization: according to the characteristics of the output variables and process conditions, set the upper and lower output constraints priority ranks, the same priority rank setting adjustments according to actual situation constraint weights.

Step 2. According to the priority ranks, judge the feasibility and adjust the soft constraints in accordance with the ranks from large to small. Under a larger priority rank if cannot find a feasible solution, the constraints of the rank will be relaxed to hard constraints, and then consider less priority rank constraints, until we find a feasible solution.

Step 3. Then the steady-state target calculation entered the stage of economy optimization or target tracking.

For Step 2, constraints of the highest priority rank $N$ are adjusted first by solving the following optimization problem:

$$
\begin{array}{ll}
\min _{\varepsilon^{N}} & J=\left(W^{N}\right)^{T} \varepsilon^{N}, \quad\left(W^{N}\right)^{T}=\left[W_{1}^{N}, \ldots, W_{2 \times n_{N}}^{N}\right] \\
\text { s.t. } & \Theta^{N} Z^{N}=b^{N},  \tag{25}\\
& \Omega^{N} Z^{N} \leq \Psi^{N},
\end{array}
$$

where

$$
\begin{gathered}
Z^{N}=\left[X_{1}^{T}, X_{2}^{T},\left(X_{3}^{1}\right)^{T}, \ldots,\left(X_{3}^{N}\right)^{T},\left(X_{4}^{1}\right)^{T}, \ldots,\right. \\
\left.\left(X_{4}^{N}\right)^{T},\left(\varepsilon_{1}^{N}\right)^{T},\left(\varepsilon_{2}^{N}\right)^{T},\left(\varepsilon_{1}^{N}\right)^{T},\left(\varepsilon_{2}^{N}\right)^{T}\right]^{T}, \\
\Omega^{N}=\text { block-diag }\left(-I_{m},-I_{m},-I_{n_{1}}, \ldots,-I_{n_{N}},\right. \\
\\
\left.-I_{n_{1}}, \ldots,-I_{n_{N}},-I_{n_{N}},-I_{n_{N}}, I_{n_{N}}, I_{n_{N}}\right),
\end{gathered}
$$

$$
\begin{align*}
& \Psi^{N}=\left[\left(0_{m \times 1}\right)^{T},\left(0_{m \times 1}\right)^{T},\left(0_{n_{1} \times 1}\right)^{T}, \ldots,\left(0_{n_{N} \times 1}\right)^{T},\right. \\
& \left(0_{n_{1} \times 1}\right)^{T}, \ldots,\left(0_{n_{N} \times 1}\right)^{T},\left(0_{n_{N} \times 1}\right)^{T},\left(0_{n_{N} \times 1}\right)^{T}, \\
& \left.\left(y_{H H L}^{N}-y_{H L}^{N}\right)^{T},\left(y_{L L}^{N}-y_{L L L}^{N}\right)^{T}\right]^{T} \text {, } \\
& b^{N}=\left[\begin{array}{c}
u_{H L}-u_{L L} \\
G_{u}^{1} u_{\infty}(k)-G_{u}^{1} u_{L L}(k)+y_{H L}^{1}-y_{\infty}^{1}(k)-G_{f}^{1} \Delta f_{\infty}(k)-e^{1} \\
\vdots \\
G_{u}^{N} u_{\infty}(k)-G_{u}^{N} u_{L L}(k)+y_{H L}^{N}-y_{\infty}^{N}(k)-G_{f}^{N} \Delta f_{\infty}(k)-e^{N} \\
G_{u}^{1} u_{H L}-G_{u}^{1} U_{\infty}(k)+Y_{\infty}^{1}(k)+G_{f}^{1} \Delta f_{\infty}(k)-y_{L L}^{1}+e^{1} \\
\vdots \\
G_{u}^{N} u_{H L}-G_{u}^{N} U_{\infty}(k)+y_{\infty}^{N}(k)+G_{f}^{N} \Delta f_{\infty}(k)-y_{L L}^{N}+e^{N} \\
0 \\
0
\end{array}\right], \\
& \Theta^{N}=\left[\begin{array}{cccccccccccc}
I_{m} & I_{m} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
G_{u}^{1} & 0 & 0 & \cdots & 0 & I_{n_{1}} & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & 0 & \cdots & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\
G_{u}^{N} & 0 & 0 & \cdots & 0 & 0 & 0 & I_{n_{N}} & -I_{n_{N}} & 0 & 0 & 0 \\
0 & G_{u}^{1} & I_{n_{1}} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & 0 & \ddots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
0 & G_{u}^{N} & 0 & 0 & I_{n_{N}} & 0 & \cdots & 0 & 0 & -I_{n_{N}} & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & I_{n_{N}} & 0 & -I_{n_{N}} & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & I_{n_{N}} & 0 & -I_{n_{N}}
\end{array}\right] . \tag{26}
\end{align*}
$$

Solving (25) may appear in three different cases, respectively: if (25) is feasible, and the optimum solution is $\varepsilon^{N}=0$, subject to $J=0$, that is, no need for soft constraints adjustment, directly solve the original problem (22); if (25) is feasible, but $\varepsilon^{N} \neq 0$, just need to relax constraints of priority ranks $N$, and further optimization solution; if (25) is infeasible, not get a feasible solution to soft constraints adjustment of the priority rank $N$, relaxing the constraints of the priority rank $N$ to hard constraints; that is,

$$
\begin{gather*}
\varepsilon_{1}^{N}=y_{H H L}^{N}-y_{H L}^{N}  \tag{27}\\
\varepsilon_{2}^{N}=y_{L L}^{N}-y_{L L L}^{N}
\end{gather*}
$$

Go to the procedure of judging rank $N-1$ constraints

$$
\begin{array}{ll}
\min _{\varepsilon^{N-1}} & J=\left(W^{N-1}\right)^{T} \varepsilon^{N-1},\left(W^{N-1}\right)^{T}=\left[W_{1}^{N-1}, \ldots, W_{2 \times n_{N-1}}^{N-1}\right] \\
\text { s.t. } & \Theta^{N-1} Z^{N-1}=b^{N-1}, \\
& \Omega^{N-1} Z^{N-1} \leq \Psi^{N-1} . \tag{28}
\end{array}
$$

For (28), the matrix form is the same with priority rank $N$, only in the corresponding position of $\varepsilon^{N-1}$ to replace $\varepsilon^{N}, b^{N-1}$ matrix is adjusted

$$
b^{N-1}=\left[\begin{array}{c}
u_{H L}-u_{L L}  \tag{29}\\
G_{u}^{1} u_{\infty}(k)-G_{u}^{1} u_{L L}(k)+y_{H L}^{1}-y_{\infty}^{1}(k)-G_{f}^{1} \Delta f_{\infty}(k)-e^{1} \\
\vdots \\
G_{u}^{N} u_{\infty}(k)-G_{u}^{N} u_{L L}(k)+y_{H L}^{N}-y_{\infty}^{N}(k)-G_{f}^{N} \Delta f_{\infty}(k)-e^{N}+\left(y_{H H L}^{N}-y_{H L}^{N}\right)^{T} \\
G_{u}^{1} u_{H L}-G_{u}^{1} u_{\infty}(k)+y_{\infty}^{1}(k)+G_{f}^{1} \Delta f_{\infty}(k)-y_{L L}^{1}+e^{1} \\
\vdots \\
G_{u}^{N} u_{H L}-G_{u}^{N} u_{\infty}(k)+y_{\infty}^{N}(k)+G_{f}^{N} \Delta f_{\infty}(k)-y_{L L}^{N}+e^{N}+\left(y_{L L}^{N}-y_{L L L}^{N}\right)^{T} \\
0 \\
0
\end{array}\right] .
$$

$N-1$ rank and $N$ rank are the same for the soft constraints adjustment processing, until the end of constraint adjustment of the priority rank 1. If all ranks of constraints are relaxed to the hard constrain and a feasible solution still can't be found, then the original problem of soft constraints adjustment is infeasible and needs to be redesigned.
3.3. Dynamic Controller Design. In engineering applications, dynamic matrix control (DMC) algorithm is one of the most widely used MPC algorithms based on the step response model of the plant. This paper adopts DMC and steady-state target calculation integration strategy.

The difference is that the general DMC algorithms have no requirements on the steady-state position of the control input, and they only require the controlled output as close as possible to arrive at its set point. However, the integration strategy DMC requires both input and output variables to approach their steady-state targets $\left(\mathbf{u}_{s}, \mathbf{y}_{s}\right)$ as far as possible. The algorithm has three basic characteristics: predictive model, receding horizon optimization, and feedback correction [24].
3.3.1. Predictive Model. Based on system process step response model, at the current time $k$, the future $P$-step prediction output can be written as follows:

$$
\begin{equation*}
\widetilde{\mathbf{y}}_{P M}(k)=\widetilde{\mathbf{y}}_{P 0}(k)+\mathbf{A} \Delta \mathbf{u}_{M}(k), \tag{30}
\end{equation*}
$$

where $P$ denotes the prediction horizon, $M$ is the control horizon, $\mathbf{A}$ is the prediction matrix composed by the corresponding step response coefficients, $\widetilde{\mathbf{y}}_{P 0}$ is the initial output prediction value when control action starting from the present time does not change, $\Delta \mathbf{u}_{M}(k)$ is the prediction incremental in $M$ control horizon, and $\widetilde{\mathbf{y}}_{P M}(k)$ is the future $P$ step prediction output under $M$-step control action change. Among them

$$
\begin{array}{cc}
\tilde{y}_{P M}(k)=\left[\begin{array}{c}
\tilde{y}_{1, P M}(k) \\
\vdots \\
\tilde{y}_{p, P M}(k)
\end{array}\right], & \tilde{y}_{P 0}(k)=\left[\begin{array}{c}
\tilde{y}_{1, P 0}(k) \\
\vdots \\
\tilde{y}_{p, P 0}(k)
\end{array}\right], \\
\Delta \mathbf{u}_{M}(k)=\left[\begin{array}{c}
\Delta u_{1, M}(k) \\
\vdots \\
\Delta u_{m, M}(k)
\end{array}\right], & \mathbf{A}=\left[\begin{array}{ccc}
A_{11} & \cdots & A_{1 m} \\
\vdots & \ddots & \vdots \\
A_{p 1} & \cdots & A_{p m}
\end{array}\right] .
\end{array}
$$

3.3.2. Receding Horizon Optimization. In the receding horizon optimization process, control increment can be obtained in every execution cycle by minimizing the following performance index:

$$
\begin{align*}
\min _{\Delta \mathbf{u}_{M}(k)} J(k)= & \left\|\mathbf{w}(k)-\tilde{\mathbf{y}}_{P M}(k)\right\|_{\mathbf{Q}}^{2}+\|\boldsymbol{\varepsilon}(k)\|_{\mathrm{S}}^{2}  \tag{32}\\
& +\left\|\mathbf{u}_{M}(k)-\mathbf{u}_{\infty}\right\|_{\mathrm{T}}^{2}+\left\|\Delta \mathbf{u}_{M}(k)\right\|_{\mathbf{R}}^{2} .
\end{align*}
$$

Subject to the model

$$
\begin{equation*}
\tilde{\mathbf{y}}_{P M}(k)=\tilde{\mathbf{y}}_{P 0}(k)+\mathbf{A} \Delta \mathbf{u}_{M}(k) . \tag{33}
\end{equation*}
$$

Subject to bound constraints

$$
\mathbf{y}_{\min }-\boldsymbol{\varepsilon} \leq \widetilde{\mathbf{y}}_{P M}(k) \leq \mathbf{y}_{\max }+\boldsymbol{\varepsilon}
$$

$$
\begin{gather*}
\mathbf{u}_{\min } \leq \mathbf{u}_{M} \leq \mathbf{u}_{\max } \\
\Delta \mathbf{u}_{\min } \leq \Delta \mathbf{u}_{M}(k) \leq \Delta \mathbf{u}_{\max } \tag{34}
\end{gather*}
$$

where $\boldsymbol{\varepsilon}$ denotes the slack variables, guaranteeing the feasibility of the DMC optimization, and $w(k)=\left[w_{1}(k), \ldots, w_{p}(k)\right]^{T}$ is the setpoint of controlled output obtained from upper SSO layer. $\mathbf{Q}, \mathbf{R}$ are the weight coefficient matrix

$$
\begin{gather*}
\mathbf{Q}=\operatorname{block}-\operatorname{diag}\left(Q_{1}, \ldots, Q_{p}\right) \\
\mathbf{Q}_{i}=\operatorname{diag}\left(q_{i}(1), \ldots, q_{i}(P)\right), \quad i=1, \ldots, p  \tag{35}\\
\mathbf{R}=\operatorname{block}-\operatorname{diag}\left(R_{1}, \ldots R_{m}\right) \\
\mathbf{R}_{j}=\operatorname{diag}\left(r_{i}(1), \ldots, r_{i}(M)\right), \quad j=1, \ldots, m
\end{gather*}
$$

Through the necessary conditions of extreme value $\partial J / \partial \Delta u_{M}(k)=0$, the optimal increment of control input can be obtained:

$$
\begin{equation*}
\Delta \mathbf{u}_{M}(k)=\left(\mathbf{A}^{T} \mathbf{Q} \mathbf{A}+\mathbf{R}\right)^{-1} \mathbf{A}^{T} \mathbf{Q}\left[\mathbf{w}(k)-\widetilde{\mathbf{y}}_{P 0}(k)\right] \tag{36}
\end{equation*}
$$

The instant increment can be calculated as follows:

$$
\begin{equation*}
\Delta \mathbf{u}(k)=\mathbf{L D}\left[\mathbf{w}(k)-\widetilde{\mathbf{y}}_{P 0}(k)\right] \tag{37}
\end{equation*}
$$

where $\mathbf{D}=\left(\mathbf{A}^{T} \mathbf{Q A}+\mathbf{R}\right)^{-1} \mathbf{A}^{T} \mathbf{Q}$; remark the operation of only the first element with

$$
L=\left[\begin{array}{ccccccccc}
1 & 0 & \cdots & 0 & & & & & 0  \tag{38}\\
& & & & \ddots & & & & \\
& & 0 & & & 1 & 0 & \cdots & 0
\end{array}\right]
$$

3.3.3. Feedback Correction. The difference between the process sample values by the present moment $k$ and prediction values of (30) is

$$
e(k+1)=\left[\begin{array}{c}
e_{1}(k+1)  \tag{39}\\
\vdots \\
e_{p}(k+1)
\end{array}\right]=\left[\begin{array}{c}
y_{1}(k+1)-\tilde{y}_{1,1}(k+1 \mid k) \\
\vdots \\
y_{p}(k+1)-\tilde{y}_{p, 1}(k+1 \mid k)
\end{array}\right],
$$

where $\tilde{y}_{i, 1}(k+1 \mid k)$ is the first element of $\tilde{y}_{i, P M}(k+1 \mid k)$, and the corrected output prediction value can be obtained using the error vector; that is,

$$
\begin{equation*}
\widetilde{\mathbf{y}}_{\mathrm{cor}}(k+1)=\widetilde{\mathbf{y}}_{N 1}(k)+\mathbf{H} e(k+1), \tag{40}
\end{equation*}
$$

where $\widetilde{\mathbf{y}}_{N 1}(k)=\widetilde{\mathbf{y}}_{N 0}(k)+\mathbf{A}_{N} \Delta \mathbf{u}, \widetilde{\mathbf{y}}_{N 0}(k)$ is the future $N$ moment initial prediction value when all of the input remained unchanged at the time $k ; \widetilde{\mathbf{y}}_{N 1}(k)$ is the future $N$ moment output prediction value under one-step control input action; $H$ is the error correct matrix. Then using a shift matrix $S$, next time the initial prediction value can be obtained, which is

$$
\begin{equation*}
\tilde{y}_{\mathrm{N} 0}(k+1)=S \tilde{y}_{\mathrm{cor}}(k+1), \tag{41}
\end{equation*}
$$



Figure 3: Two-layer predictive framework of biodiesel process.
where

$$
S=\left[\begin{array}{ccccc}
0 & 1 & & & 0  \tag{42}\\
& 0 & 1 & & \\
& & \ddots & \ddots & \\
& & & 0 & 1 \\
0 & & & & 1
\end{array}\right]_{N * N} .
$$

## 4. Control System Design

In the biodiesel reactor control, multiloops are necessary to stabilize the plant. One loop is needed to maintain the set point of specifying the product purity, and another loop is needed to ensure an optimal yield of biodiesel and to minimize the generation of unwanted by-products even in the presence of disturbances.

To achieve these goals, the control loop configurations analysis is meaningful. Based on the analysis of Mjalli et al. [4], the favorable pairings are as follows: the biodiesel concentration $\left(C_{E}\right)$ is maintained by manipulating reactant flow rate $\left(F_{o}\right)$, the reactor temperature $(T)$ is maintained by manipulating coolant flow rate $\left(F_{c}\right)$, respectively, and the effect of stirred rotational speed on the reactor output is insignificant, and it would be regarded as one of disturbances to the control system. The relative gain array (RGA) shows that there are some interactions among the controlled and manipulated variables which make two-layer predictive controller better qualified.

Consequently, the two-layer predictive controller is designed to handle a $2 \times 2$ system of inputs and outputs. The controlled output variables include biodiesel concentration $\left(C_{E}\right)$ and reactor temperature $(T)$; the manipulated variables include reactant flow rate $\left(F_{o}\right)$ and coolant flow rate $\left(F_{c}\right)$. It is very important for a reactor to handle the disturbances in the feed concentration and initial temperatures, as these disturbances heavily change the system performance.

The design of the control loop based on the two-layer predictive control strategy for the biodiesel reactor is shown in Figure 3. The SSO layer searches the optimal output setpoints $C_{E s s}$ and $T_{s s}$ according to the economic optimization goal of the actual production process. The MPC layer selects the real-time control actions $\Delta u$ to complete the dynamic tracking control.

## 5. Simulation Results and Analysis

5.1. Model Identification. For the two-layer predictive control scheme to be successful, process modeling plays a key role in capturing the varying dynamics of the system. Section 4 shows that the biodiesel process is a two-input two-output multivariable process. The process nonlinear model was programmed and simulated in Matlab as a function. Simulation results show system is open stable process.

Firstly, generalized binary noise (GBN) signal is selected as the excitation signal. GBN signals switch between $a$ and $-a$ according to the following rules:

$$
\begin{gather*}
P[u(t)=-u(t-1)]=p_{s w}, \\
P[u(t)=u(t-1)]=1-p_{s w}, \tag{43}
\end{gather*}
$$

where $p_{s w}$ is transition probability; $T_{\min }$ is defined as the sampling time of the signal held constant; $T_{s w}$ is time interval of twice conversion. The average conversion time and power spectrum are, respectively,

$$
\begin{gather*}
E T_{s w}=\frac{T_{\min }}{p_{s w}} \\
\Phi_{u}(\omega)=\frac{\left(1-q^{2}\right) T_{\min }}{1-2 q \cos T_{\min } \omega+q^{2}}, \quad q=1-2 p_{s w} \tag{44}
\end{gather*}
$$

Next, least squares (LS) identification method is used to estimate the process model parameters. Suppose an MIMO plant with $m$ input $p$ output, for the $i$ th output of the finite impulse response (FIR) model, is described as

$$
\begin{equation*}
y_{i}(k)=\sum_{j=1}^{m} \sum_{l=1}^{N} h_{i j l} u_{j}(k-l) . \tag{45}
\end{equation*}
$$

Consider experimental tests of collecting input sequence

$$
\begin{array}{cccc}
u_{1}(1) & u_{1}(2) & \cdots & u_{1}(L) \\
\vdots & & & \vdots  \tag{46}\\
u_{m}(1) & u_{m}(2) & \cdots & u_{m}(L)
\end{array}
$$

and output sequence

$$
\begin{array}{cccc}
y_{1}(1) & y_{1}(2) & \cdots & y_{1}(L) \\
\vdots & & & \vdots  \tag{47}\\
y_{p}(1) & y_{p}(2) & \cdots & y_{p}(L) .
\end{array}
$$



Figure 4: Biodiesel concentration prediction result and relative error under reactor flow rate $F_{o}$ action.


Figure 5: Reactor temperature prediction result and relative error under reactor flow rate $F_{o}$ action.

Consider matching between data and models; the introduction of residuals for each output can be independently expressed as follows:

$$
\begin{equation*}
y_{i}(k)=\varphi(k) \theta_{i}+e(k) . \tag{48}
\end{equation*}
$$

$$
\begin{gather*}
y_{i}=\left[\begin{array}{c}
y_{i}(N+1) \\
y_{i}(N+2) \\
\vdots \\
y_{i}(L)
\end{array}\right], \quad e=\left[\begin{array}{c}
e(N+1) \\
e(N+2) \\
\vdots \\
e(L)
\end{array}\right],  \tag{50}\\
\Phi=\left[\begin{array}{ccccccc}
u_{1}(N) & u_{1}(N-1) & \cdots & u_{1}(1) & & u_{m}(N) & u_{m}(N-1) \\
u_{1}(N+1) & u_{1}(N) & \cdots & u_{1}(2) & u_{m}(N+1) & u_{m}(N) & \cdots \\
\vdots & \vdots & & \vdots & \cdots & \vdots & u_{m}(1) \\
u_{1}(L-1) & u_{1}(L-2) & u_{1}(L-N) & u_{m}(L-1) & u_{m}(L-2) & u_{m}(L-N)
\end{array}\right] .
\end{gather*}
$$

Minimize the squared residuals

$$
\begin{equation*}
\min J=e^{T} e=[y-\Phi \theta]^{T}[y-\Phi \theta] . \tag{51}
\end{equation*}
$$

Obtain the optimal estimate

$$
\begin{equation*}
\widehat{\theta}=\left[\Phi^{T} \Phi\right]^{-1} \Phi^{T} y \tag{52}
\end{equation*}
$$

Matrix form is written as

$$
\begin{equation*}
y_{i}=\Phi \theta_{i}+e \tag{49}
\end{equation*}
$$

where

For the model predictive controller design, the FIR model of system identification needs to be further converted into finite step response (FSR) model. The relationship between FSR coefficients and FIR coefficients is as follows:

$$
\begin{equation*}
g_{j}=\sum_{i=1}^{j} h_{j} . \tag{53}
\end{equation*}
$$



Figure 6: Biodiesel concentration prediction result and relative error under reactor flow rate $F_{c}$ action.

## Coefficients matrix of FSR is

$$
G_{l}^{u}=\left[\begin{array}{cccc}
s_{11 l} & s_{12 l} & \cdots & s_{1 m l}  \tag{54}\\
s_{21 l} & s_{22 l} & \cdots & s_{2 m l} \\
\vdots & \vdots & \ddots & \vdots \\
s_{p 1 l} & s_{p 2 l} & \cdots & s_{p m l}
\end{array}\right] .
$$

Finally, (11)-(19) are used to create a steady-state mathematical model of two-layer prediction control. The concrete simulation process is as follows.

In the work, GBN as the excitation signal was added to the model input to produce output data. The parameters of GBN signal applied to the first input are $T_{s w}=65, a m p=0.1$ the parameters of GBN applied to the second input are $T_{s w}=$ $65, a m p=0.005$, both the conversion probabilities are taken to be $P_{s w}=1 / T_{s w}$. Simulation time $t=2000 \mathrm{~s}$, and sample time equals 2 s , under each input excitation, corresponding to two sets of output data each set of data capacity is 1000 . Among them, the former 500 data as model identification, the remaining data are used as model validations, and FSR model length value is taken as 200.

Under the action of two inputs, reactant flow rate $F_{o}$ and coolant flow rate $F_{c}$, respectively, predicted value, actual value, and the relative error of two outputs biodiesel concentration $C_{E}$ and reactor temperature $T$ were shown in Figures $4,5,6$, and 7 . Figures $4-7$ show that relative error is small enough, and the model can describe $C_{E}$ and $T$ change trends under $F_{o}$ and $F_{c}$.

Figures 8 and 9 give the two output step response curves under two input $F_{o}, F_{c}$ action, respectively, further shows the multiple-input multiple-output system is open-loop stable


Figure 7: Reactor temperature prediction result and relative error under reactor flow rate $F_{c}$ action.


Figure 8: Step response curve of biodiesel concentration and reactor temperature, respectively, under $F_{o}$ action.
and the step response model has been identified successfully. The FSR model will be utilized to represent the actual process in latter optimization and controller design.
5.2. Dynamic Simulation. To validate the effectiveness and immunity in two-layer predictive control, the models obtained in Section 5.1 are used in the simulations.


Figure 9: Step response curve of biodiesel concentration and reactor temperature, respectively, under $F_{c}$ action.

The reaction rate constants come from [18] under the common industrial conditions of $6: 1$ methanol/oil mole ratio, $1.0 \mathrm{wt} \%$ catalyst KOH , and 600 rpm stirrer rotational speed. These kinetics parameters can be considered as constants. The initial operating conditions refer to the literature [4] the validated data. According to these parameters and reaction conditions, the simulation of biodiesel transesterification reactor can be carried out.

The economic optimization method described in (22) is adopted as SSO whose main parameters are selected as follows: the cost coefficients of control input in steady-state optimization are set to $[1 ;-1]$, the input $F_{o}$ is constrained between 0 and $0.2 \mathrm{~m}^{3} / \mathrm{s}$, the input $F_{c}$ is constrained between 0 and $0.1 \mathrm{~m}^{3} / \mathrm{s}$, and the output $C_{E}$ is constrained between $3.0536 \mathrm{kmol} / \mathrm{m}^{3}$ and $3.196 \mathrm{kmol} / \mathrm{m}^{3}$, the output $T$ is constrained between 337.77 K and 338.25 K .

The parameters of the dynamic control layer adopted the unconstrained DMC algorithm: the modeling time domain $N=200$, prediction horizon $P=200$, control horizon $M=$ 20. The weight coefficient values of weight matrix $Q$ and $R$ equal to 10 and 1000 , respectively.

Conventional PID controller has also been designed in this simulation for comparison of performance to two-layer predictive controller. The parameters of PID controller for $C_{E}$ with $F_{o}$ control loop are $k p=-6 e-5, k i=-0.05$, and $k d=0$; the parameters for $T$ with $F_{c}$ control loop are $k p=-0.02$, $k i=-0.001$, and $k d=0$. The simulations of general PID controller and two-layer predictive controller are compared to validate the performance of the latter algorithm, whose results are shown in Figures 10 and 11.

As Figures 10 and 11 show, the two-layer predictive controller starts running at the time $t=0$. The results of steady state optimization are

$$
\begin{equation*}
y_{s s}=[3.196,337.77], \quad u_{s s}=[0.073,0.0062] . \tag{55}
\end{equation*}
$$



Figure 10: Biodiesel concentration and controller moves of twolayer predictive controller and PID controller.

The optimized values as the setpoints were send to the lower layer DMC. In the beginning, the closed loop response of the two-layer predictive controller was a little sluggish in bringing the biodiesel concentration back the optimum steady-state values, this is because that the algorithm enter the constraint adjustment stage based on the priority strategy which adjusting the upper limit and lower limit to be handled. About At the time $t=400$, the response gradually becomes stable. It can be seen that the two-layer predictive controller preceded the PID controller in terms of the ability to attain lower overshoot, smaller oscillation, and faster response time.

Considering the actual application, the control input is also an important indicator of good or bad controller. From Figures 10 and 11, the two-layer predictive controller has much more stable controller moves than does PID that meets the practical implementation constrains.


Figure 11: Reactor temperature and controller moves of two-layer predictive controller and PID controller.

To challenge the stability of two-layer predictive controller, some disturbances were exerted alone and at the same time. The chosen disturbance variables include coolant input temperature ( $T_{c 0}$ ), feed temperature $\left(T_{0}\right)$, triglyceride initial concentration ( $C_{T G 0}$ ), and stirrer rotational speed ( $N$ ). After the system has attained the steady state, The nominal values of $T_{c 0}, T_{0}$ were increased 3 K , respectively, and $C_{\mathrm{TG} 0}, N$ were increased $5 \%$, respectively, at the time $t=1000$ s. Figures 12 and 13 show the biodiesel concentration and reactor temperature profiles when these disturbance variables were introduced.

Figures 12 and 13 showed satisfactory rejection of all disturbances. Two-layer predictive controller was able to bring back the controlled variables to their setpoints in less than 1000 s , and overshoot was within the acceptable range. For the biodiesel concentration loop, the initial concentration


Figure 12: Biodiesel concentration and controller moves of four individual disturbance variables effects.
$C_{\mathrm{TG} 0}$ has the highest effect, with an overshoot of less than $0.01 \mathrm{kmol} / \mathrm{m}^{3}$. For the reactor temperature loop, the feed temperature $T_{0}$ has the largest effect, with an overshoot of less than 0.33 K . For the two loops, the stirrer rotational speed almost has no effect on the controlled variables.

## 6. Conclusions

Biodiesel transesterification reactor control has become very important in recent years due to the difficulty in controlling the complex and highly nonlinear dynamic behavior. In this paper, a novel two-layer predictive control scheme for a continuous biodiesel transesterification reactor has been proposed. The SSO layer achieved optimal output setpoints according to the local economic optimization goal of the actual production process, and the MPC layer realized the

(a)


$$
\text { --- } T_{c 0} \text { increase 3K } \quad \cdots \cdots C_{T G 0} \text { increase } 5 \%
$$

(b)

Figure 13: Reactor temperature and controller moves of four individual disturbance variables effects.
dynamic tracking control. The main aim was to optimize and control the biodiesel concentration and reactor temperature in order to obtain the product of the highest quality at the lower cost. With steady-state optimum target calculation and DMC algorithm implement, the performance of the two-layer predictive controller was superior to that of a conventional PID controller. The two-layer predictive control is not only stable but also tracks set points more efficiently with minimal overshoots and shorter settling times. Moreover, it exhibits good disturbance rejection characteristics.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (61034008) and the Science Research

Foundation of Liaoning Provincial Department of Education (L2012145).

## References

[1] D. Y. C. Leung, X. Wu, and M. K. H. Leung, "A review on biodiesel production using catalyzed transesterification," Applied Energy, vol. 87, no. 4, pp. 1083-1095, 2010.
[2] S. Shahla, N. G. Cheng, and R. Yusoff, "An overview on transesterification of natural oils and fats," Biotechnology and Bioprocess Engineering, vol. 15, no. 6, pp. 891-904, 2010.
[3] N. N. A. N. Yusuf, S. K. Kamarudin, and Z. Yaakub, "Overview on the current trends in biodiesel production," Energy Conversion and Management, vol. 52, no. 7, pp. 2741-2751, 2011.
[4] F. S. Mjalli, L. K. San, K. C. Yin, and M. A. Hussain, "Dynamics and control of a biodiesel transesterification reactor," Chemical Engineering and Technology, vol. 32, no. 1, pp. 13-26, 2009.
[5] T. Eevera, K. Rajendran, and S. Saradha, "Biodiesel production process optimization and characterization to assess the suitability of the product for varied environmental conditions," Renewable Energy, vol. 34, no. 3, pp. 762-765, 2009.
[6] C. S. Bildea and A. A. Kiss, "Dynamics and control of a biodiesel process by reactive absorption," Chemical Engineering Research and Design, vol. 89, no. 2, pp. 187-196, 2011.
[7] L. Zong, S. Ramanathan, and C.-C. Chen, "Fragment-based approach for estimating thermophysical properties of fats and vegetable oils for modeling biodiesel production processes," Industrial and Engineering Chemistry Research, vol. 49, no. 2, pp. 876-886, 2010.
[8] Y. K. Ho, F. S. Mjalli, and H. K. Yeoh, "Multivariable adaptive predictive model based control of a biodiesel transesterification reactor," Journal of Applied Sciences, vol. 10, no. 12, pp. 1019-1027, 2010.
[9] H. Y. Kuen, F. S. Mjalli, and Y. H. Koon, "Recursive least squaresbased adaptive control of a biodiesel transesterification reactor," Industrial and Engineering Chemistry Research, vol. 49, no. 22, pp. 11434-11442, 2010.
[10] W. A. Wali, A. I. Al-Shamma, K. H. Hassan, and J. D. Cullen, "Online genetic-ANFIS temperature control for advanced microwave biodiesel reactor," Journal of Process Control, vol. 22, pp. 1256-1272, 2012.
[11] P. T. Benavides and U. Diwekar, "Optimal control of biodiesel production in a batch reactor-part I: deterministic control," Fuel, vol. 94, pp. 211-217, 2012.
[12] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," Control Engineering Practice, vol. 11, no. 7, pp. 733-764, 2003.
[13] T. A. Johansen and A. Grancharova, "Approximate explicit constrained linear model predictive control via orthogonal search tree," IEEE Transactions on Automatic Control, vol. 48, no. 5, pp. 810-815, 2003.
[14] T. Zou, B. C. Ding, and D. Zhang, Model Predictive Control Engineering Applications Introduction, Chemical Industry Press, Beijing, China, 2010.
[15] A. Nikandrov and C. L. E. Swartz, "Sensitivity analysis of LPMPC cascade control systems," Journal of Process Control, vol. 19, no. 1, pp. 16-24, 2009.
[16] H. Noureddini and D. Zhu, "Kinetics of transesterification of soybean oil," Journal of the American Oil Chemists' Society, vol. 74, no. 11, pp. 1457-1463, 1997.
[17] A.-F. Chang and Y. A. Liu, "Integrated process modeling and product design of biodiesel manufacturing," Industrial and Engineering Chemistry Research, vol. 49, no. 3, pp. 1197-1213, 2010.
[18] G. Vicente, M. Martínez, and J. Aracil, "Kinetics of Brassica carinata oil methanolysis," Energy and Fuels, vol. 20, no. 4, pp. 1722-1726, 2006.
[19] R. Scattolini, "Architectures for distributed and hierarchical Model Predictive Control-a review," Journal of Process Control, vol. 19, no. 5, pp. 723-731, 2009.
[20] T. Zou, H. Q. Li, B. C. Ding, and D. D. Wang, "Compatibility and uniqueness analyses of steady state solution for multi-variable predictive control systems," Acta Automatica Sinica, vol. 39, pp. 519-529, 2013.
[21] D. E. Kassmann, T. A. Badgwell, and R. B. Hawkins, "Robust steady-state target calculation for model predictive control," AIChE Journal, vol. 46, no. 5, pp. 1007-1024, 2000.
[22] Y. G. Xi and H. Y. Gu, "Feasibility analysis of constrained multi-objective multi-degree-of-freedom optimization control in industrial processes," Acta Automatica Sinica, vol. 24, pp. 727732, 1998.
[23] T. Zou, H. Q. Li, X. X. Zhang, Y. Gu, and H. Y. Su, "Feasibility and soft constraint of steady state target calculation layer in LPMPC and QP-MPC cascade control systems," in Proceedings of the International Symposium on Advanced Control of Industrial Processes (ADCONIP '11), pp. 524-529, May 2011.
[24] Y. G. Xi, Predictive Control, National Defense Industry Press, Beijing, China, 1993.

## Research Article

# Research on the WSN Node Localization Based on TOA 

Qing-hui Wang, Ting-ting Lu, Meng-long Liu, and Li-feng Wei<br>School of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China<br>Correspondence should be addressed to Qing-hui Wang; wangqh8008@vip.sina.com

Received 11 August 2013; Accepted 23 September 2013
Academic Editor: Tao Zou
Copyright © 2013 Qing-hui Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Regarding the tracking of moving target in the large-scale fixed scene, a new routing algorithm of LAODV in the principle of TOA localization is proposed. Then, the participation field of the fixed node based on the node location information is properly controlled, while the routing request area is reduced through combination of AODV and LAR during transmission of the location information. Simulation results show that the proposed algorithm renders satisfactory performance in terms of average delay reduction from end to end, packet loss rate, and routing overhead. As a result, the delay and system overhead during localization could be minimized.


## 1. Introduction

LBS service is developing along with rapid economic growth and great market potential. Also, wireless localization technology is one of the favorite research topics in recent years, where the position of the event or the node position is crucial for detection of sensor node [1, 2]. Normally, the wireless localization can be realized through satellite technology or ground technology [3]. Due to constraints of the node, such as the limited electric energy, huge quantity, and specific application environments, it is difficult to obtain the node coordinate with artificial measurement or allocation. Thus, implementation of expensive satellite wireless localization seems infeasible [4]. As for the ground wireless localization technology, the target localization is realized through multipoint coordination, the advantages of which include the distributed feature, low complexity, high precision, and good generality. The TOA-based (Time of Arrival) localization technology has been widely applied in the ground wireless localization [5-7]. Also, the moving node will send the location package to the neighboring node periodically, while the fixed node will respond back to the moving node once received. Therefore, there are alot of fixed nodes participating in the localization process, which will generate heavy communication overhead and reduced localization bandwidth.

There is a high demand for low-cost high-precision localization scheme in various fields, such as industrial field, safety production management, security, and training. In
the security monitoring, the current patrol system could be totally replaced, and the real-time position and route of the security personnel are clearly visible, so that prompt action of emergency can be taken. Moreover, the "virtual fence" is set up based on the geographic location of the visitors and the authorized condition to track the visitor. Then, entrance permission to particular region is conditionally issued. With increasing coalmine accidents, safety production situation is highly desired, as the administrative staff can monitor the location and situation of each miner remotely through the electronic tag wear by the miner or equipped on the delivery vehicles [8]. Furthermore, in the training of firemen, the training personnel in the building can be located and tracked through the allocated node in each floor. Such specific subject simulation can significantly improve the training quality and effectiveness.

In this paper, the TOA-based distance measurement method is proposed to localize the moving node. During the distance measurement and localization, limited nodes are chosen, so that the localization delay and system overhead could be reduced. During data transmission, the location information of the node is applied to restrict the routing request region through combination of AODV (Ad hoc OnDemand Distance Vector Routing) with LAR (LocationAided Routing). As a result, the transmit time of the RREQ (Route Request) package is reduced along with improved system transmission performance, which is verified by simulation study with NS2 emulator.


Figure 1: Round trip travel time.

## 2. TOA Localization Algorithm

TOA [9] refers to the time for the signal to travel from the sending node to the receiving node. Given signal transmission speed, by measuring the signal transmission time to calculate the distance between two nodes. In this paper, by measuring RTOF (Roundtrip-Time Of Flight) to calculate the distance between two nodes, as shown in Figure 1 the equation for the distance between two nodes is

$$
\begin{equation*}
d=\frac{\left[\left(T_{3}-T_{0}\right)+\left(T_{2}-T_{1}\right)\right] \times V}{2} \tag{1}
\end{equation*}
$$

where $T_{0}$ is the moment when the sending node sends signals, $T_{1}$ is the moment when the receiving node receives signals, $T_{2}$ is the moment when the receiving node sends response signals, $T_{3}$ is the moment when the sending node receives response signals, and $d$ is the distance between the sending node and receiving node.

In the large-scale fixed field, the moving node is required for localization, while the fixed node is the one with known information. The neighboring nodes are those within the radius of sensor node communication. The distance from the moving node to the fixed one can be obtained with TOA distance measurement method, as described earlier. When the distance from the moving node to at least three fixed nodes is known, the location of moving node can be calculated with trilateration [10].

## 3. AODV Protocol

AODV is the routing protocol based on the distance vector algorithm, which integrates the target serial number of DSDV and the on-demand routing discovery in DSR [11]. This protocol mainly includes routing discovery and routing maintenance, where the former is only requested upon need to save the routing overdue.

When the source node communicating with other nodes fails to reach the routing of destination node, it requires the grouping of RREQ. After other nodes receive this RREQ, whether such information exists or not is checked. Then, the


Figure 2: Node location distribution.
information should be abandoned when necessary. Otherwise, it should record the RREQ in this routing table and broadcast this RREQ continuously until some central node reaches the routing of destination node, or the routing request grouping reaches the destination node.

## 4. LAODV Algorithm

In this paper, localization of the moving node is desired, while the fixed one provides assistance. First, the moving node broadcasts a data package with location information of the moving nodes to the neighboring node periodically. Then, the distance between this node and the moving one should be calculated once the fixed nodes receive it within the broadcasting range, and decision of whether response information back to the moving nodes should be sent is made. In this algorithm, only four fixed nodes near to the moving node participate, so that the quantity of the involved nodes is limited. Therefore, the average delay and system overhead of the localization are minimized.

In AODV protocol, the node requests for the routing of the destination node through broadcasting RREQ message gradually. Such a flooding routing method will unfortunately generate substantial RREQ message, resulting in a tremendous signal conflict and protocol overhead. The proposed protocol will add the thought of LAR protocol [12], which will transfer the data package selectively during the process of routing discovery within the appointed routing area.

As shown in Figure 2, node S is the moving node, D is the gateway, others are fixed nodes, and their location coordinates are known. Each fixed node broadcasts its location information to the whole network, after other fixed nodes receive it, record information to their location information table, so that each fixed node has the position information of the other nodes. The moving node $S$ broadcasts a location package, including the location coordinate of $S$ calculated in the last moment. After the fixed nodes receive this package within the broadcasting scope, it can calculate the distance between the node and other fixed nodes with the moving node S . Through comparison, to judge whether the node is one of the four nodes closest to the moving node $S$, if it is, just send a response information, including the location coordinates of fixed nodes. In this case, only nodes A, B, P, and $M$ send response information.

Moving node S gets the location coordinates of $\mathrm{A}, \mathrm{B}$, P , and M which are $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right),\left(x_{3}, y_{3}\right)$, and $\left(x_{4}, y_{4}\right)$, respectively through the response information, and through the TOA-based distance measurement method can get the distances between them and the moving node S as $d_{1}, d_{2}, d_{3}$, and $d_{4}$, respectively. Given the coordinate of the moving node S by $(x, y)$, the following equation can calculate the location coordinate of the moving node:

$$
\begin{align*}
& \sqrt{\left(x-x_{1}\right)^{2}+\left(y-y_{1}\right)^{2}}=d_{1} \\
& \sqrt{\left(x-x_{2}\right)^{2}+\left(y-y_{2}\right)^{2}}=d_{2}  \tag{2}\\
& \sqrt{\left(x-x_{3}\right)^{2}+\left(y-y_{3}\right)^{2}}=d_{3} \\
& \sqrt{\left(x-x_{4}\right)^{2}+\left(y-y_{4}\right)^{2}}=d_{4} .
\end{align*}
$$

Meanwhile, the source node $S$ sends its location coordinate to the destination node D . As the routing information of $D$ is not reachable, node $S$ will send RREQ to seek for the routing to D with its location information included in RREQ. At first, a rectangular request region with two apexes of nodes S and D is determined. When node P out of the region receives the request package, decision of whether it is within the request region will be made, and the invalid grouping could be abandoned according to the location information. Then, node M within the region will transfer the grouping in the request region with its own location information included in the RREQ package. The new request region is established by the source node and destination node, until the node within the region reaches the radio-frequency scope of node D and the routing discovery is successful. Hence, the source node could send its own location coordinate information to the gateway according to this new route.

## 5. The Specific Process of Algorithm

(1) Each fixed nodes broadcasts a location package including its location information to the whole network. After other fixed node receive it, record information to their location information table.
(2) The moving node broadcasts a location package, including the location coordinate calculated in the last moment.
(3) After fixed nodes receive the package, calculate the distance between the node and other fixed nodes with the moving node.
(4) To determine whether the node is one of the four nodes closest to the moving node, if it is, just send a response information.
(5) Given four fixed node coordinates, and through the TOA-based distance measurement method can get the distances between them and the moving node, again through the trilateration it can calculate the moving node coordinate.


Figure 3: NS2 simulation flow chart.

## 6. Simulation and Analysis

6.1. Principle of NS2. In this study, the NS2 network emulator is adopted for performance verification [13]. NS2 simulation process in general is shown in Figure 3.
6.2. The Simulation Scene Set. As shown in Figure 4, a rectangular simulation field of $1000 \mathrm{~m} * 1000 \mathrm{~m}$ with 33 nodes is assumed, 26 of which are fixed nodes and the others are moving nodes. The distance of the adjacent fixed nodes is 150 m , and four nodes form the square grid. The highest movement speed of the node is $13 \mathrm{~m} / \mathrm{s}$, and the simulation time is set as 200 seconds. The CBR data flow will be established from every moving node to the gateway, and the data package of 512 bytes will be sent with a speed of 4 packet/s. More simulation parameters are illustrated in Table 1.

After simulation scene has been setted, writing TCL script simulation is carried out on the protocol. After the simulation will be get trace and nam files, the trace file is a data storage, nam file is the whole process of simulation of dynamic demonstration. In order to analyze a large amount of data of the trace file, we need to write the gawk program for the extraction and processing of effective data (calculate the


Figure 4: Simulation scene and node distribution.

Table 1: Parameter setting for simulation environment.

| Parameter | Parameter value |
| :--- | :---: |
| channelType | Channel/wireless channel |
| phyType | Phy/wirelessPhy |
| macType | Mac/802_11 |
| ifqType | Queue/drop tail/PriQueue |
| llType | LL |
| antType | Antenna/omni antenna |
| businessType | CBR |
| nodeQuantity | 33 |

average end-to-end delay, Package loss probability, etc.). In the end, using gnuplot drawing tools will be extracting data into two-dimensional graphics which can be more intuitive to analyze the protocol performance.
6.3. Analysis of Location Delay and System Overhead. In this paper, average localization delay refers to the average time for the moving node to send localization package until receiving response package, the system overhead refers to the proportion between the fixed nodes send response package number and the moving nodes send localization package number. Figures 5 and 6 show that the average localization delay of the algorithm is 0.00021 , the system overhead is 4.8 , while the AODV and system overhead are 0.00033 and 7.7 , respectively. Obviously, the new algorithm is advantageous in terms of localization delay and system overhead. This is mainly because the quantity of the fixed nodes participating


Figure 5: Average localization delay.


Figure 6: System overhead during location process.
in location is limited, and the relative distance from the fixed nodes to the moving nodes is short.
6.4. Performance Analysis of Data Transmission. In this paper, the performance index, including the average end-to-end delay of the network, the packet loss probability and routing overhead, is evaluated.
(1) Average End-to-End Delay. The average time for the moving node to send the data package until the gateway receives the data package successfully, which is related to smoothness of the network. The smaller the delay is, the smoother the network will be. The unit is second.

Figure 7 compares the average end-to-end delay of LAODV and AODV. Clearly, the delay of the LAODV


Figure 7: Average end-to-end delay.
protocol is smaller than that of the AODV protocol, as the location information of the node is added into the improved protocol. Through the calculation of the routing request area, it improves the seeking efficiency of the node for routing. When the flexible routing is interrupted due to the moving node, it will establish a new routing towards the destination node rapidly through the discovery mechanism of LAODV. Then, the seeking time for the route could be reduced, and the smoothness of the data transmission in the network is guaranteed.
(2) Package Loss Probability. The proportion of the total lost data grouping is against the sent data grouping in the network. Then, the successful data transmission proportion in the whole network can be known, as well as the proportion of the lost data package during the transmission process due to the link failure. This parameter is related to the efficiency of the data transmission.

Figure 8 shows comparison of the packet loss probability between the two protocols. It can be seen that the packet loss probability decreases gradually after the routes are established, and the packet loss probability of LAODV is smaller than that of AODV.
(3) Routing Overhead. The proportion between the total number of bytes and total message bytes of all the messages controlled by the route (including RREQ, RREPs and RERR messages). As for the grouping transmitted by various routes, a single bounce will trigger a message transmission. The routing overhead is mainly used for balancing the efficiency of the routing protocol. The smaller the routing overhead is, the narrower bandwidth is needed to reach the destination node.

Figure 9 shows the comparison of the routing overhead for two protocols. The routing overhead of LAODV is obviously smaller than that of AODV, as LAODV sets the routing request region utilizing the node location information. Then,


Figure 8: Packet loss probability.


Figure 9: Routing overhead.
the quantity of the transferred RREQ is smaller than that with AODV protocol, and the redundant routing request information is reduced. As a result, the routing overhead decreases accordingly.

## 7. Conclusion

In this paper, an approach for tracking the moving target and transmitting the location information in the large-scale fixed scene is developed. During the localization process, the quantity of fixed nodes participating in location is limited, so that the network overhead is reduced and the bandwidth of the system location by taking advantage of the location information of the located nodes is optimized. During the process
of uploading the location information of moving nodes to the gateway, the quantity of request package transferred during the routing discovery with the node location information is also properly controlled. The effectiveness of the algorithm has been verified by the simulation results.

## References

[1] C. Savarese, J. M. Rabaey, and J. Beutel, "Locationing in distributed ad-hoc wireless sensor networks," in Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing (ICASSP '01), vol. 4, pp. 2037-2040, IEEE, May 2001.
[2] F. Viani, P. Rocca, G. Oliveri, D. Trinchero, and A. Massa, "Localization, tracking, and imaging of targets in wireless sensor networks: an invited review," Radio Science, vol. 46, no. 5, Article ID RS5002, 2011.
[3] P. Keikhosrokiani, N. Mustaffa, N. Zakaria et al., "Wireless positioning techniques and location-based services: a literature review," in Multimedia and Ubiquitous Engineering, pp. 785-797, Springer, Amsterdam, The Netherlands, 2013.
[4] W. Chen, "Cooperative limiting localization schemes for wireless sensor networks," in Proceedings of the 1st International Conference on Innovative Computing, Information and Control (ICICIC '06), vol. 2, pp. 425-428, IEEE, September 2006.
[5] J. Shen, A. F. Molisch, and J. Salmi, "Accurate passive location estimation using TOA measurements," IEEE Transactions on Wireless Communications, vol. 11, no. 6, pp. 2182-2192, 2012.
[6] J. He, S. Li, K. Pahlavan et al., "A realtime testbed for performance evaluation of indoor TOA location system," in Proceedings of the IEEE International Conference on Communications (ICC '12), pp. 482-486, IEEE, 2012.
[7] H. Jhi, J. Chen, C. Lin, and C. Huang, "A factor-graph-based TOA location estimator," IEEE Transactions on Wireless Communications, vol. 11, no. 5, pp. 1764-1773, 2012.
[8] S. Jiping, "Personnel position monitoring technology and system in underground mine," Coal Science and Technology, vol. 38, no. 11, pp. 1-4, 2010.
[9] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," Wireless Networks, vol. 8, no. 2-3, pp. 187-197, 2002.
[10] F. Izquierdo, M. Ciurana, F. Barceló, J. Paradells, and E. Zola, "Performance evaluation of a TOA-based trilateration method to locate terminals in WLAN," in Proceedings of the 2006 1st International Symposium on Wireless Pervasive Computing, pp. 1-6, IEEE, January 2006.
[11] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," Tech. Rep. RFC 3561, Internet RFCs, 2003.
[12] Y. Ko and N. H. Vaidya, "Location-Aided Routing (LAR) in mobile ad hoc networks," Wireless Networks, vol. 6, no. 4, pp. 307-321, 2000.
[13] A. Tuteja, R. Gujral, and S. Thalia, "Comparative performance analysis of DSDV, AODV and DSR routing protocols in MANET using NS2," in Proceedings of the International Conference on Advances in Computer Engineering (ACE '10), pp. 330333, IEEE, Bangalore, India, June 2010.

# Adaptive Semidiscrete Finite Element Methods for Semilinear Parabolic Integrodifferential Optimal Control Problem with Control Constraint 

Zuliang Lu ${ }^{1,2}$<br>${ }^{1}$ School of Mathematics and Statistics, Chongqing Three Gorges University, Chongqing 404000, China<br>${ }^{2}$ College of Civil Engineering and Mechanics, Xiangtan University, Xiangtan 411105, China<br>Correspondence should be addressed to Zuliang Lu; zulianglux@126.com

Received 27 April 2013; Revised 5 August 2013; Accepted 27 August 2013
Academic Editor: Dewei Li
Copyright © 2013 Zuliang Lu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The aim of this work is to study the semidiscrete finite element discretization for a class of semilinear parabolic integrodifferential optimal control problems. We derive a posteriori error estimates in $L^{2}\left(J ; L^{2}(\Omega)\right)$-norm and $L^{2}\left(J ; H^{1}(\Omega)\right)$-norm for both the control and coupled state approximations. Such estimates can be used to construct reliable adaptive finite element approximation for semilinear parabolic integrodifferential optimal control problem. Furthermore, we introduce an adaptive algorithm to guide the mesh refinement. Finally, a numerical example is given to demonstrate the theoretical results.


## 1. Introduction

With the advances of scientific computing, optimal control problems are now widely used in multidisciplinary applications such as physical, biological, medicine, engineering design, finance, fluid mechanics, and socioeconomic systems. As a result, more and more people will benefit greatly by learning to solve the optimal control problems numerically. Realizing such growing needs, books and papers on optimal control put more weight on numerical methods.

In modeling a wide range of problems for physical, economic, and social processes, optimal control problems described by integrodifferential equations play an important role. Parabolic integrodifferential optimal control problems are very important model in engineering numerical simulation, for example, biology mechanics, nuclear reaction dynamics, heat conduction in materials with memory, viscoelasticity, and so forth. Finite element approximation of optimal control problems has a very important status in the numerical methods for these problems. The finite element approximation of optimal control problem by piecewise constant functions was well investigated by Falk [1] and Geveci [2]. The finite element methods for semilinear elliptic optimal control problems were discussed by Casas et al. in [3]. In [4],
the author studied the finite element discretization for a class of quadratic boundary optimal control problems governed by nonlinear elliptic equations and obtained a posteriori error estimates for the coupled state and control approximation. Many introductions about the numerical computation method for optimal control problems can be found in [5-8].

As one of important kinds of optimal control problems, parabolic integrodifferential optimal control problem is widely used in scientific and engineering computing. The literature in this aspect was huge, see; for example, [9]. In [10], Brunner and Yan analyzed finite element Galerkin discretization for a class of optimal control problems governed by integral equations and integrodifferential equations and derived a priori error estimates and a posteriori error estimators for the approximation solutions.

Adaptive finite element method is the most important method to boost accuracy of the finite element discretization. It ensures a higher density of nodes in certain area of the given domain, where the solution is discontinuous or more difficult to approximate, using an a posteriori error indicator. A posteriori error estimates are computable quantities in terms of the discrete solution and they measure the actual discrete errors without the knowledge of exact solutions. They are essential in designing algorithms for mesh which
equidistribute the computational effort and optimize the computation. The literature in this was huge. In [11], the authors presented an a posteriori error analysis for finite element approximation of distributed convex elliptic optimal control problems. They derived a posteriori error estimates for the coupled state and control approximations. Such estimates can be used to construct reliable adaptive finite element approximation schemes for control problems. In [12], Verfürth gave a general framework for deriving a posteriori error estimates for approximate solutions of nonlinear elliptic equations. He obtained a posteriori error estimates, which can easily be computed from the given data of the problem and the computed numerical solution and which give global upper and local lower bounds on the error of the numerical solution. Some of techniques directly relevant to our work can be found in [11, 12]. Recently, in [13-16], we derived a priori error estimates and a posteriori error estimates for optimal control problems using mixed finite element methods. Although a posteriori error estimates of finite element approximation were widely used in numerical simulations, they have not yet been utilized in nonlinear parabolic integrodifferential optimal control problem.

In this paper, we adopt the standard notation $W^{m, p}(\Omega)$ for Sobolev spaces on $\Omega$ with a norm $\|\cdot\|_{m, p}$ given by $\|v\|_{m, p}^{p}=\sum_{|\alpha| \leq m}\left\|D^{\alpha} v\right\|_{L^{p}(\Omega)}^{p}$ and a seminorm $|\cdot|_{m, p}$ given by $|v|_{m, p}^{p}=\sum_{|\alpha|=m}\left\|D^{\alpha} v\right\|_{L^{p}(\Omega)}^{p}$. We set $W_{0}^{m, p}(\Omega)=\{v \in$ $\left.W^{m, p}(\Omega):\left.v\right|_{\partial \Omega}=0\right\}$. For $p=2$, we denote $H^{m}(\Omega)=$ $W^{m, 2}(\Omega), H_{0}^{m}(\Omega)=W_{0}^{m, 2}(\Omega)$, and $\|\cdot\|_{m}=\|\cdot\|_{m, 2},\|\cdot\|=$ $\|\cdot\|_{0,2}$. We denote by $L^{s}\left(0, T ; W^{m, p}(\Omega)\right)$ the Banach space of all $L^{s}$ integrable functions from $J$ to $W^{m, p}(\Omega)$ with norm $\|v\|_{L^{s}\left(J ; W^{m, p}(\Omega)\right)}=\left(\int_{0}^{T}\|v\|_{W^{m, p}(\Omega)}^{s} d t\right)^{1 / s}$ for $s \in[1, \infty)$, and the standard modification for $s=\infty$. The details can be found in [9].

In this paper, we derive a posteriori error estimates for a class of semilinear parabolic integrodifferential optimal control problems. To the best of our knowledge in the context of semilinear parabolic integrodifferential optimal control problems, these estimates are new. We consider the following semilinear parabolic integrodifferential optimal control problems:

$$
\begin{equation*}
\min _{u(t) \in K \subset U}\left\{\int_{0}^{T}\left(\frac{1}{2}\left\|y-y_{0}\right\|^{2}+\frac{\alpha}{2}\|u\|^{2}\right) d t\right\} \tag{1}
\end{equation*}
$$

subject to the state equation

$$
\begin{align*}
& y_{t}= \operatorname{div}(A \nabla y(x, t)) \\
&-\int_{0}^{t} \operatorname{div}(\psi(t, \tau) \nabla y(x, \tau)) d \tau+\phi(y) \\
&=f+B u, \quad x \in \Omega, t \in J  \tag{2}\\
& y(x, t)=0, \quad x \in \partial \Omega, t \in J \\
& y(x, 0)=y_{0}(x), \quad x \in \Omega
\end{align*}
$$

where the bounded open set $\Omega \subset \mathbb{R}^{2}$ is 2 regular convex polygon with boundary $\partial \Omega, J=(0, T], f \in L^{2}(\Omega), \psi=$
$\psi(x, t, \tau)=\psi_{i, j}(x, t, \tau)_{2 \times 2} \in C^{\infty}\left(0, T ; L^{2}(\bar{\Omega})\right)^{2 \times 2}, y_{0} \in H^{1}(\Omega)$, $\alpha$ is a positive constant, and $B$ is a continuous linear operator from $K$ to $L^{2}(\Omega)$. For any $R>0$, the function $\phi(\cdot) \in$ $W^{2, \infty}(-R, R), \phi^{\prime}(y) \in L^{2}(\Omega)$ for any $y \in L^{2}\left(J ; H_{0}^{1}(\Omega)\right)$, and $\phi^{\prime}(y) \geq 0$. We assume that the coefficient matrix $A(x)=$ $\left(a_{i, j}(x)\right)_{2 \times 2} \in\left(W^{1, \infty}(\Omega)\right)^{2 \times 2}$ is a symmetric positive definite matrix and there is a constant $c>0$ satisfying for any vector $\mathbf{X} \in \mathbb{R}^{2}, \mathbf{X}^{t} A \mathbf{X} \geq c\|\mathbf{X}\|_{\mathbb{R}^{2}}^{2}$. Here, $K$ denotes the admissible set of the control variable defined by

$$
\begin{equation*}
K=\left\{u(x, t) \in U=L^{2}\left(J ; L^{2}(\Omega)\right): u(x, t) \geq 0, t \in J\right\} . \tag{3}
\end{equation*}
$$

Assume that there are constants $c$ and $C$, such that for all $t$ and $\tau$ in $[0, T]$ :

$$
\begin{gather*}
a(z, z) \geq c\|z\|_{1, \Omega}^{2}, \quad \forall z \in V,  \tag{4}\\
|a(z, w)| \leq C\|z\|_{1, \Omega} \cdot\|w\|_{1, \Omega}, \quad \forall z, w \in V,  \tag{5}\\
|\psi(t, \tau ; z, w)| \leq C\|z\|_{1, \Omega} \cdot\|w\|_{1, \Omega}, \quad \forall z, w \in V . \tag{6}
\end{gather*}
$$

The plan of this paper is as follows. In the next section, we present the finite element discretization for semilinear parabolic integrodifferential optimal control problems. A posteriori error estimates are established for the optimal control problems in Section 3. In Section 4, we introduce an adaptive algorithm to guide the mesh refinement. In Section 5, a numerical example is given to demonstrate our theoretical results. Finally, we analyze the conclusion and future work in Section 6.

## 2. Finite Elements for Integrodifferential Optimal Control

We will now describe the finite element discretization of semilinear parabolic integrodifferential optimal control problems (1)-(2). Let $V=H_{0}^{1}(\Omega)$ and $W=L^{2}(\Omega)$. Let

$$
\begin{align*}
a(y, w) & =\int_{\Omega}(A \nabla y) \cdot \nabla w, \quad \forall y, w \in V, \\
\psi(t, \tau ; z, w) & =(\psi(t, \tau) \nabla z, \nabla w), \quad \forall z, w \in V \times V, \\
(u, v) & =\int_{\Omega} u v, \quad \forall(u, v) \in W \times W,  \tag{7}\\
\left(f_{1}, f_{2}\right) & =\int_{\Omega} f_{1} f_{2}, \quad \forall\left(f_{1}, f_{2}\right) \in W \times W
\end{align*}
$$

Then, the semilinear parabolic integrodifferential optimal control problems (1)-(2) can be restated as

$$
\begin{equation*}
\min _{u(t) \in K}\left\{\int_{0}^{T}\left(\frac{1}{2}\left\|y-y_{0}\right\|^{2}+\frac{\alpha}{2}\|u\|^{2}\right) d t\right\} \tag{8}
\end{equation*}
$$

subject to

$$
\begin{gather*}
\left(y_{t}, w\right)+a(y, w)+\int_{0}^{t} \psi(t, \tau ; y(\tau), w) d \tau+(\phi(y), w) \\
=(f+B u, w), \quad \forall w \in V, t \in J \\
y(x, 0)=y_{0}(x), \quad x \in \Omega \tag{9}
\end{gather*}
$$

where the inner product in $L^{2}(\Omega)$ or $L^{2}(\Omega)^{2}$ is indicated by $(\cdot, \cdot)$. From Yanik and Fairweather [17], we know that the above weak form has at least one solution in $y \in W(0, T)=$ $\left\{w \in L^{2}\left(0, T ; H^{1}(\Omega)\right), w_{t}^{\prime} \in L^{2}\left(0, T ; H^{-1}(\Omega)\right)\right\}$.

It is well known (see, e.g., [11]) that the optimal control problem has a solution $(y, u)$, and if a pair $(y, u)$ is the solution of (8)-(9), then there is a costate $p \in V$ such that triplet $(y, p, u)$ satisfies the following optimality conditions:

$$
\begin{gather*}
\left(y_{t}, w\right)+a(y, w) \\
+\int_{0}^{t} \psi(t, \tau ; y(\tau), w) d \tau+(\phi(y), w)  \tag{10}\\
=(f+B u, w), \quad \forall w \in V, \\
y(x, 0)=y_{0}(x), \quad x \in \Omega,  \tag{11}\\
-\left(p_{t}, w\right)+a(q, p)+\int_{t}^{T} \psi(\tau, t ; q, p(\tau)) d \tau+\left(\phi^{\prime}(y) p, q\right) \\
=\left(y-y_{0}, q\right), \quad \forall q \in V,  \tag{12}\\
p(x, T)=0, \quad x \in \Omega,  \tag{13}\\
\int_{0}^{T}\left(\alpha u+B^{*} p, v-u\right)_{U} d t \geq 0, \quad \forall v \in K, \tag{14}
\end{gather*}
$$

where $B^{*}$ is the adjoint operator of $B$. In the rest of the paper, we will simply write the product as $(\cdot, \cdot)$ whenever no confusion should be caused.

Let us consider the finite element approximation of the optimal control problem (8)-(9). Again here we consider only $n$-simplex elements and conforming finite elements.

Let $\mathscr{T}^{h}$ be regular partition of $\Omega$. Associated with $\mathscr{T}^{h}$ is a finite dimensional subspace $V_{h}$ of $C(\bar{\Omega})$, such that $\left.\chi\right|_{\tau}$ are polynomials of $m$-order $(m \geq 1)$ for all $\chi \in V_{h}$ and $\tau \in \mathscr{T}^{h}$. It is easy to see that $V_{h} \subset V$. Let $h_{\tau}$ denote the maximum diameter of the element $\tau$ in $\mathscr{T}^{h}, h=\max _{\tau \in \mathscr{T}^{h}}\left\{h_{\tau}\right\}$.

Due to the limited regularity of the optimal control $u$ in general, there will be no advantage in considering higherorder finite element spaces for the control. So, we only consider the piecewise constant finite element space for the approximation of the control, though higher-order finite element spaces will be used to approximate the state and the costate. Let $P_{0}(\tau)$ denote all the 0 -order polynomial over $\tau$. Therefore, we take $K_{h}=\left\{u \in K:\left.u(x, t)\right|_{\tau} \in\right.$ $\left.P_{0}(\tau)\right\}$. In addition, $C$ or $c$ denotes a general positive constant independent of $h$.

By the definition of finite element subspace, the finite element discretization of (8)-(9) is as follows: compute $\left(y_{h}, u_{h}\right) \in V_{h} \times K_{h}$ such that

$$
\begin{gather*}
\min _{u_{h} \in K_{h}}\left\{\int_{0}^{T}\left(\frac{1}{2}\left\|y_{h}-y_{0}\right\|^{2}+\frac{\alpha}{2}\left\|u_{h}\right\|^{2}\right) d t\right\}  \tag{15}\\
\left(y_{h t}, w_{h}\right)+a\left(y_{h}, w_{h}\right)+\int_{0}^{t} \psi\left(t, \tau ; y_{h}(\tau), w_{h}\right) d \tau  \tag{16}\\
+\left(\phi\left(y_{h}\right), w_{h}\right)=\left(f+B u_{h}, w_{h}\right), \\
y_{h}(x, 0)=y_{0}^{h}(x), \quad x \in \Omega, \tag{17}
\end{gather*}
$$

where $w_{h} \in V_{h}, y_{0}^{h} \in V_{h}$ is an approximation of $y_{0}$.
Again, it follows that the optimal control problems (15)(17) have a solution $\left(y_{h}, u_{h}\right)$, and if a pair $\left(y_{h}, u_{h}\right)$ is the solution of (15)-(17), then there is a costate $p_{h} \in V_{h}$ such that triplet ( $y_{h}, p_{h}, u_{h}$ ) satisfies the following optimality conditions:

$$
\begin{gather*}
\left(y_{h t}, w_{h}\right)+a\left(y_{h}, w_{h}\right)+\int_{0}^{t} \psi\left(t, \tau ; y_{h}(\tau), w_{h}\right) d \tau  \tag{18}\\
+\left(\phi\left(y_{h}\right), w_{h}\right)=\left(f+B u_{h}, w_{h}\right) \\
y_{h}(x, 0)=y_{0}^{h}(x), \quad x \in \Omega  \tag{19}\\
-\left(p_{h t}, w_{h}\right)+a\left(q_{h}, p_{h}\right)+\int_{t}^{T} \psi\left(\tau, t ; q_{h}, p_{h}(\tau)\right) d \tau  \tag{20}\\
+\left(\phi^{\prime}\left(y_{h}\right) p_{h}, q_{h}\right)=\left(y_{h}-y_{0}, q_{h}\right) \\
p_{h}(x, T)=0, \quad x \in \Omega  \tag{21}\\
\int_{0}^{T}\left(\alpha u_{h}+B^{*} p_{h}, v_{h}-u_{h}\right) d t \geq 0 \tag{22}
\end{gather*}
$$

where $w_{h}, q_{h} \in V_{h}$, and $v_{h} \in K_{h}$.
In the rest of the paper, we will use some intermediate variables. For any control function $u_{h} \in K$, we first define the state solution $\left(y\left(u_{h}\right), p\left(u_{h}\right)\right)$ satisfying

$$
\begin{gather*}
\left(y_{t}\left(u_{h}\right), w\right)+a\left(y\left(u_{h}\right), w\right) \\
+\int_{0}^{t} \psi\left(t, \tau ; y\left(u_{h}\right)(\tau), w\right) d \tau  \tag{23}\\
+\left(\phi\left(y\left(u_{h}\right)\right), w\right)=\left(f+B u_{h}, w\right),  \tag{24}\\
\forall w \in V, \quad y\left(u_{h}\right)(x, 0)=y_{0}(x), \quad x \in \Omega, \\
-\left(p_{t}\left(u_{h}\right), q\right)+a\left(q, p\left(u_{h}\right)\right)+\int_{t}^{T} \psi\left(\tau, t ; q, p\left(u_{h}\right)(\tau)\right) d \tau  \tag{25}\\
+\left(\phi^{\prime}\left(y\left(u_{h}\right)\right) p\left(u_{h}\right), q\right)
\end{gather*}
$$

$$
\begin{equation*}
=\left(y\left(u_{h}\right)-y_{0}, q\right), \quad \forall q \in V, \quad p\left(u_{h}\right)(x, T)=0, \quad x \in \Omega . \tag{26}
\end{equation*}
$$

Now, we restate the following well-known estimates in [9].

Lemma 1. Let $\hat{\pi}_{h}$ be the Clément-type interpolation operator defined in [9]. Then for any $v \in H^{1}(\Omega)$ and all element $\tau$,

$$
\begin{align*}
& \left\|v-\hat{\pi}_{h} v\right\|_{L^{2}(\tau)}+h_{\tau}\left\|\nabla\left(v-\hat{\pi}_{h} v\right)\right\|_{L^{2}(\tau)} \\
& \leq C h_{\tau} \sum_{\bar{\tau}^{\prime} \cap \bar{\tau} \neq \emptyset}|v|_{L^{2}\left(\tau^{\prime}\right)},  \tag{27}\\
& \left\|v-\hat{\pi}_{h} v\right\|_{L^{2}(l)} \leq C h_{l}^{1 / 2} \sum_{l \subset \bar{\tau}^{\prime}}|\nabla v|_{L^{2}\left(\tau^{\prime}\right)},
\end{align*}
$$

where $l$ is the edge of the element.
For $\varphi \in W_{h}$, we will write

$$
\begin{align*}
\phi(\varphi)-\phi(\rho) & =-\widetilde{\phi}^{\prime}(\varphi)(\rho-\varphi)  \tag{28}\\
& =-\phi^{\prime}(\rho)(\rho-\varphi)+\widetilde{\phi}^{\prime \prime}(\varphi)(\rho-\varphi)^{2}
\end{align*}
$$

where

$$
\begin{gather*}
\widetilde{\phi}^{\prime}(\varphi)=\int_{0}^{1} \phi^{\prime}(\varphi+s(\rho-\varphi)) d s \\
\widetilde{\phi}^{\prime \prime}(\varphi)=\int_{0}^{1}(1-s) \phi^{\prime \prime}(\rho+s(\varphi-\rho)) d s \tag{29}
\end{gather*}
$$

are bounded functions in $\bar{\Omega}[12]$.

## 3. A Posteriori Error Estimates

In this section, we will obtain a posteriori error estimates for semilinear parabolic integrodifferential optimal control problems. Firstly, we estimate the error $\left\|u-u_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}$.

For given $u \in K$, let $M$ be the inverse operator of the state equation (10), such that $y(u)=M B u$ is the solution of the state equation (10). Similarly, for given $u_{h} \in K_{h}, y_{h}\left(u_{h}\right)=$ $M_{h} B u_{h}$ is the solution of the discrete state equation (16). Let

$$
\begin{align*}
S(u) & =\frac{1}{2}\left\|M B u-y_{0}\right\|^{2}+\frac{\alpha}{2}\|u\|^{2}, \\
S_{h}\left(u_{h}\right) & =\frac{1}{2}\left\|M_{h} B u_{h}-y_{0}\right\|^{2}+\frac{\alpha}{2}\left\|u_{h}\right\|^{2} . \tag{30}
\end{align*}
$$

It is clear that $S$ and $S_{h}$ are well defined and continuous on $K$ and $K_{h}$. Also the functional $S_{h}$ can be naturally extended on $K$. Then (8) and (15) can be represented as

$$
\begin{gather*}
\min _{u \in K}\{S(u)\}  \tag{31}\\
\min _{u_{h} \in K_{h}}\left\{S_{h}\left(u_{h}\right)\right\} . \tag{32}
\end{gather*}
$$

It can be shown that

$$
\begin{gather*}
\left(S^{\prime}(u), v\right)=\left(\alpha u+B^{*} p, v\right), \\
\left(S^{\prime}\left(u_{h}\right), v\right)=\left(\alpha u_{h}+B^{*} p\left(u_{h}\right), v\right),  \tag{33}\\
\left(S_{h}^{\prime}\left(u_{h}\right), v\right)=\left(\alpha u_{h}+B^{*} p_{h}, v\right),
\end{gather*}
$$

where $p\left(u_{h}\right)$ is the solution of (23)-(25).

In many applications, $S(\cdot)$ is uniformly convex near the solution $u$ (see, e.g., [18]). The convexity of $S(\cdot)$ is closely related to the second-order sufficient conditions of the control problems, which was assumed in many studies on numerical methods of the problems. If $S(\cdot)$ is uniformly convex, then there is a $c>0$, such that

$$
\begin{equation*}
\int_{0}^{T}\left(S^{\prime}(u)-S^{\prime}\left(u_{h}\right), u-u_{h}\right) d t \geq c\left\|u-u_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}^{2} \tag{34}
\end{equation*}
$$

where $u$ and $u_{h}$ are the solutions of (31) and (32), respectively. We will assume the above inequality throughout this paper.

Let $p\left(u_{h}\right)$ be the solution of (23)-(25); we establish the following error estimate.

Theorem 2. Let $u$ and $u_{h}$ be the solutions of (31) and (32), respectively. Assume that $K_{h} \subset K$. In addition, assume that $\left.\left(S_{h}^{\prime}\left(u_{h}\right)\right)\right|_{\tau} \in H^{s}(\tau)$, for all $\tau \in \mathscr{T}_{h},(s=0,1)$, and there is a $v_{h} \in K_{h}$ such that

$$
\begin{equation*}
\left|\left(S_{h}^{\prime}\left(u_{h}\right), v_{h}-u\right)\right| \leq C \sum_{\tau \in \mathscr{T}_{h}} h_{\tau}\left\|S_{h}^{\prime}\left(u_{h}\right)\right\|_{H^{s}(\tau)}\left\|u-u_{h}\right\|_{L^{2}(\tau)}^{s} \tag{35}
\end{equation*}
$$

Then, one has

$$
\begin{equation*}
\left\|u-u_{h}\right\|_{L^{2}\left(j ; L^{2}(\Omega)\right)}^{2} \leq C \eta_{1}^{2}+C\left\|p_{h}-p\left(u_{h}\right)\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \tag{36}
\end{equation*}
$$

where

$$
\begin{equation*}
\eta_{1}^{2}=\int_{0}^{T} \sum_{\tau \in \mathscr{T}_{h}} h_{\tau}^{1+s}\left\|\alpha u_{h}+B^{*} p_{h}\right\|_{H^{s}(\tau)}^{1+s} d t \tag{37}
\end{equation*}
$$

Proof. It follows from (31) and (32) that

$$
\begin{gather*}
\int_{0}^{T}\left(S^{\prime}(u), u-v\right) \leq 0, \quad \forall v \in K  \tag{38}\\
\int_{0}^{T}\left(S_{h}^{\prime}\left(u_{h}\right), u_{h}-v_{h}\right) \leq 0, \quad \forall v_{h} \in K_{h} \subset K . \tag{39}
\end{gather*}
$$

Then it follows from (35), (39), and the Schwartz inequality, that

$$
\begin{aligned}
& c\left\|u-u_{h}\right\|_{L^{2}\left(j ; L^{2}(\Omega)\right)}^{2} \\
& \leq \int_{0}^{T}\left(S^{\prime}(u)-S^{\prime}\left(u_{h}\right), u-u_{h}\right) d t \\
& \leq-\int_{0}^{T}\left(S^{\prime}\left(u_{h}\right), u-u_{h}\right) d t \\
& = \\
& \quad \int_{0}^{T}\left\{\left(S_{h}^{\prime}\left(u_{h}\right), u_{h}-u\right)\right. \\
& \left.\quad+\left(S_{h}^{\prime}\left(u_{h}\right)-S^{\prime}\left(u_{h}\right), u-u_{h}\right)\right\} d t
\end{aligned}
$$

$$
\begin{align*}
\leq & \int_{0}^{T}\left\{\left(S_{h}^{\prime}\left(u_{h}\right), v_{h}-u\right)\right. \\
& \left.+\left(S_{h}^{\prime}\left(u_{h}\right)-S^{\prime}\left(u_{h}\right), u-u_{h}\right)\right\} d t \\
\leq & C \int_{0}^{T}\left\{\sum_{\tau \in \mathscr{T}_{h}} h_{\tau}^{1+s}\left\|S_{h}^{\prime}\left(u_{h}\right)\right\|_{H^{s}(\tau)}^{1+s}\right. \\
& \left.+\left\|S_{h}^{\prime}\left(u_{h}\right)-S^{\prime}\left(u_{h}\right)\right\|_{L^{2}(\Omega)}^{2}\right\} d t \\
& +\frac{c}{2}\left\|u-u_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}^{2} \tag{40}
\end{align*}
$$

It is not difficult to show that

$$
\begin{equation*}
S_{h}^{\prime}\left(u_{h}\right)=\alpha u_{h}+B^{*} p_{h}, \quad S^{\prime}\left(u_{h}\right)=\alpha u_{h}+B^{*} p\left(u_{h}\right) \tag{41}
\end{equation*}
$$

where $p\left(u_{h}\right)$ is defined in (23)-(26). Thanks to (11), it is easy to derive

$$
\begin{align*}
\int_{0}^{T} & \left\|S_{h}^{\prime}\left(u_{h}\right)-S^{\prime}\left(u_{h}\right)\right\|_{L^{2}(\Omega)} d t \\
& =\int_{0}^{T}\left\|B^{*}\left(p_{h}-p\left(u_{h}\right)\right)\right\|_{L^{2}(\Omega)} d t  \tag{42}\\
& \leq C\left\|p_{h}-p\left(u_{h}\right)\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)} \\
& \leq C\left\|p_{h}-p\left(u_{h}\right)\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)} .
\end{align*}
$$

Then, by the estimates (40) and (42), we can prove the requested result (36).

Now, we estimate the error $\left\|y\left(u_{h}\right)-y_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}$.
Theorem 3. Let $\left(y\left(u_{h}\right), p\left(u_{h}\right)\right)$ and $\left(y_{h}, p_{h}\right)$ be the solutions of (23)-(26) and (18)-(22), respectively. Then

$$
\begin{equation*}
\left\|y\left(u_{h}\right)-y_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \leq C \sum_{i=2}^{4} \eta_{i}^{2} \tag{43}
\end{equation*}
$$

where

$$
\begin{aligned}
& \eta_{2}^{2}=\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{\tau}^{2} \int_{\tau}\left(f+B u_{h}-y_{h t}+\operatorname{div}\left(A \nabla y_{h}\right)\right. \\
&+\int_{0}^{t} \operatorname{div}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) d \tau \\
&\left.-\phi\left(y_{h}\right)\right)^{2} d t
\end{aligned}
$$

$$
\begin{align*}
\eta_{3}^{2}=\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{l} \int_{\partial \tau} & {\left[\left(A \nabla y_{h}\right) \cdot n\right.} \\
& \left.+\int_{0}^{t}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) \cdot n d \tau\right]^{2} d l d t \\
\eta_{4}^{2}=\| & y_{h}(x, 0)-y_{0}(x) \|_{L^{2}(\Omega)}^{2} \tag{44}
\end{align*}
$$

where $l$ is a face of an element $\tau, h_{l}$ is the size of face $l$, and $\left[\left(A \nabla y_{h}\right) \cdot n\right]$ is the $A$-normal derivative jump over the interior face $l$, defined by

$$
\begin{equation*}
\left[\left(A \nabla y_{h}\right) \cdot n\right]_{l}=\left(\left.A \nabla y_{h}\right|_{\tau_{l}^{1}}-\left.A \nabla y_{h}\right|_{\tau_{l}^{2}}\right) \cdot n \tag{45}
\end{equation*}
$$

where $n$ is the unit normal vector on $l=\bar{\tau}_{l}^{1} \cap \bar{\tau}_{l}^{2}$ outwards $\tau_{l}^{1}$. For convenience, one defines $\left[\left(A \nabla y_{h}\right) \cdot n\right]_{l}=0$ when $l \subset \partial \Omega$.

Proof. Let $e_{I}^{y}$ be the Clément-type interpolator of $e^{y}$ defined in Lemma 1. It follows from (18) and (23) that we have

$$
\begin{align*}
& \left(\left(y_{h}-y\left(u_{h}\right)\right)_{t}, w_{h}\right)+a\left(y_{h}-y\left(u_{h}\right), w_{h}\right) \\
& \quad+\int_{0}^{t} \psi\left(t, \tau ;\left(y_{h}-y\left(u_{h}\right)\right)(\tau), w_{h}\right) d \tau  \tag{46}\\
& \quad+\left(\phi\left(y_{h}\right)-\phi\left(y\left(u_{h}\right)\right), w_{h}\right)=0, \quad \forall w_{h} \in V_{h} .
\end{align*}
$$

Let $e^{y}=y_{h}-y\left(u_{h}\right)$; by using (46), then we obtain

$$
\begin{aligned}
&\left(\left(y_{h}-y\left(u_{h}\right)\right)_{t}, e^{y}\right)+a\left(y_{h}-y\left(u_{h}\right), e^{y}\right) \\
& \quad+\int_{0}^{t} \psi\left(t, \tau ;\left(y_{h}-y\left(u_{h}\right)\right)(\tau), e^{y}\right) d \tau \\
&+\left(\phi\left(y_{h}\right)-\phi\left(y\left(u_{h}\right)\right), e^{y}\right) \\
&=\left(\left(y_{h}-y\left(u_{h}\right)\right)_{t}, e^{y}-e_{I}^{y}\right)+a\left(y_{h}-y\left(u_{h}\right), e^{y}-e_{I}^{y}\right) \\
&+\int_{0}^{t} \psi\left(t, \tau ;\left(y_{h}-y\left(u_{h}\right)\right)(\tau), e^{y}-e_{I}^{y}\right) d \tau \\
&+\left(\phi\left(y_{h}\right)-\phi\left(y\left(u_{h}\right)\right), e^{y}-e_{I}^{y}\right) \\
&=\left(y_{h t}, e^{y}-e_{I}^{y}\right)+a\left(y_{h}, e^{y}-e_{I}^{y}\right) \\
&+\int_{0}^{t} \psi\left(t, \tau ; y_{h}(\tau), e^{y}-e_{I}^{y}\right) d \tau \\
&+\left(\phi\left(y_{h}\right), e^{y}-e_{I}^{y}\right)-\left(f+B u_{h}, e^{y}-e_{I}^{y}\right) \\
&= \sum_{\tau} \int_{\tau}\left(y_{h t}-\operatorname{div}\left(A \nabla y_{h}\right)\right. \\
& \quad-\int_{0}^{t} \operatorname{div}\left(\psi(t, \tau) \nabla y_{h}\right) d \tau \\
&\left.\quad+\phi\left(y_{h}\right)-f-B u_{h}, e^{y}-e_{I}^{y}\right)
\end{aligned}
$$

$$
\begin{align*}
+\sum_{\tau} \int_{\partial \tau} & {\left[\left(A \nabla y_{h}\right) \cdot n\right.} \\
& \left.\left.+\int_{0}^{t}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) \cdot n d \tau\right]\left(e^{y}-e_{I}^{y}\right)\right] \tag{47}
\end{align*}
$$

Then, we have

$$
\begin{align*}
& \frac{1}{2} \frac{d}{d t}\left\|y_{h}-y\left(u_{h}\right)\right\|_{0, \Omega}^{2}+c\left\|y_{h}-y\left(u_{h}\right)\right\|_{1, \Omega}^{2} \\
& \leq\left(\left(y_{h}-y\left(u_{h}\right)\right)_{t}, e^{y}\right)+a\left(y_{h}-y\left(u_{h}\right), e^{y}\right) \\
& \quad+\left(\phi\left(y_{h}\right)-\phi\left(y\left(u_{h}\right)\right), e^{y}\right) \\
& \leq \sum_{\tau} \int_{\tau}\left(y_{h t}-\operatorname{div}\left(A \nabla y_{h}\right)-\int_{0}^{t} \operatorname{div}\left(\psi(t, \tau) \nabla y_{h}\right) d \tau\right. \\
& \left.\quad+\phi\left(y_{h}\right)-f-B u_{h}, e^{y}-e_{I}^{y}\right) \\
& \quad+\sum_{\tau} \int_{\partial \tau}\left[\left(A \nabla y_{h}\right) \cdot n+\int_{0}^{t}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) \cdot n d \tau\right] \\
& \quad \times\left(e^{y}-e_{I}^{y}\right) \\
& \quad-\int_{0}^{t} \psi\left(t, \tau ;\left(y_{h}-y\left(u_{h}\right)\right)(\tau), e^{y}\right) d \tau . \tag{48}
\end{align*}
$$

By integrating time from 0 to $t$ in the above inequality, combining (6) and the Schwartz inequality, we have

$$
\begin{align*}
& \frac{1}{2}\left\|y_{h}-y\left(u_{h}\right)\right\|_{0, \Omega}^{2}+c \int_{0}^{t}\left\|y_{h}-y\left(u_{h}\right)\right\|_{1, \Omega}^{2} d \tau \\
& \leq C \int_{0}^{t} \sum_{\tau} h_{\tau}^{2} \int_{\tau}\left(y_{h t}-\operatorname{div}\left(A \nabla y_{h}\right)\right. \\
& -\int_{0}^{t} \operatorname{div}\left(\psi(t, \tau) \nabla y_{h}\right) d \tau \\
& \left.+\phi\left(y_{h}\right)-f-B u_{h}\right)^{2} d \tau \\
& +\int_{0}^{t} \sum_{\tau} h_{l} \int_{\partial \tau}\left[\left(A \nabla y_{h}\right) \cdot n\right. \\
& \left.\quad+\int_{0}^{t}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) \cdot n d \tau\right]^{2} d \tau \\
& \quad+\delta \int_{0}^{t}\left\|y_{h}-y\left(u_{h}\right)\right\|_{1, \Omega}^{2} d \tau \\
& \quad+C \int_{0}^{t} \int_{\tau}\left\|\left(y_{h}-y\left(u_{h}\right)\right)(s)\right\|_{1, \Omega}^{2} d s d \tau \\
& \quad+\left\|y_{h}(x, 0)-y_{0}(x)\right\|_{L^{2}(\Omega)}^{2} \tag{49}
\end{align*}
$$

Since $\delta$ is small enough, then from (49) and the Gronwall inequality, we have

$$
\begin{align*}
& \int_{0}^{t}\left\|y_{h}-y\left(u_{h}\right)\right\|_{1, \Omega}^{2} d \tau \\
& \leq C \int_{0}^{t} \sum_{\tau} h_{\tau}^{2} \int_{\tau}\left(y_{h t}-\operatorname{div}\left(A \nabla y_{h}\right)\right. \\
& \\
& \quad-\int_{0}^{t} \operatorname{div}\left(\psi(t, \tau) \nabla y_{h}\right) d \tau \\
& \\
& \left.+\quad \phi\left(y_{h}\right)-f-B u_{h}\right)^{2} d \tau  \tag{50}\\
& +C \int_{0}^{t} \sum_{\tau} h_{l} \int_{\partial \tau}\left[\left(A \nabla y_{h}\right) \cdot n+\int_{0}^{t}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right)\right. \\
& \\
& \quad n d \tau]^{2} d \tau+\left\|y_{h}(x, 0)-y_{0}(x)\right\|_{L^{2}(\Omega)}^{2}
\end{align*}
$$

So, by using the inequality (50) we obtain

$$
\begin{equation*}
\left\|y_{h}-y\left(u_{h}\right)\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \leq C \sum_{i=2}^{4} \eta_{i}^{2} \tag{51}
\end{equation*}
$$

This completes the proof.
Analogous to Theorem 3, we can prove the following estimates.

Theorem 4. Let $\left(y\left(u_{h}\right), p\left(u_{h}\right)\right)$ and $\left(y_{h}, p_{h}\right)$ be the solutions of (23)-(26) and (18)-(22), respectively. Then

$$
\begin{equation*}
\left\|p\left(u_{h}\right)-p_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \leq C \sum_{i=2}^{6} \eta_{i}^{2} \tag{52}
\end{equation*}
$$

where

$$
\begin{align*}
& \eta_{5}^{2}=\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{\tau}^{2} \int_{\tau}\left(y_{h}-y_{0}+p_{h t}+\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
&+\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau \\
&\left.\quad \phi^{\prime}\left(y_{h}\right) p_{h}\right)^{2} d \tau d t \\
& \eta_{6}^{2}=\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{l} \int_{\partial \tau}\left[\left(A^{*} \nabla p_{h}\right) \cdot n\right. \\
&\left.+\int_{0}^{t}\left(\left(\psi^{*}(t, \tau) \nabla p_{h}(\tau)\right) \cdot n\right) d \tau\right]^{2} d l d t \tag{53}
\end{align*}
$$

where $\eta_{2}-\eta_{4}$ are defined in Theorem 3, $l$ is a face of an element $\tau$, and $\left[\left(A^{*} \nabla p_{h}\right) \cdot n\right]$ is the $A$-normal derivative jump over the interior face $l$, defined by

$$
\begin{equation*}
\left[\left(A^{*} \nabla p_{h}\right) \cdot n\right]_{l}=\left(\left.A^{*} \nabla p_{h}\right|_{\tau_{l}^{1}}-\left.A^{*} \nabla p_{h}\right|_{\tau_{l}^{2}}\right) \cdot n \tag{54}
\end{equation*}
$$

where $n$ is the unit normal vector on $l=\bar{\tau}_{l}^{1} \cap \bar{\tau}_{l}^{2}$ outwards $\tau_{l}^{1}$. For convenience, one defines $\left[\left(A \nabla p_{h}\right) \cdot n\right]_{l}=0$ when $l \subset \partial \Omega$.

Proof. Let $e^{p}=p\left(u_{h}\right)-p_{h}$, and let $e_{I}^{p}=\hat{\pi}_{h} e^{p}$, where $\hat{\pi}_{h}$ is the Clément-type interpolator defined in Lemma 1. Then, from (20) and (25), we obtain

$$
\begin{align*}
& -\left(q_{h},\left(p_{h}-p\left(u_{h}\right)\right)_{t}\right)+a\left(q_{h}, p_{h}-p\left(u_{h}\right)\right) \\
& \quad+\int_{t}^{T} \psi\left(\tau, t ; q_{h}(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau  \tag{55}\\
& \quad+\left(\phi^{\prime}\left(y_{h}\right) p_{h}-\phi^{\prime}\left(y\left(u_{h}\right)\right) p\left(u_{h}\right), q_{h}\right) \\
& \quad=\left(y_{h}-y\left(u_{h}\right), q_{h}\right), \quad \forall q_{h} \in V_{h} .
\end{align*}
$$

Namely,

$$
\begin{align*}
& -\left(q_{h},\left(p_{h}-p\left(u_{h}\right)\right)_{t}\right)+a\left(q_{h}, p_{h}-p\left(u_{h}\right)\right) \\
& \quad+\int_{t}^{T} \psi\left(\tau, t ; q_{h}(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& \quad+\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}-p\left(u_{h}\right)\right), q_{h}\right)  \tag{56}\\
& \quad=\left(y_{h}-y\left(u_{h}\right), q_{h}\right) \\
& \quad-\left(\left(\phi^{\prime}\left(y_{h}\right)-\phi^{\prime}\left(y\left(u_{h}\right)\right)\right) p\left(u_{h}\right), q_{h}\right)
\end{align*}
$$

By using (56), we obtain

$$
\begin{aligned}
- & \left(e^{p},\right. \\
& \left.\left(p_{h}-p\left(u_{h}\right)\right)_{t}\right)+a\left(e^{p}, p_{h}-p\left(u_{h}\right)\right) \\
& +\int_{t}^{T} \psi\left(\tau, t ; e^{p}(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& +\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}-p\left(u_{h}\right)\right), e^{p}\right) \\
= & -\left(e^{p}-e_{I}^{p},\left(p_{h}-p\left(u_{h}\right)\right)_{t}\right)+a\left(e^{p}-e_{I}^{p}, p_{h}-p\left(u_{h}\right)\right) \\
& +\int_{t}^{T} \psi\left(\tau, t ;\left(e^{p}-e_{I}^{p}\right)(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& +\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}-p\left(u_{h}\right)\right), e^{p}-e_{I}^{p}\right) \\
& -\left(e_{I}^{p},\left(p_{h}-p\left(u_{h}\right)\right)_{t}\right)+a\left(e_{I}^{p}, p_{h}-p\left(u_{h}\right)\right) \\
& +\int_{t}^{T} \psi\left(\tau, t ; e_{I}^{p}(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& +\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}-p\left(u_{h}\right)\right), e_{I}^{p}\right) \\
= & \left.-\left(e^{p}-e_{I}^{p},\left(p_{h}-p\left(u_{h}\right)\right)\right)_{t}\right)+a\left(e^{p}-e_{I}^{p}, p_{h}-p\left(u_{h}\right)\right) \\
& +\int_{t}^{T} \psi\left(\tau, t ;\left(e^{p}-e_{I}^{p}\right)(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& +\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}-p\left(u_{h}\right)\right), e^{p}-e_{I}^{p}\right)+\left(y_{h}-y\left(u_{h}\right), e_{I}^{p}\right) \\
& -\left(\left(\phi^{\prime}\left(y_{h}\right)-\phi^{\prime}\left(y\left(u_{h}\right)\right)\right) p\left(u_{h}\right), e_{I}^{p}\right)
\end{aligned}
$$

$$
\begin{align*}
&=-\left(e^{p}-e_{I}^{p}, p_{h t}\right)+a\left(e^{p}-e_{I}^{p}, p_{h}\right) \\
&+\int_{t}^{T} \psi\left(\tau, t ;\left(e^{p}-e_{I}^{p}\right)(t), p_{h}\right)(\tau) d \tau \\
&+\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}\right), e^{p}-e_{I}^{p}\right) \\
&+\left(e^{p}-e_{I}^{p}, p_{t}\left(u_{h}\right)\right)-a\left(e^{p}-e_{I}^{p}, p\left(u_{h}\right)\right) \\
&-\int_{t}^{T} \psi\left(\tau, t ;\left(e^{p}-e_{I}^{p}\right)(t), p\left(u_{h}\right)(\tau)\right) d \tau \\
&-\left(\phi^{\prime}\left(y_{h}\right)\left(p\left(u_{h}\right)\right), e^{p}-e_{I}^{p}\right)+\left(y_{h}-y\left(u_{h}\right), e_{I}^{p}\right) \\
&-\left(\left(\phi^{\prime}\left(y_{h}\right)-\phi^{\prime}\left(y\left(u_{h}\right)\right)\right) p\left(u_{h}\right), e_{I}^{p}\right) \\
&= \sum_{\tau} \int_{\tau}\left(-y_{h}+y_{0}-p_{h t}-\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
&\left.\quad-\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau+\phi^{\prime}\left(y_{h}\right) p_{h}\right) \\
& \times\left(e^{p}-e_{I}^{p}\right) \\
&+\sum_{\tau} \int_{\partial \tau}\left(\left(A^{*} \nabla p_{h}\right) \cdot n+\int_{0}^{t}\left(\left(\psi^{*}(t, \tau) \nabla p_{h}(\tau)\right) \cdot n\right) d \tau\right) \\
& \quad \times\left(e^{p}-e_{I}^{p}\right)+\left(y_{h}-y\left(u_{h}\right), e_{I}^{p}\right) \\
&-\left(\left(\phi^{\prime}\left(y_{h}\right)-\phi^{\prime}\left(y\left(u_{h}\right)\right)\right) p\left(u_{h}\right), e^{p}\right) . \tag{57}
\end{align*}
$$

Then, we have

$$
\begin{align*}
& -\frac{1}{2} \frac{d}{d t}\left\|p_{h}-p\left(u_{h}\right)\right\|_{0, \Omega}^{2}+c\left\|p_{h}-p\left(u_{h}\right)\right\|_{1, \Omega}^{2} \\
& \leq-\left(e^{p},\left(p_{h}-p\left(u_{h}\right)\right)_{t}\right)+a\left(e^{p}, p_{h}-p\left(u_{h}\right)\right) \\
& \quad+\left(\phi^{\prime}\left(y_{h}\right)\left(p_{h}-p\left(u_{h}\right)\right), e^{p}\right) \\
& \leq \sum_{\tau} \int_{\tau}\left(-y_{h}+y_{0}-p_{h t}-\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
& \quad-\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau \\
& \left.\quad+\phi^{\prime}\left(y_{h}\right) p_{h}\right)\left(e^{p}-e_{I}^{p}\right) \\
& +\sum_{\tau} \int_{\partial \tau}\left(\left(A^{*} \nabla p_{h}\right) \cdot n\right. \\
& \left.\quad+\int_{t}^{T}\left(\left(\psi^{*}(t, \tau) \nabla p_{h}(\tau)\right) \cdot n\right) d \tau\right) \\
& \quad \times\left(e^{p}-e_{I}^{p}\right)+\left(y_{h}-y\left(u_{h}\right), e_{I}^{p}\right) \\
& -\left(\left(\phi^{\prime}\left(y_{h}\right)-\phi^{\prime}\left(y\left(u_{h}\right)\right)\right) p\left(u_{h}\right), e^{p}\right) \\
& +\int_{t}^{T} \psi\left(\tau, t ; e^{p}(t),\left(p_{h}-p\left(u_{h}\right)\right)(\tau)\right) d \tau . \tag{58}
\end{align*}
$$

By integrating time from $t$ to $T$ in the above inequality, combining (6) and the Schwartz inequality, we have

$$
\begin{align*}
& \frac{1}{2}\left\|p_{h}-p\left(u_{h}\right)\right\|_{0, \Omega}^{2}+c \int_{t}^{T}\left\|p_{h}-p\left(u_{h}\right)\right\|_{1, \Omega}^{2} d \tau \\
& \leq C \int_{t}^{T} \sum_{\tau} h_{\tau}^{2} \int_{\tau}\left(y_{h}-y_{0}+p_{h t}+\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
& \quad+\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau \\
& \left.\quad-\phi^{\prime}\left(y_{h}\right) p_{h}\right)^{2} d \tau \\
& \quad+\int_{t}^{T} \sum_{\tau} h_{l} \int_{\partial \tau}\left[\left(A^{*} \nabla p_{h}\right) \cdot n\right. \\
& \left.\quad+\int_{t}^{T}\left(\left(\psi^{*}(t, \tau) \nabla p_{h}(\tau)\right) \cdot n\right) d \tau\right]^{2} d \tau \\
& \quad+\delta \int_{t}^{T}\left\|p_{h}-p\left(u_{h}\right)\right\|_{1, \Omega}^{2} d \tau \\
& \quad+C \int_{t}^{T} \int_{\tau}\left\|\left(p_{h}-p\left(u_{h}\right)\right)(s)\right\|_{1, \Omega}^{2} d s d \tau \\
& \quad+C \int_{t}^{T} \int_{\tau}\left\|y_{h}-y\left(u_{h}\right)\right\|_{0, \Omega}^{2} d \tau . \tag{59}
\end{align*}
$$

Since $\delta$ is small enough, then from (59) and the Gronwall inequality, we have

$$
\begin{aligned}
& \int_{t}^{T}\left\|p_{h}-p\left(u_{h}\right)\right\|_{1, \Omega}^{2} d \tau \\
& \leq C \int_{t}^{T} \sum_{\tau} h_{\tau}^{2} \int_{\tau}\left(y_{h}-y_{0}+p_{h t}+\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
& +\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau \\
& \left.\quad-\phi^{\prime}\left(y_{h}\right) p_{h}\right)^{2} d \tau \\
& \quad+\int_{t}^{T} \sum_{\tau} h_{l} \int_{\partial \tau}\left[\left(A^{*} \nabla p_{h}\right) \cdot n\right. \\
& \left.\quad+\int_{t}^{T}\left(\left(\psi^{*}(t, \tau) \nabla p_{h}(\tau)\right) \cdot n\right) d \tau\right]^{2} d \tau \\
& \quad+\left\|y\left(u_{h}\right)-y_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}^{2}
\end{aligned}
$$

$$
\begin{align*}
& \leq C \int_{t}^{T} \sum_{\tau} h_{\tau}^{2} \int_{\tau}\left(y_{h}-y_{0}+p_{h t}+\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
&+\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau \\
&\left.-\phi^{\prime}\left(y_{h}\right) p_{h}\right)^{2} d \tau \\
&+\int_{t}^{T} \sum_{\tau} h_{l} \int_{\partial \tau}\left[\left(A^{*} \nabla p_{h}\right) \cdot n\right. \\
&\left.+\int_{t}^{T}\left(\left(\psi^{*}(t, \tau) \nabla p_{h}(\tau)\right) \cdot n\right) d \tau\right]^{2} d \tau \\
&+\left\|y\left(u_{h}\right)-y_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \tag{60}
\end{align*}
$$

Finally, combine inequality (60) and Theorem 3 to obtain

$$
\begin{equation*}
\left\|p\left(u_{h}\right)-p_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \leq C \sum_{i=2}^{6} \eta_{i}^{2} \tag{61}
\end{equation*}
$$

This completes the proof.
Hence, we combine Theorems 2-4 to conclude.
Theorem 5. Let $(y, p, u)$ and $\left(y_{h}, p_{h}, u_{h}\right)$ be the solutions of (10)-(14) and (18)-(22), respectively. Then

$$
\begin{gather*}
\left\|u-u_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}^{2}+\left\|y-y_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \\
+\left\|p-p_{h}\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} \leq C \sum_{i=1}^{6} \eta_{i}^{2} \tag{62}
\end{gather*}
$$

where $\eta_{1}-\eta_{6}$ are defined in Theorems 2-4, respectively.
Proof. From (10)-(14) and (23)-(26), we obtain the error equations

$$
\begin{align*}
& \left(y_{t}-y_{t}\left(u_{h}\right), w\right)+a\left(y-y\left(u_{h}\right), w\right) \\
& \quad+\int_{0}^{t} \psi\left(t, \tau ;\left(y-y\left(u_{h}\right)\right)(\tau), w_{h}\right) d \tau \\
& \quad+\left(\phi(y)-\phi\left(y\left(u_{h}\right)\right), w\right)=\left(B\left(u-u_{h}\right), w\right) \\
& -\left(p_{t}-p_{t}\left(u_{h}\right), q\right)+a\left(q, p-p\left(u_{h}\right)\right) \\
& \quad+\int_{t}^{T} \psi\left(\tau, t ; q_{h}(t),\left(p-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& \quad+\left(\phi^{\prime}(y) p-\phi^{\prime}\left(y\left(u_{h}\right)\right) p\left(u_{h}\right), q\right)=\left(y-y\left(u_{h}\right), q\right) \tag{63}
\end{align*}
$$

for all $w \in V$ and $q \in V$. Thus, it follows from (63) that

$$
\begin{aligned}
\left(y_{t}-\right. & \left.y_{t}\left(u_{h}\right), w\right)+a\left(y-y\left(u_{h}\right), w\right) \\
& +\int_{0}^{t} \psi\left(t, \tau ;\left(y-y\left(u_{h}\right)\right)(\tau), w_{h}\right) d \tau \\
& +\left(\phi(y)-\phi\left(y\left(u_{h}\right)\right), w\right)=\left(B\left(u-u_{h}\right), w\right)
\end{aligned}
$$

$$
\begin{align*}
- & \left(p_{t}-p_{t}\left(u_{h}\right), q\right)+a\left(q, p-p\left(u_{h}\right)\right) \\
& +\int_{t}^{T} \psi\left(\tau, t ; q_{h}(t),\left(p-p\left(u_{h}\right)\right)(\tau)\right) d \tau \\
& +\left(\phi^{\prime}\left(y\left(u_{h}\right)\right)\left(p-p\left(u_{h}\right)\right), q\right) \\
& =\left(y-y\left(u_{h}\right), q\right)+\left(\widetilde{\phi}^{\prime \prime}\left(y\left(u_{h}\right)\right)\left(y\left(u_{h}\right)-y\right) p, q\right) . \tag{64}
\end{align*}
$$

By using the stability results in [17, 19], then we can prove that

$$
\begin{align*}
\left\|y-y\left(u_{h}\right)\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} & \leq C\left\|u-u_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}^{2} \\
\left\|p-p\left(u_{h}\right)\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2} & \leq\left\|y-y\left(u_{h}\right)\right\|_{L^{2}\left(J ; H^{1}(\Omega)\right)}^{2}  \tag{65}\\
& \leq C\left\|u-u_{h}\right\|_{L^{2}\left(J ; L^{2}(\Omega)\right)}^{2}
\end{align*}
$$

Finally, combining Theorems 2-4 and (65) leads to (62).

## 4. An Adaptive Algorithm

In this section, we introduce an adaptive algorithm to guide the mesh refinement process. A posteriori error estimates which have been derived in Section 3 are used as an error indicator to guide the mesh refinement in adaptive finite element method.

Now, we discuss the adaptive mesh refinement strategy. The general idea is to refine the mesh such that the error indicator like $\eta$ is equally distributed over the computational mesh. Assume that an a posteriori error estimator $\eta$ has the form of $\eta^{2}=\sum_{\tau_{i}} \eta_{\tau_{i}}^{2}$, where $\tau_{i}$ is the finite elements. At each iteration, an average quantity of all $\eta_{\tau_{i}}^{2}$ is calculated, and each $\eta_{\tau_{i}}^{2}$ is then compared with this quantity. The element $\tau_{i}$ is to be refined if $\eta_{\tau_{i}}^{2}$ is larger than this quantity. As $\eta_{\tau_{i}}^{2}$ represents the total approximation error over $\tau_{i}$, this strategy makes sure that higher density of nodes is distributed over the area where the error is higher.

Based on this principle, we define an adaptive algorithm of the semilinear parabolic integrodifferential optimal control problems (1)-(2) as follows: starting from initial triangulations $\mathscr{T}_{h_{0}}$ of $\Omega$, we construct a sequence of refined triangulation $\mathscr{T}_{h_{j}}$ as follows. Given $\mathscr{T}_{h_{j}}$, we compute the solutions ( $y_{h}, p_{h}, u_{h}$ ) of the system (18)-(22) and their error estimator as follows:

$$
\begin{aligned}
\eta_{\tau}^{2}= & \int_{0}^{T} \sum_{\tau \in \mathscr{T}_{h}} h_{\tau}^{1+s}\left\|\alpha u_{h}+B^{*} p_{h}\right\|_{H^{s}(\tau)}^{1+s} d t \\
& +\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{\tau}^{2} \int_{\tau}\left(f+B u_{h}-y_{h t}+\operatorname{div}\left(A \nabla y_{h}\right)\right. \\
& +\int_{0}^{t} \operatorname{div}\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) d \tau \\
& \left.-\phi\left(y_{h}\right)\right)^{2} d \tau d t
\end{aligned}
$$

$$
\begin{align*}
&+\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{l} \int_{\partial \tau} {\left[\left(A \nabla y_{h}\right) \cdot n\right.} \\
&\left.+\int_{0}^{t}\left(\left(\psi(t, \tau) \nabla y_{h}(\tau)\right) \cdot n\right) d \tau\right]^{2} d l d t \\
&+\left\|y_{h}(x, 0)-y_{0}(x)\right\|_{L^{2}(\Omega)}^{2} \\
&+\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}} h_{\tau}^{2} \int_{\tau}\left(y_{h}-y_{0}+p_{h t}+\operatorname{div}\left(A^{*} \nabla p_{h}\right)\right. \\
&+\int_{t}^{T} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p_{h}(\tau)\right) d \tau \\
&+\int_{0}^{T} \sum_{\tau \in \mathscr{T}^{h}}^{\sum_{l} h_{l} \int_{\partial \tau}[ } \begin{aligned}
& \left(A^{*} \nabla p_{h}\right) \cdot n \\
& \left.\left.+\int_{0}^{t}\left(\left(\psi_{h}^{*}\right) p_{h}\right) d \tau d t, \tau\right) \nabla p_{h}(\tau)\right) \\
& E_{j}=\sum_{\tau \in \mathscr{T}_{h}} \eta_{\tau}^{2}
\end{aligned}
\end{align*}
$$

Then, we adopt the following mesh refinement strategy: all the triangles $\tau \in \mathscr{T}_{h_{j}}$ satisfying $\eta_{\tau}^{2} \geq \rho E_{j} / n$ are divided into four new triangles in $\mathscr{T}_{h_{j+1}}$ by joining the midpoints of the edges, where $n$ is the number of the elements of $\mathscr{T}_{h_{j}}$ and $\rho$ is a given constant. In order to maintain the new triangulation $\mathscr{T}_{h_{j+1}}$ to be regular and conformal, some additional triangles need to be divided into two or four new triangles depending on whether they have one or more neighbors which have been refined. Then, we obtain the new mesh $\mathscr{T}_{h_{j+1}}$. The above procedure will continue until $E_{j} \leq$ tol, where tol is a given tolerance error.

## 5. Numerical Example

In this section, we will give a numerical example to illustrate our theoretical results. Our numerical example is the following semilinear parabolic integrodifferential optimal control problem:

$$
\begin{aligned}
& \min _{u(t) \in K}\left\{\int_{0}^{1}\left(\frac{1}{2}\left\|y-y_{0}\right\|^{2}+\frac{1}{2}\left\|u-u_{0}\right\|^{2}\right) d t\right\} \\
& y_{t}-\operatorname{div}(\nabla y(x, t))-\int_{0}^{t} \operatorname{div}(\psi(t, \tau) \nabla y(x, \tau)) d \tau \\
& \quad+\phi(y)=f+u, \quad x \in \Omega, t \in J
\end{aligned}
$$

$$
\begin{gather*}
y(x, t)=0, \quad x \in \partial \Omega, t \in J \\
y(x, 0)=0, \quad x \in \Omega \\
-p_{t}-\operatorname{div}(\nabla p(x, t))-\int_{t}^{1} \operatorname{div}\left(\psi^{*}(\tau, t) \nabla p(x, \tau)\right) d \tau  \tag{67}\\
+\phi^{\prime}(y) p=y-y_{0}, \quad x \in \Omega, t \in J \\
p(x, t)=0, \quad x \in \partial \Omega, t \in J \\
p(x, 1)=0, \quad x \in \Omega
\end{gather*}
$$

In this example, we choose the domain $\Omega=[0,1] \times[0,1]$. Let $\Omega$ be partitioned into $\mathscr{T}_{h}$ as described Section 2 . For the constrained optimization problem:

$$
\begin{equation*}
\min _{u \in K \subset U} \int_{0}^{1} S(u) d t \tag{68}
\end{equation*}
$$

where $S(u)=(1 / 2)\left\|y-y_{0}\right\|^{2}+(1 / 2)\left\|u-u_{0}\right\|^{2}$ is a convex functional on $U$ and $K=\{u \in U: u \geq 0$ a.e. in $\Omega \times J\}$; the iterative scheme reads ( $n=0,1,2, \ldots$ )

$$
\begin{gather*}
b\left(u_{n+1 / 2}, v\right)=b\left(u_{n}, v\right)-\rho_{n}\left(S^{\prime}\left(u_{n}\right), v\right), \quad \forall v \in U,  \tag{69}\\
u_{n+1}=P_{K}^{b}\left(u_{n+1 / 2}\right)
\end{gather*}
$$

where $b(\cdot, \cdot)$ is a symmetric and positive definite bilinear form such that there exist constants $c_{0}$ and $c_{1}$ satisfying

$$
\begin{gather*}
|b(u, v)| \leq c_{1}\|u\|_{U}\|v\|_{U}, \quad \forall u, v \in U, \\
b(u, u) \geq c_{0}\|u\|_{U}^{2}, \tag{70}
\end{gather*}
$$

and the projection operator $P_{K}^{b} U \rightarrow K$ is defined: for given $w \in U$ find $P_{K}^{b} w \in K$ such that

$$
\begin{equation*}
b\left(P_{K}^{b} w-w, P_{K}^{b} w-w\right)=\min _{u \in K} b(u-w, u-w) \tag{71}
\end{equation*}
$$

The bilinear form $b(\cdot, \cdot)$ provides suitable preconditioning for the projection algorithm. An application of (69) to the discretized semilinear parabolic integrodifferential optimal control problem yields the following algorithm:

$$
\begin{gather*}
b\left(u_{n+1 / 2}, v_{h}\right)=b\left(u_{n}, v_{h}\right)-\rho_{n}\left(u_{n}+p_{n}, v_{h}\right), \quad \forall v_{h} \in K_{h}, \\
\int_{0}^{1}\left(\left(y_{t}, w\right)+a(y, w)+\int_{0}^{t} \psi(t, \tau ; y(\tau), w) d \tau\right. \\
+(\phi(y), w)) d t=\int_{0}^{1}(f+u, w) d t, \quad \forall w \in V \\
\int_{0}^{1}\left(-\left(p_{t}, q\right)+a(q, p)+\int_{t}^{1} \psi(\tau, t ; q, p(\tau)) d \tau\right. \\
\left.+\left(\phi^{\prime}(y) p, q\right)\right) d t=\int_{0}^{1}\left(y-y_{0}, q\right) d t, \quad \forall q \in V \\
u_{n+1}=P_{K}^{b}\left(u_{n+1 / 2}\right), \quad u_{n+1 / 2}, u_{n} \in K_{h} . \tag{72}
\end{gather*}
$$

The main computational effort is to solve the state and costate equations and to compute the projection $P_{K}^{b} u_{n+1 / 2}$. In this paper, we use a fast algebraic multigrid solver to solve the state and costate equations. Then, it is clear that the key to saving computing time is how to compute $P_{K}^{b} u_{n+1 / 2}$ efficiently. For the piecewise constant elements, $K_{h}=\left\{u_{h}: u_{h} \geq 0\right\}$ and $b(u, v)=(u, v)_{U}$; then

$$
\begin{equation*}
\left.P_{K}^{b} u_{n+1 / 2}\right|_{\tau}=\max \left(0,\left.\operatorname{avg}\left(u_{n+1 / 2}\right)\right|_{\tau}\right) \tag{73}
\end{equation*}
$$

where $\left.\operatorname{avg}\left(u_{n+1 / 2}\right)\right|_{\tau}$ is the average of $u_{n+1 / 2}$ over $\tau$.
In solving our discretized optimal control problem, we use the preconditioned projection gradient method with $b(u, v)=(u, v)_{U}$ and a fixed step size $\rho=0.9$. We now briefly describe the solution algorithm to be used for solving the numerical example in this section as follows.
(1) Solve the discretized optimization problem with the projection gradient method on the current meshes and calculate the error estimators $\eta_{i}$.
(2) Adjust the meshes using the estimators and update the solution on new meshes, as described.

Now, we give a numerical example to illustrate our theoretical results.

Example 1. Let $\psi(t, \tau)=1, \phi(y)=y^{5}$. We choose the state function by

$$
\begin{equation*}
y\left(x_{1}, x_{2}\right)=2 \sin \pi x_{1} \sin \pi x_{2} \sin \pi t . \tag{74}
\end{equation*}
$$

The function $f$ is given by $f(x)=y_{t}-\operatorname{div}(\nabla y(x, t))-$ $\int_{0}^{t} \operatorname{div}(\nabla y(x, \tau)) d \tau+y^{5}-u$. The costate function can be chosen

$$
\begin{equation*}
p\left(x_{1}, x_{2}\right)=\sin \pi x_{1} \sin \pi x_{2} \sin \pi t \tag{75}
\end{equation*}
$$

The function $y_{0}$ is given by $y_{0}(x)=y+p_{t}+\operatorname{div}(\nabla p(x, t))+$ $\int_{t}^{1} \operatorname{div}(\nabla p(x, \tau)) d \tau-5 y^{4} p$. We assume that

$$
\begin{gather*}
\lambda= \begin{cases}0.8, & x_{1}+x_{2}>1.0, \\
0.3, & x_{1}+x_{2} \leq 1.0,\end{cases}  \tag{76}\\
u_{0}\left(x_{1}, x_{2}\right)=1-\sin \frac{\pi x_{1}}{2}-\sin \frac{\pi x_{2}}{2}+\lambda .
\end{gather*}
$$

Thus, the control function is given by

$$
\begin{equation*}
u\left(x_{1}, x_{2}\right)=\max \left(u_{0}-p, 0\right) \tag{77}
\end{equation*}
$$

In this example, the control function $u$ has a strong discontinuity introduced by $u_{0}$. The control function $u$ is discretized by piecewise constant functions, whereas the state $y$ and the costate $p$ were approximated by piecewise linear functions. In Table 1, numerical results of $u, y$, and $p$ on uniform and adaptive meshes are presented. It can be found that the adaptive meshes generated using our error indicators can save substantial computational work, in comparison with the uniform meshes. On the other hand, for the discontinuous control variable $u$, the accuracy has become better from the uniform meshes to the adaptive meshes in Table 1.

Table 1: Numerical results on uniform and adaptive meshes.

|  | On uniform mesh |  |  | On adaptive mesh |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $u$ | $y$ | $p$ | $u$ | $y$ | $p$ |
| Nodes | 8097 | 8097 | 8097 | 1102 | 1969 | 1969 |
| Sides | 23968 | 23968 | 23968 | 3143 | 5744 | 5744 |
| Elements | 15872 | 15872 | 15872 | 2042 | 3776 | 3776 |
| Dofs | 15872 | 15872 | 15872 | 2042 | 3776 | 3776 |
| Total $L^{2}$ error | $4.312 e-03$ | $5.457 e-3$ | $2.869 e-3$ | $4.018 e-03$ | $5.365 e-3$ | $2.768 e-3$ |

## 6. Conclusion and Future Works

In this paper, we discuss the semi-discrete finite element methods of the semilinear parabolic integrodifferential optimal control problems (1)-(2). We have established a posteriori error estimates for each the state, the costate, and the control approximation. The posteriori error estimates for those problems by finite element methods seem to be new.

In our future work, we will use the mixed finite element method to deal with nonlinear parabolic integrodifferential optimal control problems. Furthermore, we will consider a posteriori error estimates and superconvergence of mixed finite element solution for nonlinear parabolic integrodifferential optimal control problems.

## Acknowledgments

The author expresses his thanks to the referees for their helpful suggestions, which led to improvements of the presentation. This work is supported by the National Science Foundation of China (11201510), Mathematics TianYuan Special Funds of the National Natural Science Foundation of China (11126329), China Postdoctoral Science Foundation funded project (2011M500968), Natural Science Foundation Project of CQ CSTC (cstc2012jjA00003), Natural Science Foundation of Chongqing Municipal Education Commission (KJ121113), and Science and Technology Project of Wanzhou District of Chongqing (2013030050).

## References

[1] R.S. Falk, "Approximation of a class of optimal control problems with order of convergence estimates," Journal of Mathematical Analysis and Applications, vol. 44, pp. 28-47, 1973.
[2] T. Geveci, "On the approximation of the solution of an optimal control problem governed by an elliptic equation," RAIRO Analyse Numérique, vol. 13, no. 4, pp. 313-328, 1979.
[3] E. Casas, M. Mateos, and F. Tröltzsch, "Error estimates for the numerical approximation of boundary semilinear elliptic control problems," Computational Optimization and Applications, vol. 31, no. 2, pp. 193-219, 2005.
[4] Z. Lu, "A posteriori error estimates of finite element methods for nonlinear quadratic boundary optimal control problem," Numerical Analysis and Applications, vol. 4, pp. 210-222, 2011.
[5] Y. Chen and W. Liu, "A posteriori error estimates for mixed finite element solutions of convex optimal control problems," Journal of Computational and Applied Mathematics, vol. 211, no. 1, pp. 76-89, 2008.
[6] D. A. French and J. T. King, "Approximation of an elliptic control problem by the finite element method," Numerical Functional Analysis and Optimization, vol. 12, no. 3-4, pp. 299-314, 1991.
[7] M. D. Gunzburger and L. S. Hou, "Finite-dimensional approximation of a class of constrained nonlinear optimal control problems," SIAM Journal on Control and Optimization, vol. 34, no. 3, pp. 1001-1043, 1996.
[8] P. Neittaanmäki and D. Tiba, Optimal Control of Nonlinear Parabolic Systems: Theory, Algorithms and Applications, Marcel Dekker, New York, NY, USA, 1994.
[9] J.-L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer, Berlin, Germany, 1971.
[10] H. Brunner and N. Yan, "Finite element methods for optimal control problems governed by integral equations and integrodifferential equations," Applied Numerical Mathematics, vol. 47, pp. 173-187, 2003.
[11] W. Liu and N. Yan, "A posteriori error estimates for distributed convex optimal control problems," Numerische Mathematik, vol. 101, pp. 1-27, 2005.
[12] R. Verfürth, "A posteriori error estimates for nonlinear problems. Finite element discretizations of elliptic equations," Mathematics of Computation, vol. 62, no. 206, pp. 445-475, 1994.
[13] Y. Chen and Z. Lu, "Error estimates of fully discrete mixed finite element methods for semilinear quadratic parabolic optimal control problem," Computer Methods in Applied Mechanics and Engineering, vol. 199, no. 23-24, pp. 1415-1423, 2010.
[14] Y. Chen and Z. Lu, "Error estimates for parabolic optimal control problem by fully discrete mixed finite element methods," Finite Elements in Analysis and Design, vol. 46, no. 11, pp. 957965, 2010.
[15] Z. Lu and Y. Chen, "A posteriori error estimates of triangular mixed finite element methods for semilinear optimal control problems," Advances in Applied Mathematics and Mechanics, vol. 1, no. 2, pp. 242-256, 2009.
[16] Z. Lu and Y. Chen, " $L^{\infty}$-error estimates of triangular mixed finite element methods for optimal control problem govern by semilinear elliptic equation," Numerical Analysis and Applications, vol. 2, pp. 74-86, 2009.
[17] E. G. Yanik and G. Fairweather, "Finite element methods for parabolic and hyperbolic partial integro-differential equations," Nonlinear Analysis. Theory, Methods \& Applications, vol. 12, no. 8, pp. 785-809, 1988.
[18] W. Liu and N. Yan, "A posteriori error estimates for distributed convex optimal control problems," Advances in Computational Mathematics, vol. 15, no. 1-4, pp. 285-309, 2001.
[19] I. H. Sloan and V. Thomée, "Time discretization of an integrodifferential equation of parabolic type," SIAM Journal on Numerical Analysis, vol. 23, no. 5, pp. 1052-1061, 1986.

## Research Article

# Switching Signal Design for Exponential Stability of Uncertain Discrete-Time Switched Time-Delay Systems 

Zheng-Fan Liu, Chen-Xiao Cai, and Yun Zou<br>School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China<br>Correspondence should be addressed to Chen-Xiao Cai; ccx5281@vip.163.com

Received 4 June 2013; Accepted 9 August 2013
Academic Editor: Weihai Zhang
Copyright © 2013 Zheng-Fan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

The switching signal design for exponential stability with $H_{\infty}$ performance of uncertain switched linear discrete-time systems with interval time-varying delay is considered. Systems with norm-bounded parameter uncertainties are considered. By taking a new Lyapunov-Krasovskii (LK) function, sufficient conditions for the existence of a class of stabilizing switching laws are derived in terms of linear matrix inequalities (LMIs) to guarantee the considered switched time-delay system to be exponentially stable. The resulting stability criteria are of fewer matrix variables and are less conservative than some existing ones. In addition, numerical examples are illustrated to show the main improvement.


## 1. Introduction

Switched system is represented as the family of subsystems with switching rule which is concerned with various environmental factors and different controllers. During the last decades, there has been increasing interest in the stability analysis and control design for the switched systems (see, e.g., [1-6]). The design of switching signal is one of the three basic problems in stability analysis and the design of switched systems [2,3]. Switching signal included timedriven switching and state-driven switching. In the first case, time-driven switching depends on time, and many effective methods have been developed such as dwell time method [7, 8], average dwell time method [9,10], and mode-dependent average dwell time method [11, 12]. In the case of state-driven switching, a switching action takes place when the system state hits a switching surface, and the study of the stability is based on a number of methods, including piecewise Lyapunov function [13] and convex combination [14].

On the other hand, time delay is one of the instability sources for dynamical systems and is a common phenomenon in many industrial and engineering systems. During the last two decades, much attention has been paid to the problem of stability analysis and controller synthesis for time-delay systems (see, e.g., [15-18]). A switched system with time-delay individual subsystems is called a switched
time-delay system [19-28]. Switched time-delay systems have various applications in practical engineering systems, such as power systems and power electronics [1, 21]. Many important results on the dynamical behavior have been reported for switched time-delay system [8-12].

Recently, increasing attention has been devoted to the problem of delay-dependent stability of switched delay systems [22-26]. In [23-25], a switching signal design technique is proposed to guarantee the asymptotic stability of switched systems with interval time-varying delay. Based on a discrete LK functional, in [26] a switching rule for the asymptotic stability and stabilization for a class of discretetime switched systems with interval time-varying delays is designed via linear matrix inequalities. However, the $\operatorname{term} \sum_{s=k+1-d(k)}^{k-d_{m}} x^{T}(s) Q x(s)=\sum_{s=k+1-d_{M}}^{k-d_{m}} x^{T}(s) Q x(s)-$ $\sum_{s=k+1-d_{M}}^{k-d(k)} x^{T}(s) Q x(s)$ is estimated as $\sum_{s=k+1-d_{M}}^{k-d_{m}} x^{T}(s) Q x(s)$ for any $0<d_{m} \leq d(k) \leq d_{M}, Q=Q^{T}>0$, which may lead to considerable conservativeness. Following the work, free weighting matrix and additional nonnegative inequality approaches have been used to improve the conservativeness for the obtained results in [27]. However, many free weighting matrices were introduced, which made the stability result complicated. $H_{\infty}$ concept was proposed to reduce the effect of the disturbance input on the regulated output to
within a prescribed level. In [29], $H_{\infty}$ controls were proposed for uncertain discrete switched systems under arbitrary switching signal. In [30], a switching signal design for $H_{\infty}$ performance of uncertain discrete switched systems with interval delay and linear fractional perturbations is considered. Nevertheless, the criteria still leave some room for improvement in accuracy as well as complexity due to the method used and offer motivation for further investigation.

Motivated by the above literatures, in this paper, by using an improved discrete LK function combined with LMIs technique, a new switching signal design approach is developed to guarantee the $H_{\infty}$ performance for uncertain discrete switched systems with interval time-varying delay to be exponentially stable. Numerical examples are given to show the effectiveness of the proposed method which can be easily solved by using MATLAB LMI control toolbox.

The remaining part of the paper is organized as follows. In Section 2, some preliminaries are introduced. In Section 3, the main results for uncertain switched linear discrete-time systems with interval time-varying delay are presented to be exponential stability. In Section 4, some numerical examples are given to illustrate the effectiveness and the merit of the proposed method. The last section concludes the work.

Notations. We use standard notations throughout the paper. $\lambda_{\text {min }}(M)\left(\lambda_{\max }(M)\right)$ stands for the minimal (maximum) eigenvalue of $M . M^{T}$ is the transpose of the matrix $M$. The relation $M>N(M<N)$ means that the matrix $M-N$ is positive (negative) definite. $\|x\|$ denotes the Euclidian-norm of the vector $x \in R^{n}$. $R^{n}$ represents the $n$-dimensional real Euclidean space. $R^{n \times m}$ is the set of all real $n \times m$ matrices. $\operatorname{diag}\{\cdots\}$ stands for a block-diagonal matrix. In symmetric block matrices or long matrix expressions, we use an asterisk "*" to represent a term that is induced by symmetry. $I$ denotes the identity matrix.

## 2. Problem Description and Preliminaries

Consider the following uncertain linear discrete-time switched time-delay system:

$$
\begin{align*}
x(k+1)= & {\left[A_{\sigma}+\Delta A_{\sigma}(k)\right] x(k) } \\
& +\left[A_{d \sigma}+\Delta A_{d \sigma}(k)\right] x(k-d(k)) \\
& +\left[B_{\sigma}+\Delta B_{\sigma}(k)\right] w(k), \\
y(k)= & {\left[C_{\sigma}+\Delta C_{\sigma}(k)\right] x(k) }  \tag{1}\\
& +\left[C_{d \sigma}+\Delta C_{d \sigma}(k)\right] x(k-d(k)) \\
& +\left[D_{\sigma}+\Delta D_{\sigma}(k)\right] w(k), \\
x(l)= & \phi(l), \quad l=k_{0}-d_{M}, \ldots, k_{0}
\end{align*}
$$

where $x(k) \in R^{n}$ denotes the system state vector. $y(k) \in R^{m}$ is the measured output. $w(k) \in R^{p}$ is the disturbance input which belongs to $l_{2}[0, \infty) . \phi(l) \in R^{n}$ is a vector-valued initial function. The switching signal $\sigma: Z \rightarrow \mathcal{Z}=\{1,2, \ldots, N\}$ is a piecewise constant function. $\sigma=i$ means that the $i$ th
subsystem is activated. $N$ is the number of subsystems of the switched system. The system matrices $A_{i}, A_{d i}, B_{i}, C_{i}, C_{d i}$, and $D_{i}$ are a set of known real matrices with appropriate dimensions. $\Delta A_{i}(k), \Delta A_{d i}(k), \Delta B_{i}(k), \Delta C_{i}(k), \Delta C_{d i}(k)$, and $\Delta D_{i}(k)$ are real-valued unknown matrices representing time-varying parameter uncertainties, and are assumed to be of the form

$$
\begin{gather*}
{\left[\begin{array}{ccc}
\Delta A_{i}(k) & \Delta A_{d i}(k) & \Delta B_{i}(k) \\
\Delta C_{i}(k) & \Delta C_{d i}(k) & \Delta D_{i}(k)
\end{array}\right]=\left[\begin{array}{l}
M_{1 i} \\
M_{2 i}
\end{array}\right]}  \tag{2}\\
\Delta_{1 i}(k)\left[\begin{array}{lll}
N_{1 i} & N_{2 i} & N_{3 i}
\end{array}\right]
\end{gather*}
$$

where $M_{1 i}, M_{2 i}, N_{1 i}, N_{2 i}$, and $N_{3 i}$ are known real constant matrices and $\Delta_{1 i}(k): N \rightarrow R^{l_{1} \times l_{2}}$ is unknown time-varying matrix function satisfying

$$
\begin{equation*}
\Delta_{1 i}^{T}(k) \Delta_{1 i}(k) \leq I \tag{3}
\end{equation*}
$$

The parameter uncertainties $\Delta A_{i}(k), \Delta A_{d i}(k), \Delta B_{i}(k)$, $\Delta C_{i}(k), \Delta C_{d i}(k)$, and $\Delta D_{i}(k)$ are said to be admissible if both (2) and (3) hold.

For given finite positive integer $d_{m}$ and $d_{M}$, time-varying delay $d(k)$ satisfying

$$
\begin{equation*}
0<d_{m} \leq d(k) \leq d_{M}, \quad \forall k \in N^{+} \tag{4}
\end{equation*}
$$

Now we present the following definitions and lemmas that are useful in deriving the principal contribution of this paper.

Definition 1 (see [20]). The system (1) is said to be exponentially stable if there exist a switching function $\sigma(\cdot)$ and positive number $c$ such that any solution $x(k, \phi)$ of the system satisfies

$$
\begin{equation*}
\|x(k)\| \leq c \lambda^{k-k_{0}}\|\phi\|_{s}, \quad \forall k \geq k_{0} \tag{5}
\end{equation*}
$$

for any initial conditions $\left(k_{0}, \phi\right) \in R^{+} \times C^{n} . c>0$ is the decay coefficient, $0<\lambda \leq 1$ is the decay rate, and $\|\phi\|_{s}=\sup \{\|\phi(l)\|$, $\left.l=k_{0}-d_{M}, k_{0}-d_{M}+1, \ldots, k_{0}\right\}$.

Definition 2 (see [26]). The system of matrices $\left\{L_{i}\right\}(i \in U)$ is said to be strictly complete if, for every $x \in R^{n} \backslash\{0\}$, there is $i \in U$ such that $x^{T} L_{i} x<0$.

It is easy to see that the system of matrices $\left\{L_{i}\right\}(i \in U)$ is strictly complete if and only if $\bigcup_{i=1}^{N} \Omega_{i}=R^{n} \backslash\{0\}$, where $\Omega_{i}=$ $\left\{x \in R^{n}: x^{T} L_{i} x<0\right\}(i \in \mho)$.

Lemma 3 (see [31]). The system of matrices $\left\{L_{i}\right\}(i \in \mathcal{J})$ is strictly complete if there exist $\alpha_{i} \geq 0, \sum_{i=1}^{N} \alpha_{i}=1$ such that $\sum_{i=1}^{N} \alpha_{i} L_{i}<0$. If $N=2$; then the above condition is also necessary for the strict completeness.

Lemma 4 (see [17]). For any matrix $R=R^{T}>0$ integers $a \leq b$, if vector function $\xi(k):\{-b,-b+1, \ldots,-a\} \rightarrow R^{n}$, then

$$
(a-b) \sum_{s=k-b}^{k-a-1} z^{T}(s) R z(s) \leq \xi^{T}(k)\left[\begin{array}{cc}
-R & R  \tag{6}\\
* & -R
\end{array}\right] \xi(k)
$$

where

$$
\begin{gather*}
z(k)=x(k+1)-x(k) \\
\xi^{T}(k)=\left[\begin{array}{ll}
x^{T}(k-a) & x^{T}(k-b)
\end{array}\right] . \tag{7}
\end{gather*}
$$

Lemma 5 (see [32], Schur's complement). Let $M, P$, and $Q$ be given matrices such that $Q>0$. Then

$$
\left[\begin{array}{cc}
P & M  \tag{8}\\
* & -Q
\end{array}\right]<0 \Longleftrightarrow P+M Q^{-1} M^{T}<0 .
$$

The objective of this paper is to design a reasonable switching rule for discrete-time switched system (1) with time-varying delay satisfying (4) to guarantee that the system while be exponentially stable.

## 3. Main Result

In this section, we will divide the whole state space $R^{n}$ into $N$ subregions and then define a particular quadratic function in each sub region. Via an appropriate designed switching rule, the particular quadratic function in each sub region will decrease along the system trajectory with the corresponding subsystem. Consequently, the whole switched system remains exponentially stable.

Firstly, we will introduce the switching regions and the corresponding switching law. Given $P>0$ and $U>0$, define the domains by

$$
\begin{equation*}
\Omega_{i}\left(P, U, A_{i}\right)=\left\{x(k) \in R^{n}: x^{T}(k) Y_{i} x(k)<0\right\}, \tag{9}
\end{equation*}
$$

where $Y_{i}=A_{i}^{T} P A_{i}-U, i \in U$.
From the similar proof of [26, 27], it can be easily obtained that

$$
\begin{equation*}
\bigcup_{i=1}^{N} \Omega_{i}=R^{n} \backslash\{0\} \tag{10}
\end{equation*}
$$

Construct the following switching region:

$$
\begin{align*}
& \bar{\Omega}_{1}=\Omega_{1}, \quad \bar{\Omega}_{2}=\Omega_{2} \backslash \bar{\Omega}_{1}, \quad \bar{\Omega}_{3}=\Omega_{3} \backslash \bar{\Omega}_{1} \backslash \bar{\Omega}_{2}, \ldots, \\
& \bar{\Omega}_{N}=\Omega_{N} \backslash \bar{\Omega}_{1} \backslash \cdots \backslash \bar{\Omega}_{N-1} \tag{11}
\end{align*}
$$

We can obtain $\bigcup_{i=1}^{N} \bar{\Omega}_{i}=R^{n} \backslash\{0\}$ and $\bar{\Omega}_{i} \cap \bar{\Omega}_{j}=\phi$, for all $i \neq j$, where $\phi$ is an empty set.

After dividing the whole state space $R^{n}$ into $N$ sub regions, we construct the switching signal as follows:

$$
\begin{equation*}
\sigma(x(k))=i, \quad \forall x(k) \in \bar{\Omega}_{i}(i \in \mathcal{J}) \tag{12}
\end{equation*}
$$

In order to discuss robust stability of system (1), first, we consider the following nominal system without parametric uncertainties:

$$
\begin{gather*}
x(k+1)=A_{i} x(k)+A_{d i} x(k-d(k))+B_{i} w(k),  \tag{13}\\
y(k)=C_{i} x(k)+C_{d i} x(k-d(k))+D_{i} w(k) .
\end{gather*}
$$

The following theorem gives a sufficient condition for the existence of an admissible reasonable switching rule for system (13) with disturbance input $w(k)=0$ to be exponentially stable.

Theorem 6. For some constants $\gamma \in(0,1]$ and $0 \leq \alpha_{i} \leq 1, i \in$ $U, \sum_{i=1}^{N} \alpha_{i}=1$, if there exist positive definite symmetric matrices $P, U, Q_{j}, R_{j}, j=1,2,3$, such that the following LMIs hold:

$$
\begin{gather*}
\Xi=\left[\begin{array}{ccccc}
\phi_{11} & \phi_{12} & \phi_{13} & 0 & \phi_{15} \\
* & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} \\
* & * & \phi_{33} & 0 & 0 \\
* & * & * & \phi_{44} & 0 \\
* & * & * & * & -W
\end{array}\right]<0,  \tag{14}\\
\sum_{i=1}^{N} \alpha_{i} A_{i}^{T} P A_{i}<U \tag{15}
\end{gather*}
$$

then the system (13) with time-varying delay satisfying (4) and $w(k)=0$ is globally exponentially stable with convergence rate $\lambda=\sqrt{\gamma}$ by the switching signal designed by (12).

Here

$$
\begin{align*}
\phi_{11}= & U-\gamma P+Q_{1}+Q_{3}-\frac{\gamma^{d_{m}}}{d_{m}}\left(R_{1}+R_{3}\right) \\
\phi_{12}= & A_{i}^{T} P B_{i}, \quad \phi_{13}=\frac{\gamma^{d_{m}}}{d_{m}}\left(R_{1}+R_{3}\right), \\
\phi_{15}= & \left(A_{i}-I\right)^{T} W^{T}, \\
\phi_{22}= & B_{i}^{T} P B_{i}-\gamma^{d_{M}} Q_{3}-\frac{\gamma^{d_{M}}}{d_{M}-d_{m}}\left(2 R_{2}+R_{3}\right), \\
\phi_{23}= & \frac{\gamma^{d_{M}}}{d_{M}-d_{m}}\left(R_{2}+R_{3}\right) \\
\phi_{24}= & \frac{\gamma^{d_{M}}}{d_{M}-d_{m}} R_{2},  \tag{16}\\
\phi_{25}= & B_{i}^{T} W^{T}, \\
\phi_{33}= & \gamma^{d_{m}}\left(Q_{2}-Q_{1}\right)-\frac{\gamma^{d_{m}}}{d_{m}}\left(R_{1}+R_{3}\right) \\
& -\frac{\gamma^{d_{M}}}{d_{M}-d_{m}}\left(R_{2}+R_{3}\right) \\
\phi_{44}= & -\gamma^{d_{M}} Q_{2}-\frac{\gamma^{d_{M}}}{d_{M}-d_{m}} R_{2} \\
W= & \left(d_{M}-d_{m}\right) R_{2}+d_{m} R_{1}+d_{M} R_{3} .
\end{align*}
$$

Proof. Consider the following LK function for system (1):

$$
\begin{equation*}
V(k)=V_{1}(k)+V_{2}(k)+V_{3}(k) . \tag{17}
\end{equation*}
$$

Here

$$
\begin{align*}
V_{1}(k)= & x^{T}(k) P x(k),  \tag{18}\\
V_{2}(k)= & \sum_{s=k-d_{m}}^{k-1} \gamma^{k-1-s} x^{T}(s) Q_{1} x(s) \\
& +\sum_{s=k-d_{M}}^{k-d_{m}-1} \gamma^{k-1-s} x^{T}(s) Q_{2} x(s)  \tag{19}\\
& +\sum_{s=k-d(k)}^{k-1} \gamma^{k-1-s} x^{T}(s) Q_{3} x(s), \\
V_{3}(k)= & \sum_{\theta=-d_{m}}^{-1} \sum_{s=k+\theta}^{k-1} \gamma^{k-1-s} z^{T}(s) R_{1} z(s) \\
& +\sum_{\theta=-d_{M}}^{-d_{m}-1} \sum_{s=k+\theta}^{k-1} \gamma^{k-1-s} z^{T}(s) R_{2} z(s)  \tag{20}\\
& +\sum_{\theta=-d(k)}^{-1} \sum_{s=k+\theta}^{k-1} \gamma^{k-1-s} z^{T}(s) R_{3} z(s),
\end{align*}
$$

where $P, Q_{j}, R_{j}, j=1,2,3$ are positive definite symmetric matrices and $\gamma \in(0,1]$.

Now, we will show the decay estimation of $V(k)$ in (17) along the state trajectory of system (1). To this end, define

$$
\begin{equation*}
V(k+1)-\gamma V(k)=\sum_{j=1}^{3} \widetilde{\Delta} V_{j}(k) . \tag{21}
\end{equation*}
$$

Here

$$
\begin{aligned}
\widetilde{\Delta} V_{1}(k)= & x^{T}(k)\left(A_{i}^{T} P A_{i}-\gamma P\right) x(k) \\
& +2 x^{T}(k) A_{i}^{T} P B_{i} x(k-d(k)) \\
& +x^{T}(k-d(k)) B_{i}^{T} P B_{i} x(k-d(k)), \\
\widetilde{\Delta} V_{2}(k)= & x^{T}(k)\left(Q_{1}+Q_{3}\right) x(k) \\
& -\gamma^{d_{m}} x^{T}\left(k-d_{m}\right)\left(Q_{1}-Q_{2}\right) x\left(k-d_{m}\right) \\
& -\gamma^{\tau(k)} x^{T}(k-d(k)) Q_{3} x(k-d(k)) \\
& -\gamma^{d_{M}} x^{T}\left(k-d_{M}\right) Q_{2} x\left(k-d_{M}\right) \\
\leq & x^{T}(k)\left(Q_{1}+Q_{3}\right) x(k) \\
& -\gamma^{d_{M}} x^{T}(k-d(k)) Q_{3} x(k-d(k))
\end{aligned}
$$

$$
\begin{align*}
& -\gamma^{d_{m}} x^{T}\left(k-d_{m}\right)\left(Q_{1}-Q_{2}\right) x\left(k-d_{m}\right) \\
& -\gamma^{d_{M}} x^{T}\left(k-d_{M}\right) Q_{2} x\left(k-d_{M}\right), \tag{23}
\end{align*}
$$

$$
\begin{align*}
\tilde{\Delta} V_{3}(k)= & z^{T}(k)\left(\left(d_{M}-d_{m}\right) R_{2}+d_{m} R_{1}+d(k) R_{3}\right) z(k) \\
& -\sum_{s=k-d_{m}}^{k-1} \gamma^{k-s} z^{T}(s) R_{1} z(s) \\
& -\sum_{s=k-d_{M}}^{k-d_{m}-1}\left\{\gamma^{k-s} z^{T}(s) R_{2} z(s)\right\} \\
& -\sum_{s=k-d(k)}^{k-1} \gamma^{k-s} z^{T}(s) R_{3} z(s) \tag{24}
\end{align*}
$$

while

$$
\begin{aligned}
& -\sum_{s=k-d(k)}^{k-1} \gamma^{k-s} z^{T}(s) R_{3} z(s) \\
& =-\sum_{s=k-d_{m}}^{k-1}\left\{\gamma^{k-s} z^{T}(s) R_{3} z(s)\right\} \\
& -\sum_{s=k-d(k)}^{k-d_{m}-1} \gamma^{k-s} z^{T}(s) R_{3} z(s), \\
& -\sum_{s=k-d_{M}}^{k-d_{m}-1} \gamma^{k-s} z^{T}(s) R_{2} z(s)
\end{aligned}
$$

$$
\begin{gathered}
=-\sum_{s=k-d(k)}^{k-d_{m}-1}\left\{\gamma^{k-s} z^{T}(s) R_{2} z(s)\right\} \\
-\sum_{s=k-d_{M}}^{k-d(k)-1} \gamma^{k-s} z^{T}(s) R_{2} z(s) .
\end{gathered}
$$

So (24) could be

$$
\begin{align*}
\tilde{\Delta} V_{3}(k) \leq & z^{T}(k)\left(\left(d_{M}-d_{m}\right) R_{2}+d_{m} R_{1}+d_{M} R_{3}\right) z(k) \\
& -\sum_{s=k-d_{m}}^{k-1} \gamma^{k-s} z^{T}(s)\left(R_{1}+R_{3}\right) z(s) \\
& -\sum_{s=k-d(k)}^{k-d_{m}-1} \gamma^{k-s} z^{T}(s)\left(R_{2}+R_{3}\right) z(s) \\
& -\sum_{s=k-d_{M}}^{k-d(k)-1} \gamma^{k-s} z^{T}(s) R_{2} z(s) \tag{26}
\end{align*}
$$

From Lemma 4, we have

$$
\begin{align*}
& -\sum_{s=k-d_{m}}^{k-1} \gamma^{k-s} z^{T}(s)\left(R_{1}+R_{3}\right) z(s) \\
& \quad \leq \frac{\gamma^{d_{m}}}{d_{m}} \xi_{1}^{T}(k)\left[\begin{array}{cc}
-R_{1}-R_{3} & R_{1}+R_{3} \\
* & -R_{1}-R_{3}
\end{array}\right] \xi_{1}(k) \\
& -\sum_{s=k-d(k)}^{k-d_{m}-1} \gamma^{k-s} z^{T}(s)\left(R_{2}+R_{3}\right) z(s) \\
& \quad \leq \frac{\gamma^{d_{M}}}{d_{M}-d_{m}} \xi_{2}^{T}(k)\left[\begin{array}{cc}
-R_{2}-R_{3} & R_{2}+R_{3} \\
* & -R_{2}-R_{3}
\end{array}\right] \xi_{2}(k), \\
& -\sum_{s=k-d_{M}}^{k-d(k)-1} \gamma^{k-s} z^{T}(s) R_{2} z(s) \\
& \quad \leq \frac{\gamma^{d_{M}}}{d_{M}-d_{m}} \xi_{3}^{T}(k)\left[\begin{array}{cc}
-R_{2} & R_{2} \\
* & -R_{2}
\end{array}\right] \xi_{3}(k) \tag{27}
\end{align*}
$$

where

$$
\begin{gather*}
\xi_{1}^{T}(k)=\left[\begin{array}{ll}
x^{T}(k) & x^{T}\left(k-d_{m}\right)
\end{array}\right], \\
\xi_{2}^{T}(k)=\left[\begin{array}{lll}
x^{T}\left(k-d_{m}\right) & x^{T}(k-d(k))
\end{array}\right],  \tag{28}\\
\xi_{3}^{T}(k)=\left[\begin{array}{ll}
x^{T}(k-d(k)) & x^{T}\left(k-d_{M}\right)
\end{array}\right] .
\end{gather*}
$$

Combining (21)-(27), it yields

$$
\begin{align*}
& V(k+1)-\gamma V(k) \\
& \quad \leq \xi^{T}(k) \\
& \quad \times\left[\begin{array}{cccc}
\phi_{11}+A_{i}^{T} W A_{i} & \phi_{12}+A_{i}^{T} W A_{i} & \phi_{13} & 0 \\
* & \phi_{22}+B_{i}^{T} W B_{i} & \phi_{23} & \phi_{24} \\
* & * & \phi_{33} & 0 \\
* & * & * & \phi_{44}
\end{array}\right] \xi(k) \\
& \quad+x^{T}(k) Y_{i} x(k), \tag{29}
\end{align*}
$$

where

$$
\xi^{T}(k)=\left[\begin{array}{lll}
x^{T}(k) & x^{T}(k-d(k)) & x^{T}\left(k-d_{m}\right) \tag{30}
\end{array} x^{T}\left(k-d_{M}\right)\right] .
$$

By Lemma 3 and condition (15), we know that the system of matrices $Y_{i}=A_{i}^{T} P A_{i}-U, i \in \mho$ is strictly complete, and the sets $\Omega_{i}$ and $\bar{\Omega}_{i}$ by (5) and (6) are well defined such that

$$
\begin{gather*}
\bigcup_{i=1}^{N} \Omega_{i}=R^{n} \backslash\{0\}, \quad \bigcup_{i=1}^{N} \mid \bar{\Omega}_{i}=R^{n} \backslash\{0\},  \tag{31}\\
\bar{\Omega}_{i} \cap \bar{\Omega}_{j}=\Phi, \quad i \neq j .
\end{gather*}
$$

Therefore, for any $x(k) \in R^{n}, k>0$, there always exists an $i \in\{1,2, \ldots, N\}$ such that $x(k) \in \bar{\Omega}_{i}$. Choosing the switching
rule (12), by Schur's complement of [32] with condition (14), leads to

$$
\begin{equation*}
V(k+1)-\gamma V(k) \leq 0 \Longleftrightarrow V(k) \leq \gamma^{k} V(0) \tag{32}
\end{equation*}
$$

By the system LK function (17), there always exist two positive constants $c_{1}, c_{2}$ such that

$$
\begin{equation*}
c_{1}\|x(k)\|^{2} \leq V(k), \quad V(0) \leq c_{2}\|x(0)\|_{s}^{2} . \tag{33}
\end{equation*}
$$

Here

$$
\begin{gather*}
c_{1}=\lambda_{\min }(P) \\
c_{2}=\lambda_{\max }(P)+\sum_{j=1}^{3}\left(\lambda_{\max }\left(Q_{j}\right)+\lambda_{\max }\left(R_{j}\right)\right) \tag{34}
\end{gather*}
$$

From (32) and (33), one obtains

$$
\begin{equation*}
\|x(k)\| \leq \sqrt{\frac{c_{2}}{c_{1}}} \gamma^{k / 2}\|x(0)\|_{s} . \tag{35}
\end{equation*}
$$

By Definition 1, we know that the system (1) is exponentially stable with decay rate $\lambda=\sqrt{\gamma}$. This completes the proof.

In order to determine the exponentially stability with $H_{\infty}$ performance $\kappa$ of system (1), we need to introduce the following definition.

Definition 7 (see [30]). Consider system (1) with the switching signal in (12) and the following conditions.
(i) With $w(k)=0$, the system (1) is exponentially stable with convergence rate $0<\gamma<1$.
(ii) With zero initial conditions, the signals $w(k)$ and $y(k)$ are bounded by

$$
\begin{equation*}
\sum_{k=0}^{\infty} \gamma^{-2 k} y^{T}(k) y(k) \leq \kappa^{2} \sum_{k=0}^{\infty} \gamma^{-2 k} w^{T}(k) w(k) \tag{36}
\end{equation*}
$$

for all $w \in L_{2}(\gamma, 0, \infty), w \neq 0$ for constants $\kappa>0$ and $0<\gamma<1$. In the above conditions, the system (1) is exponentially stabilizable with $H_{\infty}$ performance $\kappa$ and convergence rate $\gamma$ by switching signal in (12).

Theorem 8. For some constants $\gamma \in(0,1]$ and $0 \leq \alpha_{i} \leq 1$, $\sum_{i=1}^{N} \alpha_{i}=1(i \in \mho)$, if there exist positive definite symmetric
matrices $P, U, Q_{j}, R_{j}(j=1,2,3)$ such that (15) and the following LMIs hold:

$$
\left[\begin{array}{ccccccc}
\phi_{11} & \phi_{12} & \phi_{13} & 0 & \tilde{\phi}_{15} & C_{i}^{T} & \phi_{15}  \tag{37}\\
* & \phi_{22} & \phi_{23} & \phi_{24} & \tilde{\phi}_{25} & C_{d i}^{T} & \phi_{25} \\
* & * & \phi_{33} & 0 & 0 & 0 & 0 \\
* & * & * & \phi_{44} & 0 & 0 & 0 \\
* & * & * & * & \tilde{\phi}_{55} & D_{i}^{T} & B_{i}^{T} W^{T} \\
* & * & * & * & * & -I & 0 \\
* & * & * & * & * & * & -W
\end{array}\right]<0,
$$

then the system (13) with time-varying delay satisfying (4) is globally exponentially stable with convergence rate $\lambda=\sqrt{\gamma}$ and $H_{\infty}$ performance $\kappa$ by the switching signal designed by (12).

Here

$$
\begin{gather*}
\tilde{\phi}_{15}=A_{i}^{T} P B_{i} \\
\widetilde{\phi}_{25}=A_{d i}^{T} P B_{i}  \tag{38}\\
\widetilde{\phi}_{55}=B_{i}^{T} P B_{i}-\kappa^{2}
\end{gather*}
$$

Proof. The proof is similar to that of Theorem 6. From Theorem 6, one can easily obtain

$$
\begin{align*}
& V(k+1)-\gamma V(k)+y^{T}(k) y(k)-\kappa^{2} w^{T}(k) w(k) \\
& \quad \leq \eta_{1}^{T}(k)\left[\begin{array}{ccccc}
\phi_{11} & \phi_{12} & \phi_{13} & 0 & \tilde{\phi}_{15} \\
* & \phi_{22} & \phi_{23} & \phi_{24} & \widetilde{\phi}_{25} \\
* & * & \phi_{33} & 0 & 0 \\
* & * & * & \phi_{44} & 0 \\
* & * & * & * & \tilde{\phi}_{55}
\end{array}\right] \eta_{1}(k)  \tag{39}\\
& \quad+y^{T}(k) y(k)+z^{T}(k) W z(k)
\end{align*}
$$

where

$$
\eta_{1}^{T}(k)=\left[\begin{array}{llll}
x^{T}(k) & x^{T}(k-d(k)) & x^{T}\left(k-d_{m}\right) & x^{T}\left(k-d_{M}\right)  \tag{40}\\
w^{T}(k)
\end{array}\right] .
$$

Combining (37) and (15), by Schur's complement, we have

$$
\begin{equation*}
V(k+1)-\gamma V(k)+y^{T}(k) y(k)-\kappa^{2} w^{T}(k) w(k) \leq 0 . \tag{41}
\end{equation*}
$$

By Definition 7, the system (13) with time-varying delay satisfying (4) is globally exponentially stable with convergence rate $\lambda=\sqrt{\gamma}$ and $H_{\infty}$ performance $\kappa$ by the switching signal designed by (12).

Now, we extend Theorems 6 and 8 to obtain the corresponding results for uncertain switched systems (1). Set

$$
\begin{equation*}
p(k, i)=\Delta_{i}(k)\left(N_{1 i} x(k)+N_{2 i} x(k-d(k))+N_{3 i} w(k)\right) . \tag{42}
\end{equation*}
$$

Combining (3), we have

$$
\begin{align*}
p^{T}(k) p(k) \leq & \left(N_{1 i} x(k)+N_{2 i} x(k-d(k))+N_{3 i} w(k)\right)^{T} \\
& \times\left(N_{1 i} x(k)+N_{2 i} x(k-d(k))+N_{3 i} w(k)\right) . \tag{43}
\end{align*}
$$

Then, for any $\varepsilon>0$, the following inequalities hold:

$$
\begin{align*}
& \varepsilon\left(N_{1 i} x(k)+N_{2 i} x(k-d(k))+N_{3 i} w(k)\right)^{T} \\
& \quad \times\left(\left(N_{1 i} x(k)+N_{2 i} x(k-d(k))+N_{3 i} w(k)\right)\right.  \tag{44}\\
& \quad-\varepsilon p^{T}(k, i) p(k, i) \geq 0 .
\end{align*}
$$

By (2), the system (1) using (42) can be expressed as follows:

$$
\begin{align*}
x(k+1)= & A_{i} x(k)+A_{d i} x(k-d(k)) \\
& +B_{i} w(k)+M_{1 i} p(k, i),  \tag{45}\\
y(k)= & C_{i} x(k)+C_{d i} x(k-d(k)) \\
& +D_{i} w(k)+M_{2 i} p(k, i) .
\end{align*}
$$

The following theorem provides the robust exponential stability conditions for uncertain switched systems (45) with $w(k)=0$.

Theorem 9. For some constants $\gamma \in(0,1]$ and $0 \leq \alpha_{i} \leq 1$, $i \in U, \sum_{i=1}^{N} \alpha_{i}=1$, if there exist positive definite symmetric matrices $P, U, Q_{j}, R_{j}(j=1,2,3)$ such that (15) and the following LMIs hold:

$$
\left[\begin{array}{cccccc}
\widetilde{\phi}_{11} & \widetilde{\phi}_{12} & 0 & 0 & \widetilde{\phi}_{16} & \phi_{15}  \tag{46}\\
* & \widetilde{\phi}_{22} & \phi_{23} & \phi_{24} & \widetilde{\phi}_{26} & \phi_{25} \\
* & * & \phi_{33} & 0 & 0 & 0 \\
* & * & * & \phi_{44} & 0 & 0 \\
* & * & * & * & \widetilde{\phi}_{66} & M_{1 i}^{T} W^{T} \\
* & * & * & * & * & -W
\end{array}\right]<0,
$$

then the system (45) with time-varying delay satisfying (4) and $w(k)=0$ is globally exponentially stable with convergence rate $\lambda=\sqrt{\gamma}$ by the switching signal designed by (12).

Here

Proof. The result is carried out using the techniques employed for proving Theorem 6. Combining (44) with $w(k)=0$, one can obtain the following

$$
\begin{align*}
& V(k+1)-\gamma V(k) \\
& \quad \leq z^{T}(k) W z(k) \\
& \quad+\eta_{2}^{T}(k)\left[\begin{array}{ccccc}
\tilde{\phi}_{11} & \tilde{\phi}_{12} & 0 & 0 & \tilde{\phi}_{16} \\
* & \widetilde{\phi}_{22} & \phi_{23} & \phi_{24} & \widetilde{\phi}_{26} \\
* & * & \phi_{33} & 0 & 0 \\
* & * & * & \phi_{44} & 0 \\
* & * & * & * & \tilde{\phi}_{66}
\end{array}\right] \eta_{2}^{T}(k), \tag{48}
\end{align*}
$$

where

$$
\begin{gather*}
\widetilde{\phi}_{11}=\phi_{11}+\varepsilon N_{1 i}^{T} N_{1 i}, \\
\tilde{\phi}_{12}=\phi_{12}+\varepsilon N_{1 i}^{T} N_{2 i},  \tag{47}\\
\widetilde{\phi}_{16}=A_{i}^{T} P M_{1 i}, \\
\widetilde{\phi}_{22}=\phi_{22}+\varepsilon N_{2 i}^{T} N_{2 i}, \\
\widetilde{\phi}_{26}=A_{d i}^{T} P M_{1 i}, \\
\widetilde{\phi}_{66}=M_{1 i}^{T} P M_{1 i}-\varepsilon .
\end{gather*}
$$

$$
\begin{equation*}
\eta_{2}^{T}(k)=\left[x^{T}(k) x^{T}(k-d(k)) x^{T}\left(k-d_{M}\right) x^{T}\left(k-d_{M}\right) p^{T}(k, i)\right] . \tag{49}
\end{equation*}
$$

From (46) and (15), by Schur Complement, we have $V(k+1)-$ $\gamma V(k) \leq 0$.

Now, we are going to analyze the $H_{\infty}$ performance for the uncertain system (1). The main result is given as follows.

Theorem 10. For some constants $\gamma \in(0,1]$ and $0 \leq \alpha_{i} \leq 1$, $i \in U, \sum_{i=1}^{N} \alpha_{i}=1$, if there exist positive definite symmetric matrices $P, U, Q_{j}, R_{j}(j=1,2,3)$ such that (15) and the following LMIs hold

$$
\left[\begin{array}{cccccccc}
\tilde{\phi}_{11} & \tilde{\phi}_{12} & \phi_{13} & 0 & \bar{\phi}_{15} & \tilde{\phi}_{16} & C_{i}^{T} & \phi_{15}  \tag{52}\\
* & \widetilde{\phi}_{22} & \phi_{23} & \phi_{24} & \bar{\phi}_{25} & \widetilde{\phi}_{26} & C_{d i}^{T} & \phi_{25} \\
* & * & \phi_{33} & 0 & 0 & 0 & 0 & 0 \\
* & * & * & \phi_{44} & 0 & 0 & 0 & 0 \\
* & * & * & * & \bar{\phi}_{55} & \widetilde{\phi}_{56} & D_{i}^{T} & \widetilde{\phi}_{58} \\
* & * & * & * & * & \widetilde{\phi}_{66} & M_{2 i}^{T} & \widetilde{\phi}_{68} \\
* & * & * & * & * & * & -I & 0 \\
* & * & * & * & * & * & * & -W
\end{array}\right]<0
$$

then the system (1) with time-varying delay satisfying (4) is globally exponentially stable with convergence rate $\lambda=\sqrt{\gamma}$ and $H_{\infty}$ performance $\kappa$ by the switching signal designed by (12).

Here

$$
\begin{gather*}
\bar{\phi}_{15}=\tilde{\phi}_{15}+\varepsilon N_{1 i}^{T} N_{3 i}, \quad \bar{\phi}_{25}=\tilde{\phi}_{25}+\varepsilon N_{2 i}^{T} N_{3 i}, \\
\tilde{\phi}_{58}=B_{i}^{T} W^{T}, \quad \bar{\phi}_{55}=\tilde{\phi}_{55}+\varepsilon N_{3 i}^{T} N_{3 i},  \tag{51}\\
\tilde{\phi}_{56}=B_{i}^{T} P M_{1 i}, \quad \widetilde{\phi}_{68}=M_{1 i}^{T} W^{T} .
\end{gather*}
$$

Proof. The result is carried out using the techniques employed for proving Theorems 6, 8 , and 9 . There exists matrix $\Theta$ satisfying

$$
\begin{aligned}
V(k & +1)-\gamma V(k)+y^{T}(k) y(k)-\kappa^{2} w^{T}(k) w(k) \\
& \leq \eta_{3}^{T}(k) \Theta \eta_{3}^{T}(k)+z^{T}(k) W z(k)+y^{T}(k) y(k),
\end{aligned}
$$

where

$$
\begin{equation*}
\eta_{3}^{T}(k)=\left[x^{T}(k) x^{T}(k-d(k)) x^{T}\left(k-d_{m}\right) x^{T}\left(k-d_{M}\right) w^{T}(k) p^{T}(k, i)\right] . \tag{53}
\end{equation*}
$$

Combining (50) and (15), by Schur Complement, we have $V(k+1)-\gamma V(k)+y^{T}(k) y(k)-\kappa^{2} w^{T}(k) w(k) \leq 0$.

Example 1 (see [27]). Consider system (13) with $w(k)=0$ and the following parameters:

## 4. Examples

In this section, we consider some numerical examples in $[26,27]$ to show that the proposed theorem in this paper provides less conservativeness with comparison to some existing results.

$$
\begin{array}{ll}
A_{1}=\left[\begin{array}{cc}
0.54 & 1.02 \\
-0.17 & -0.31
\end{array}\right], & A_{d 1}=\left[\begin{array}{cc}
0.18 & 0.36 \\
-0.06 & -0.12
\end{array}\right],  \tag{54}\\
A_{2}=\left[\begin{array}{cc}
-0.01 & -0.06 \\
0.01 & 0.04
\end{array}\right], & A_{d 2}=\left[\begin{array}{cc}
0.11 & 0.18 \\
-0.03 & -0.04
\end{array}\right] .
\end{array}
$$



Figure 1: The switching regions.

Table 1: Allowable delay upper bound for different decay rate with $d_{m}=1, \alpha_{1}=\alpha_{2}=0.5$.

|  | $\lambda=0.8$ | $\lambda=1$ |
| :--- | :---: | :---: |
| $[26]$ | - | $d_{M}=2$ |
| [27] | $d_{M}=8$ | $d_{M}=182$ |
| Theorem 6 | $d_{M}=10$ | $d_{M}=\infty$ |

Assume the minimum delay bound $d_{m}=1$. By Theorem 6 , via solving LMIs (14) and (15) with $\lambda=0.8$ and $\alpha_{1}=\alpha_{2}=0.5$, we have $d_{M}=10$ and

$$
P=\left[\begin{array}{cc}
13.2019 & 30.1428  \tag{55}\\
* & 72.992
\end{array}\right], \quad U=\left[\begin{array}{cc}
0.4493 & 1.1242 \\
* & 3.098
\end{array}\right] .
$$

From (9) and (11), one can obtain the switching regions

$$
\begin{gather*}
\bar{\Omega}_{1}=\left\{\left[\begin{array}{ll}
x_{1} & x_{2}
\end{array}\right]^{T} \in R:-0.0244 x_{1}^{2}-0.5572 x_{1} x_{2}\right. \\
\left.-1.4106 x_{2}^{2}<0\right\}, \quad \bar{\Omega}_{2}=R^{2} \backslash \bar{\Omega}_{1} \tag{56}
\end{gather*}
$$

The switching regions $\bar{\Omega}_{1}$ and $\bar{\Omega}_{2}$ of the system are shown in Figure 1. Select the switching signal by

$$
\begin{equation*}
\sigma(x(k))=i, \quad x(k) \in \bar{\Omega}_{i}, \quad i=1,2 . \tag{57}
\end{equation*}
$$

With the initial condition $\phi(\theta)=[1-1]^{T}, \theta=-10,-9, \ldots, 0$, and $d(k)=10$, the state responses of system are shown in Figure 2.

In order to show the improvement of our results over other recent results, some comparisons are made in Table 1. From Table 1, one can see that the results of this paper provide a larger allowable upper bound for time delay to guarantee the asymptotical or exponentially stability of system (1) with (54) by switching signal in (12).


Figure 2: The state response.

Example 2 (see [27]). Consider system (13) with $w(k)=0$ and the following parameters:

$$
\begin{array}{ll}
A_{1}=\left[\begin{array}{cc}
1.01 & 0.1 \\
0 & 0.1
\end{array}\right], & A_{d 1}=\left[\begin{array}{cc}
-0.1 & 0 \\
-0.1 & -0.1
\end{array}\right], \\
A_{2}=\left[\begin{array}{cc}
0.1 & 0 \\
0.1 & 1.02
\end{array}\right], & A_{d 2}=\left[\begin{array}{cc}
0.12 & 0 \\
0.11 & 0.11
\end{array}\right] . \tag{58}
\end{array}
$$

Let $\alpha_{1}=\alpha_{2}=0.5$, the minimum delay bound $d_{m}=1$, and decay rate $\lambda=1$. The delay upper bound obtained by [27] is 3 . However, our result obtained by the application of Theorem 6 could be infinite to guarantee the asymptotic stability of system (13) with (58) by switching signal in (12). The switching regions can be designed in the same way as in Example 1.

Example 3. Consider the discrete-time switched system (13) with $w(k)=0$ and the following parameters:

$$
\begin{array}{ll}
A_{1}=\left[\begin{array}{cc}
1.21 & 0.1 \\
0 & 0.1
\end{array}\right], & A_{d 1}=\left[\begin{array}{cc}
-0.1 & 0 \\
-0.1 & -0.1
\end{array}\right], \\
A_{2}=\left[\begin{array}{cc}
0.1 & 0 \\
0.1 & 1.02
\end{array}\right], & A_{d 2}=\left[\begin{array}{cc}
0.12 & 0 \\
0.11 & 0.11
\end{array}\right] . \tag{59}
\end{array}
$$

One can easily find that both the above subsystems are unstable. The state responses of the two subsystems are shown in Figures 3 and 4. Let $d_{m}=1$. By Theorem 6, via solving LMIs (14) and (15) with $\lambda=1$ and $\alpha_{1}=\alpha_{2}=0.5$, we have $d_{M}=5$ and

$$
\begin{align*}
& P=1.0 e+003 *\left[\begin{array}{cc}
3.5486 & 0.1556 \\
* & 0.5634
\end{array}\right] \\
& U=1.0 e+003 *\left[\begin{array}{cc}
2.6204 & 0.2602 \\
* & 0.3184
\end{array}\right] \tag{60}
\end{align*}
$$



Figure 3: State responses of the subsystems 1.

From (9) and (11), one can obtain the switching regions

$$
\begin{array}{r}
\bar{\Omega}_{i}^{1}=\left\{\left[\begin{array}{ll}
x_{1} & x_{2}
\end{array}\right]^{T} \in R: 2.5751 x_{1}^{2}+0.376 x_{1} x_{2}\right. \\
 \tag{61}\\
\left.-0.2741 x_{2}^{2}<0\right\}, \quad \bar{\Omega}_{2}=R^{2} \backslash \bar{\Omega}_{1}
\end{array}
$$

Select the switching signal by

$$
\begin{equation*}
\sigma(x(k))=i, \quad x(k) \in \bar{\Omega}_{i}, \quad i=1,2 . \tag{62}
\end{equation*}
$$

With time-varying delay $d(k)=5$ and the initial condition $\phi(\theta)=\left[\begin{array}{ll}1 & -1\end{array}\right]^{T}, \theta=-5,-4, \ldots, 0$. The state responses of system are shown in Figure 5. It can be seen from Figure 5 that the designed switching law is effective although all subsystems are unstable. However, with the same parameters and $d_{M}=5$, the results in [27] cannot find any feasible solution to guarantee the asymptotic stability of system (13) with (59).

Example 4. Consider system (45) with the following parameters:

$$
\begin{array}{cc}
A_{1}=\left[\begin{array}{cc}
1 & 0.01 \\
0 & 0
\end{array}\right], & A_{d 1}=\left[\begin{array}{cc}
0 & 0.1 \\
0 & -0.1
\end{array}\right], \\
B_{1}=\left[\begin{array}{cc}
0.1 & 0 \\
0 & 0.1
\end{array}\right], & C_{1}=\left[\begin{array}{cc}
0.01 & 0 \\
0 & 0.08
\end{array}\right], \\
C_{d 1}=\left[\begin{array}{cc}
0.02 & 0 \\
0 & 0.01
\end{array}\right], & D_{1}=\left[\begin{array}{cc}
0.1 & 0.1 \\
0 & 0.2
\end{array}\right], \\
A_{2}=\left[\begin{array}{cc}
1 & 0.01 \\
0.01 & 1.01
\end{array}\right], & A_{d 2}=\left[\begin{array}{cc}
0.1 & 0 \\
0.1 & 0.1
\end{array}\right],
\end{array}
$$



Figure 4: State responses of the subsystems 2.


Figure 5: State responses of the switched systems 2.

$$
\begin{gather*}
B_{2}=\left[\begin{array}{cc}
0.2 & 0 \\
0.1 & 0.1
\end{array}\right], \quad C_{2}=\left[\begin{array}{cc}
0.03 & 0 \\
0 & 0.05
\end{array}\right], \\
C_{d 2}=\left[\begin{array}{cc}
0.02 & 0 \\
0 & 0.02
\end{array}\right], \quad D_{2}=\left[\begin{array}{cc}
0.2 & 0 \\
0 & 0.1
\end{array}\right], \\
M_{2}=\left[\begin{array}{cc}
0.1 & 0 \\
0 & 0.05
\end{array}\right], N_{2}=\left[\begin{array}{cc}
0.01 & 0 \\
0 & 0.02
\end{array}\right], \\
M_{1}=B_{1}, \quad N_{1}=C_{d 1}, \quad N_{3}=0.1 \times B_{1} . \tag{63}
\end{gather*}
$$

For a given $d_{m}=1, \gamma=0.98, \kappa=1.53$, and $\alpha_{1}=\alpha_{2}=0.5$. We compute the upper delay bound $d_{M}=23$ such that the uncertain system is exponential stability by applying Theorem 10. However, with the same parameters and $d_{M}=23$, the results in [30] cannot find any feasible solution to guarantee
the exponentially stable with $H_{\infty}$ performance $\kappa=1.53$ of system (45). Via solving LMIs (50) and (15), we have

$$
P=\left[\begin{array}{cc}
0.6836 & 0.0942  \tag{64}\\
* & 0.6615
\end{array}\right], \quad U=\left[\begin{array}{ll}
0.3424 & 0.0067 \\
0.0067 & 0.3387
\end{array}\right] .
$$

From (9) and (11), one can obtain the switching regions

$$
\begin{gather*}
\bar{\Omega}_{1}=\left\{\left[\begin{array}{ll}
x_{1} & x_{2}
\end{array}\right]^{T} \in R: 0.3413 x_{1}^{2}+0.0002 x_{1} x_{2}\right. \\
\left.-0.3386 x_{2}^{2}<0\right\}, \quad \bar{\Omega}_{2}=R^{2} \backslash \bar{\Omega}_{1} . \tag{65}
\end{gather*}
$$

Select the switching signal by $\sigma(x(k))=i, x(k) \in \bar{\Omega}_{i}, i=1,2$. On the other side, by setting the upper delay bound $d_{M}=3$, we have the $H_{\infty}$ performance $\kappa=0.275$.

## 5. Conclusions

By using improved discrete LK function combined with LMIs technique, in this paper, we propose new criteria for the exponential stability with $H_{\infty}$ performance of uncertain switched linear discrete-time systems with interval time-varying delay. If there is a feasible solution for the proposed LMIs conditions under some given upper bounds of delay, the switching law can be designed and the exponential stability of the systems can be achieved. Finally, the obtained results are shown to be less conservative than previous ones via the numerical examples.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China (nos. 61104064 and 61174038).

## References

[1] Z. D. Sun and S. S. Ge, Switched Linear Systems-Control and Design, Springer, New York, NY, USA, 2004.
[2] D. Liberzon, "Basic problems in stability and design of switched systems," IEEE Control Systems Magazine, vol. 19, no. 5, pp. 5970, 1999.
[3] D. Liberzon, Switching in Systems and Control, Birkhäuser, Boston, Mass, USA, 2003.
[4] A. V. Savkin and R. J. Evans, Hybrid Dynamical Systems: Controller and Sensor Switching Problems, Springer, New York, NY, USA, 2002.
[5] M. Wicks, P. Peleties, and R. deCarlo, "Switched controller synthesis for the quadratic stabilization of a pair of unstable linear systems," European Journal of Control, vol. 4, no. 2, pp. 140-147, 1998.
[6] H. Lin and P. J. Antsaklis, "Stability and stabilizability of switched linear systems: a survey of recent results," IEEE Transactions on Automatic Control, vol. 54, no. 2, pp. 308-322, 2009.
[7] A. S. Morse, "Supervisory control of families of linear setpoint controllers-part 1: exact matching," IEEE Transactions on Automatic Control, vol. 41, no. 10, pp. 1413-1431, 1996.
[8] W. H. Chen and W. X. Zheng, "Delay-independent minimum dwell time for exponential stability of uncertain switched delay
systems," IEEE Transactions on Automatic Control, vol. 55, no. 10, pp. 2406-2413, 2010.
[9] L. Zhang and P. Shi, "Stability $L_{2}$-gain and asynchronous $H_{\infty}$ control of discrete-time switched systems with average dwell time," IEEE Transactions on Automatic Control, vol. 54, no. 9, pp. 2192-2199, 2009.
[10] G. Zhai, B. Hu, K. Yasuda, and A. N. Michel, "Stability analysis of switched systems with stable and unstable subsystems: an average dwell time approach," in Proceedings of the American Control Conference, pp. 200-204, 2000.
[11] X. D. Zhao, L. X. Zhang, P. Shi, and M. Liu, "Stability and stabilization of switched linear systems with mode-dependent average dwell time," IEEE Transactions on Automatic Control, vol. 57, no. 7, pp. 1809-1815, 2012.
[12] X. D. Zhao, H. Liu, and Z. H. Wang, "Weighted $H_{\infty}$ performance analysis of switched linear systems with modedependent average dwell time," International Journal of System Science, vol. 44, no. 11, pp. 2130-2139, 2013.
[13] M. Johansson and A. Rantzer, "Computation of piecewise quadratic Lyapunov functions for hybrid systems," IEEE Transactions on Automatic Control, vol. 43, no. 4, pp. 555-559, 1998.
[14] A. Bacciotti, "Stabilization by means of state space depending switching rules," Systems \& Control Letters, vol. 53, no. 3-4, pp. 195-201, 2004.
[15] J. K. Hale and S. M. Verduyn Lunel, Introduction to FunctionalDifferential Equations, Springer, New York, NY, USA, 1993.
[16] J. N. Li, H. Y. Su, Z. Z. Wu, and J. Chu, "Robust stabilization for discrete-time nonlinear singular systems with mixed time delays," Asian Journal of Control, vol. 15, no. 6, pp. 1-11, 2013.
[17] Z. G. Feng, J. Lam, and H. J. Gao, "Delay-dependent robust $H_{\infty}$ controller synthesis for discrete singular delay systems," International Journal of Robust and Nonlinear Control, vol. 21, no. 16, pp. 1880-1902, 2011.
[18] Q.-L. Han, "A discrete delay decomposition approach to stability of linear retarded and neutral systems," Automatica, vol. 45, no. 2, pp. 517-524, 2009.
[19] Y. G. Sun, L. Wang, and G. Xie, "Delay-dependent robust stability and stabilization for discrete-time switched systems with mode-dependent time-varying delays," Applied Mathematics and Computation, vol. 180, no. 2, pp. 428-435, 2006.
[20] W. A. Zhang and L. Yu, "Stability analysis for discrete-time switched time-delay systems," Automatica, vol. 45, no. 10, pp. 2265-2271, 2009.
[21] C. Meyer, S. Schroder, and R. W. de Doncker, "Solid-state circuit breakers and current limiters for medium-voltage systems having distributed power systems," IEEE Transactions on Power Electronics, vol. 19, no. 5, pp. 1333-1340, 2004.
[22] P. Yan and H. Ozbay, "Stability analysis of switched time delay systems," SIAM Journal on Control and Optimization, vol. 47, no. 2, pp. 936-949, 2008.
[23] S. Kim, S. A. Campbell, and X. Z. Liu, "Stability of a class of linear switching systems with time delay," IEEE Transactions on Circuits and Systems, vol. 53, no. 2, pp. 384-393, 2006.
[24] X. M. Sun, W. Wang, G. P. Liu, and J. Zhao, "Stability analysis for linear switched systems with time varying delay," IEEE Transactions on Systems, Man and Cybernetics B, vol. 38, pp. 528-533, 2008.
[25] V. N. Phat, "Robust stability and stabilizability of uncertain linear hybrid systems with state delays," IEEE Transactions on Circuits Systems II-Express Briefs, vol. 52, no. 2, pp. 94-98, 2005.
[26] V. N. Phat and K. Ratchagit, "Stability and stabilization of switched linear discrete-time systems with interval timevarying delay," Nonlinear Analysis: Hybrid Systems, vol. 5, no. 4, pp. 605-612, 2011.
[27] C. H. Lien, K. W. Yu, H. C. Chang, L. Y. Chung, and J. D. Chen, "Switching signal design for exponential stability of discrete switched systems with interval time-varying delay," Journal of the Franklin Institute, vol. 349, no. 6, pp. 2182-2192, 2012.
[28] C. H. Lien, K. W. Yu, Y. J. Chung, H. C. Chang, and J. D. Chen, "Switching signal design for global exponential stability of uncertain switched nonlinear systems with time-varying delay," Nonlinear Analysis: Hybrid Systems, vol. 5, no. 1, pp. 10-19, 2011.
[29] Y. G. Sun, L. Wang, and G. Xie, "Delay-dependent robust stability and control for uncertain discrete-time switched systems with mode-dependent time delays," Applied Mathematics and Computation, vol. 187, no. 2, pp. 1228-1237, 2007.
[30] C. H. Lien, K. W. Yu, L. Y. Chung, and J. D. Chen, " $H_{\infty}$ performance for uncertain discrete switched systems with interval time-varying delay via switching signal design," Applied Mathematical Modelling, vol. 37, no. 4, pp. 2484-2494, 2013.
[31] F. Uhlig, "A recurring theorem about pairs of quadratic forms and extensions," Linear Algebra and Its Applications, vol. 25, pp. 219-237, 1979.
[32] S. P. Boyd, El Ghaoui, E. Feron, and V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory, SIAM, Philadelphia, Pa, USA, 1994.

# Stability Analysis of Networked Control Systems with Random Time Delays and Packet Dropouts Modeled by Markov Chains 

Li Qiu, ${ }^{1}$ Fengqi Yao, ${ }^{2}$ and Xiaopin Zhong ${ }^{1}$<br>${ }^{1}$ Shenzhen Key Laboratory of Urban Rail Transit, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China<br>${ }^{2}$ School of Electrical Engineering \& Information, Anhui University of Technology, Mảanshan 243000, China<br>Correspondence should be addressed to Xiaopin Zhong; xzhong@szu.edu.cn

Received 27 May 2013; Revised 30 July 2013; Accepted 31 July 2013
Academic Editor: Baocang Ding
Copyright © 2013 Li Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper investigates the stability analysis problem for a class of discrete-time networked control systems (NCSs) with random time delays and packet dropouts based on unified Markov jump model. The random time delays and packet dropouts existed in feedback communication link are modeled by two independent Markov chains; the resulting closed-loop system is described by a new Markovian jump linear system (MJLS) with Markov delays. Sufficient conditions of the stochastic stability for NCSs is obtained by constructing a novel Lyapunov functional, and the mode-dependent output feedback controller design method is presented based on linear matrix inequality (LMI) technique. A numerical example is given to illustrate the effectiveness of the proposed method.


## 1. Introduction

Networked control systems (NCSs) are a type of closedloop systems, in which the control loops are closed through communication networks. Compared with the traditional control systems, the use of the communication networks bring many advantages such as low cost, reduced weight, and simple installation and maintenance as well as high efficiency, flexibility, and reliability. Consequently, NCSs are applied in a broad range such as manufacturing plants, vehicles, aircrafts, spacecrafts, and remote surgery [1]. However, the communication networks in control loops also present some constraints such as time delays and packet dropouts due to limited bandwidth; quantization errors caused by hybrid nature of NCSs; variable sampling or transmission intervals due to multiple nodes; clock asynchronization among local and remote nodes; network security and safety and network security due to shared communication networks [2, 3]. It is generally known that any of these networked-induced communication imperfections and constraints can degrade closed-loop performance or, even worse, can harm closedloop stability of NCSs. Therefore, it is important to know how these effects influence the stability properties. Recently, some
important results of NCSs have been reported in the existing literature for instance, the discussions of packet dropouts [413], time delays [14-24], quantization [25], distributed synchronization [26], communication constraints [27], stability and controller design [28-33], both data quantizations and packet losses [34], both time delays and packet dropouts [3544], and output feedback control problem [19, 45, 46].

In NCSs, time delays and packet dropouts are two important issues. To study these issues, many efforts have been made for NCSs with time delays [14-24] and packet dropouts [4-13]; for more details review, please refer to the literature therein. However, both time delays and packet dropouts exist in NCSs by the insertion of communication network in the feedback control loop. Xie and Xia [35] studied the robust fault tolerant controller design for NCSs with fast varying delay and packet dropout. Zhang and Yu [36] presented a switched system model to describe the NCSs with both delay and packet dropout, and the state feedback stabilizing controllers are designed by augmenting technique. Yu and Shi [37] addressed the two-mode-dependent state feedback controller design in NCSs with time delays and packet dropouts by augmenting the state variable approach. Dong et al. [38] studied the robust $H_{\infty}$ filtering problem
for a class of uncertain nonlinear networked systems with both stochastic time-varying communication delays and packet dropouts. Jiang et al. [39] introduced the design of observer-based controller for NCSs with network induced delay and packet dropout. Li et al. [40] considered the observer-based fault detection problem for NCSs with long time delays and packet dropout by modeling the observers system as an uncertain switched system. Li et al. [41] studied the guaranteed const control of NCSs with the S-C packet dropouts and time delays. Wang and Yang [42] considered the problem of $H_{\infty}$ controller design for NCSs with time delay and packet dropout by applying the linear estimationbased time delay and packet dropout compensation method. Liu et al. [43] investigated the receding horizon $H_{\infty}$ control problem for a class NCSs with random delay and packet disordering by using the receding optimization principle. Qiu et al. [44] considered the state feedback control problem of NCS with both time delays and packet dropouts. In [19, 45, 46], the output feedback control problem of NCSs were investigated. To the best of the authors' knowledge, up to now, little attention has been paid to the study of NCSs with random time delays and packet dropouts based on Markov jump unified model, which motivates our investigation.

In this paper, we address the unified model and stability analysis problem of NCSs with the random time delays and packet dropouts under a Markovian jump linear system (MJLS) framework. The feedback communication link random time delays and packet dropouts are modeled by two independent Markov chains, the resulting closed-loop system is modeled as a new MJLS with Markov delays. Then, we give stability analysis and output feedback controller design which are for discrete-time NCSs with both time delays and packet dropouts by the Lyapunov stability theory and linear matrix inequality method.

Notations. In the sequel, if not explicit, matrices are assumed to have appropriate dimensions. $\mathbb{R}^{n}$ and $\mathbb{R}^{n \times m}$ denote, respectively, the $n$ dimensional Euclidean space and the set of all $n \times m$ real matrices. The notations $A>0$ and $A<0$ are used to denote the positive and negative definite matrix, respectively. $\operatorname{diag}\left(A_{1}, \ldots, A_{n}\right)$ refers to a $n \times n$ diagonal matrix with $A_{i}$ as its $i$ th diagonal entry. $I$ and 0 denote the identity matrix and zero matrix with appropriate dimensions, respectively. The superscript ${ }^{T}$ denotes the transpose for vectors or matrices. $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator. The symbol $*$ denotes blocks that are readily inferred by symmetry.

## 2. Problem Description

The framework of the system over a network medium is depicted in Figure 1. Considering the same assumption in [14], the sensor, the controller, and the actuator are timedriven and are connected over a network medium. Under the assumption, it is known that the controller updates at the instant $k$ will always use the most recent data; otherwise, it will maintain the old data. In the NCS as Figure 1, network


Figure 1: The structure of the NCS with random delays and/or packet dropouts.
induced time delays and packet dropouts exist in the feedback communication link.

The discrete-time plant with a time-varying controller is described as

$$
\begin{gather*}
x_{p}(k+1)=A x_{p}(k)+B u(k),  \tag{1}\\
y(k)=C x_{p}(k)
\end{gather*}
$$

where $x_{p}(k) \in \mathbb{R}^{n}$ is the system state, $u(k) \in \mathbb{R}^{m}$ is the control input, and $y(k) \in \mathbb{R}^{p}$ is measurable output. $A, B$, and $C$ are known real constant matrices with appropriate dimensions. Random time delays and packet dropouts exist in link from sensor-to-controller (S-C), as shown in Figure 1. Here, $\tau(k)$ represents the bounded random S-C time delays. One way to model the delays $\tau(k)$ is using the finite state Markov chain as shown in [17-19]. The main advantage of the Markov model is that the dependencies between the delays are taken into account since the current time delays in real networks are usually related with the previous delays [17]. In this paper $\tau(k)$ is modeled as a homogeneous Markov chain that take values in $\bar{S}_{2}=\left\{0,1, \ldots, s_{2}\right\} . S$ denotes the network switches between the S-C. $\alpha(k)(\alpha(k)=0,1)$ denotes the states of $S$. When $S$ is in state $\alpha(k)=0$, the packet is received successfully and the $\bar{y}(k)=y(k-\tau(k))$. Whereas when $S$ is in state $\alpha(k)=1$, the packet is lost and the switch output is held at the previous value $\bar{y}(k)=\bar{y}(k-1)$. The behavior of the S-C time delays and packet dropouts can be modeled as

$$
\begin{equation*}
\bar{y}(k)=(1-\alpha(k)) y(k-\tau(k))+\alpha(k) \bar{y}(k-1), \tag{2}
\end{equation*}
$$

where

$$
\alpha(k)= \begin{cases}0, & \text { if } S \text { is closed and the packet is received, }  \tag{3}\\ 1, & \text { if } S \text { is open and the packet is lost. }\end{cases}
$$

Considering the mode-dependent output feedback controller:

$$
\begin{equation*}
u(k)=K(\alpha(k), \tau(k)) \bar{y}(k) \tag{4}
\end{equation*}
$$

where $K(\alpha(k), \tau(k))$ is the output feedback controller gain.

Let $x(k)=\left[\begin{array}{ll}x_{p}(k)^{T} & \bar{y}(k-1)^{T}\end{array}\right]^{T}$ be the augmented state vector. Under the control (4), the closed-loop system of (1) is

$$
\begin{align*}
x(k+1) & =\bar{A}(\alpha(k)) x(k)+\bar{B}(\alpha(k)) H x(k-\tau(k)),  \tag{5}\\
x(k) & =\varphi(k), \quad k=-\tau_{\max },-\tau_{\max }+1, \ldots, 0,
\end{align*}
$$

where $\bar{A}(\alpha(k))=\left[\begin{array}{c}A \alpha(k) B K(\alpha(k) \tau(k)) \\ 0 \\ \alpha(k) I\end{array}\right], \bar{B}(\alpha(k))=$ $[\underset{(1-\alpha(k)) C}{(1-\alpha(k)),(k)) C}], H=\left[\begin{array}{ll}I & 0\end{array}\right], \tau_{\max }=\max \{\tau(k)\}$, and $\varphi(k)$ is the initial condition of $x(k)$.

In system (5), $\{\alpha(k), k \in \mathbb{Z}\}$ and $\{\tau(k), k \in \mathbb{Z}\}$ are two independent discrete-time homogeneous Markov chains taking value in a finite set $\bar{S}_{1}=\{0,1\}$ and $\bar{S}_{2}=\left\{0,1, \ldots, s_{2}\right\}$ with transition probabilities:

$$
\begin{gather*}
\operatorname{Pr}\{\alpha(k+1)=j \mid \alpha(k)=i\}=\pi_{i j}, \quad \pi_{i}=\operatorname{Pr}\left(\pi_{0}=i\right), \\
\operatorname{Pr}\{\tau(k+1)=n \mid \tau(k)=m\}=\lambda_{m n}, \\
\lambda_{m}=\operatorname{Pr}\left(\lambda_{0}=m\right), \tag{6}
\end{gather*}
$$

where $\pi_{i j} \geq 0$ and $\lambda_{m n} \geq 0$ for all $i, j \in \bar{S}_{1}, m, n \in \bar{S}_{2}$ and

$$
\begin{equation*}
\sum_{j=0}^{1} \pi_{i j}=1, \quad \sum_{n=0}^{s_{2}} \lambda_{m n}=1 \tag{7}
\end{equation*}
$$

For $\alpha(k)=i, i \in \bar{S}_{1}$, when $\alpha(k)$ in mode $i=0$ and $i=1$, the $\alpha(k)$ in (5) take value $\alpha(k)=0$ and $\alpha(k)=$ 1, respectively. $\bar{A}(\alpha(k))$ and $\bar{B}(\alpha(k))$ are known constant matrices of appropriate dimensions.

Remark 1. The closed-loop system (5) is a MJLS with two Markov chains, which describe the behavior of the S-C time delays and packet dropouts, respectively. This enables us to analyze and synthesize such NCSs by applying MJLS theory. Note that modeling the S-C time delays and packet dropouts simultaneously in NCSs based on unified Markov jump model has not been done in the literature.

Definition 2 (see [19]). The system in (5) is stochastically stable if for every finite $x_{0}=x(0)$, initial mode $\alpha_{0}=\alpha(0) \in$ $\bar{S}_{1}$, and $\tau_{0}=\tau(0) \in \bar{S}_{2}$, there exists a finite $\mathscr{W}>0$ such that the following holds:

$$
\begin{equation*}
\mathbb{E}\left\{\sum_{k=0}^{\infty}\|x(k)\|^{2} \mid x_{0}, \alpha_{0}, \tau_{0}\right\}<x_{0}^{T} \mathscr{W} x_{0} . \tag{8}
\end{equation*}
$$

## 3. Main Results

By applying a new Lyapunov functional, sufficient conditions for the stochastic stability and synthesis of the modedependent output feedback controller design for system (5) will be established in this section.

Theorem 3. For system (5), given random but bounded scalar $\tau(k) \in\left[\begin{array}{ll}\tau_{\min } & \tau_{\text {max }}\end{array}\right]$, if for each mode $i \in \bar{S}_{1}, m \in \bar{S}_{2}$, there
exist matrices $P_{i, m}>0, Q_{1}>0, Q_{2}>0, Q_{3}>0, R_{1}>0$, and $R_{2}>0$ such that the following matrix inequalities:

$$
\left[\begin{array}{cccc}
\Xi_{1} & \Xi_{2} & \Xi_{3} & \Xi_{4}  \tag{9}\\
* & -\bar{P}_{i, m} & 0 & 0 \\
* & * & -R_{1} & 0 \\
* & * & * & -R_{2}
\end{array}\right]<0,
$$

where

$$
\begin{align*}
& \Xi_{1}=\left[\begin{array}{cccc}
\Pi_{i, m} & H^{T} R_{1} & 0 & 0 \\
R_{1} H & -Q_{3}-2 R_{1}-2 R_{2} & R_{2} & R_{1}+R_{2} \\
0 & R_{2} & -Q_{2}-R_{2} & 0 \\
0 & R_{1}+R_{2} & 0 & -Q_{1}-R_{1}-R_{2}
\end{array}\right] \text {, } \\
& \Xi_{2}=\left[\begin{array}{llll}
\bar{P}_{i, m} \bar{A}_{i} & \bar{P}_{i, m} \bar{B}_{i} & 0 & 0
\end{array}\right]^{T}, \\
& \Xi_{3}=\left[\begin{array}{llll}
\tau_{\max } R_{1} H\left(\bar{A}_{i}-I\right) & \tau_{\max } R_{1} H \bar{B}_{i} & 0 & 0
\end{array}\right]^{T}, \\
& \Xi_{4}=\left[\begin{array}{llll}
\widetilde{\tau} R_{2} H\left(\bar{A}_{i}-I\right) & \widetilde{\tau} R_{2} H \bar{B}_{i} & 0 & 0
\end{array}\right]^{T}, \\
& \tilde{\tau}=\tau_{\text {max }}-\tau_{\text {min }}, \\
& \bar{P}_{i, m}=\sum_{j=0}^{1} \sum_{n=0}^{s_{2}} \pi_{i j} \lambda_{m n} P_{j, n}, \\
& \Pi_{i, m}=-P_{i, m}+H^{T}\left(Q_{1}+Q_{2}\right) H \\
& +\left(\tau_{\max }-\tau_{\text {min }}+1\right) H^{T} Q_{3} H-H^{T} R_{1} H, \\
& \bar{A}_{i}=\left[\begin{array}{cc}
A & i B K(i, m) \\
0 & i I
\end{array}\right], \\
& \bar{B}_{i}=\left[\begin{array}{c}
(1-i) B K(i, m) C \\
(1-i) C
\end{array}\right], \tag{10}
\end{align*}
$$

and $H$ is defined in (5).
Hold for all $i, j \in \bar{S}_{1}$ and $m, n \in \bar{S}_{2}$; then system (5) is stochastically stable.

Proof. For the closed-loop system (5), stochastic Lyapunov functional is constructed as follows:

$$
\begin{align*}
V(x(k), \alpha(k), \tau(k)) & =\sum_{\rho=1}^{4} V_{\rho}(x(k), \alpha(k), \tau(k)) \\
& =\sum_{\rho=1}^{4} V_{\rho} \tag{11}
\end{align*}
$$

where

$$
\begin{gathered}
V_{1}=x(k)^{T} P(\alpha(k), \tau(k)) x(k), \\
V_{2}=\sum_{l=k-\tau_{\max }}^{k-1} x(l)^{T} H^{T} Q_{1} H x(l)+\sum_{l=k-\tau_{\min }}^{k-1} x(l)^{T} H^{T} Q_{2} H x(l), \\
V_{3}= \\
\sum_{\theta=-\tau_{\max }+2}^{-\tau_{\min }+1} \sum_{l=k+\theta-1}^{k-1} x(l)^{T} H^{T} Q_{3} H x(l) \\
\\
+\sum_{l=k-\tau(k)}^{k-1} x(l)^{T} H^{T} Q_{3} H x(l),
\end{gathered}
$$

$$
\begin{align*}
V_{4}= & \sum_{\theta=-\tau_{\max }+1}^{0} \sum_{l=k+\theta-1}^{k-1} \tau_{\max } \delta(l)^{T} H^{T} R_{1} H \delta(l) \\
& +\sum_{\theta=-\tau_{\max }+1}^{-\tau_{\min }} \sum_{l=k+\theta-1}^{k-1}\left(\tau_{\max }-\tau_{\min }\right) \\
& \times \delta(l)^{T} H^{T} R_{2} H \delta(l) \tag{12}
\end{align*}
$$

and $\delta(l)=x(l+1)-x(l)$. In the following when $\alpha(k)=i$ and $\tau(k)=m$, we will write $P(\alpha(k), \tau(k)), K(\alpha(k), \tau(k)), \bar{A}(\alpha(k))$ and $\bar{B}(\alpha(k))$ as $P_{i, m}, K_{i, m}, \bar{A}_{i}$, and $\bar{B}_{i}$, respectively. We denote:

$$
\begin{align*}
& \Delta V(x(k), \alpha(k), \tau(k)) \\
&=\sum_{\rho=1}^{4} \Delta V_{\rho}  \tag{13}\\
&= \sum_{\rho=1}^{4}\left[V_{\rho}(x(k+1), \alpha(k+1), \tau(k+1) \mid x(k),\right. \\
&\left.\quad \alpha(k), \tau(k))-V_{\rho}(x(k), \alpha(k), \tau(k))\right] .
\end{align*}
$$

Let $\xi(k)=\left[x(k)^{T}(H x(k-m))^{T}\left(H x\left(k-\tau_{\min }\right)\right)^{T}\right.$ $\left.\left(H x\left(k-\tau_{\max }\right)\right)^{T}\right]^{T}$. Then, along the solution of system (5) we have

$$
\begin{align*}
\mathbb{E}\left[\Delta V_{1}\right]= & x(k+1)^{T}\left[\sum_{j=0}^{1} \sum_{n=0}^{s_{2}} \pi_{i j} \lambda_{m n} P_{j, n}\right] \\
& \times x(k+1)-x^{T}(k) P_{i, m} x(k) \\
= & \xi^{T}(k)\left[\begin{array}{c}
\bar{A}_{i}^{T} \\
\bar{B}_{i}^{T} \\
0 \\
0
\end{array}\right] \bar{P}_{i, m}\left[\begin{array}{llll}
\bar{A}_{i} & \bar{B}_{i} & 0 & 0
\end{array}\right] \xi(k)  \tag{14}\\
& -x^{T}(k) P_{i, m} x(k)
\end{align*}
$$

where $\bar{P}_{i, m}$ is defined in Theorem 3.
We have

$$
\begin{aligned}
\mathbb{E}\left[\Delta V_{2}\right]= & x^{T}(k) H^{T}\left(Q_{1}+Q_{2}\right) H x(k) \\
& -x^{T}\left(k-\tau_{\max }\right) H^{T} Q_{1} H x\left(k-\tau_{\max }\right) \\
& -x^{T}\left(k-\tau_{\min }\right) H^{T} Q_{2} H x\left(k-\tau_{\min }\right) \\
\mathbb{E}\left[\Delta V_{3}\right]= & \left(\tau_{\max }-\tau_{\min }+1\right) x(k)^{T} H^{T} Q_{3} H x(k) \\
& -\sum_{\theta=k-\tau_{\max }+1}^{k-\tau_{\min }} x(l)^{T} H^{T} Q_{3} H x(l) \\
& +\left(\sum_{l=k-n+1}^{k-1}-\sum_{l=k-m+1}^{k-1}\right) x(l)^{T} H^{T} Q_{3} H x(l) \\
& -x(k-m)^{T} H^{T} Q_{3} H x(k-m)
\end{aligned}
$$

Note that

$$
\begin{align*}
\sum_{l=k-n+1}^{k-1} & x^{T}(l) H^{T} Q_{3} H x(l) \\
& =\left[\sum_{l=k-\tau_{\min }+1}^{k-1}+\sum_{l=k-n+1}^{k-\tau_{\min }}\right] x^{T}(l) H^{T} Q_{3} H x(l)  \tag{17}\\
& \leq\left[\sum_{l=k-m+1}^{k-1}+\sum_{l=k-\tau_{\max }+1}^{k-\tau_{\min }}\right] x^{T}(l) H^{T} Q_{3} H x(l) .
\end{align*}
$$

By combining (16) and (17), we have

$$
\begin{align*}
\mathbb{E}\left[\Delta V_{3}\right] \leq & \left(\tau_{\max }-\tau_{\min }+1\right) x(k)^{T} H^{T} Q_{3} H x(k) \\
& -x(k-m)^{T} H^{T} Q_{3} H x(k-m)  \tag{18}\\
\mathbb{E}\left[\Delta V_{4}\right]= & \tau_{\max }^{2} \delta^{T}(k) H^{T} R_{1} H \delta(k)
\end{align*}
$$

$$
-\sum_{l=k-\tau_{\max }}^{k-1} \tau_{\max } \delta^{T}(l) H^{T} R_{1} H \delta(l)
$$

$$
\begin{equation*}
+\left(\tau_{\max }-\tau_{\min }\right)^{2} \delta^{T}(k) H^{T} R_{2} H \delta(k) \tag{19}
\end{equation*}
$$

$$
-\sum_{l=k-\tau_{\max }}^{k-\tau_{\min }-1}\left(\tau_{\max }-\tau_{\min }\right) \delta^{T}(l) H^{T} R_{2} H \delta(l) .
$$

By Jensen's inequality, we can get

$$
\begin{aligned}
& \sum_{l=k-\tau_{\max }}^{k-1} \tau_{\max } \delta^{T}(l) H^{T} R_{1} H \delta(l) \\
&=\left(\sum_{l=k-\tau_{\max }}^{k-\tau_{k}-1}+\sum_{l=k-\tau_{k}}^{k-1}\right)\left(\tau_{\max }-\tau_{k}+\tau_{k}\right) \\
& \times \delta^{T}(l) H^{T} R_{1} H \delta(l) \\
& \geq\left(\tau_{\max }-\tau_{k}\right) \sum_{l=k-\tau_{\max }}^{k-\tau_{k}-1} \delta^{T}(l) H^{T} R_{1} H \delta(l) \\
&+\tau_{k} \sum_{l=k-\tau_{k}}^{k-1} \delta^{T}(l) H^{T} R_{1} H \delta(l) \\
& \geq\left(\sum_{l=k-\tau_{\max }}^{k-\tau_{k}-1} \delta(l)\right)^{T} H^{T} R_{1} H\left(\sum_{l=k-\tau_{\max }}^{k-\tau_{k}-1} \delta(l)\right) \\
&+\left(\sum_{l=k-\tau_{k}}^{k-1} \delta(l)\right)^{T} H^{T} R_{1} H\left(\sum_{l=k-\tau_{k}}^{k-1} \delta(l)\right)
\end{aligned}
$$

$$
\begin{align*}
\geq & \left(x\left(k-\tau_{\max }\right)-x\left(k-\tau_{k}\right)\right)^{T} H^{T} R_{1} H \\
& \times\left(x\left(k-\tau_{\max }\right)-x\left(k-\tau_{k}\right)\right) \\
& +\left(x\left(k-\tau_{k}\right)-x(k)\right)^{T} H^{T} R_{1} H\left(x\left(k-\tau_{k}\right)-x(k)\right) \\
\geq & \xi(k)^{T}\left\{\left[\begin{array}{cccc}
H^{T} R_{1} H & -H^{T} R_{1} & 0 & 0 \\
-R_{1} H & 2 R_{1} & 0 & -R_{1} \\
0 & 0 & 0 & 0 \\
0 & -R_{1} & 0 & R_{1}
\end{array}\right]\right\} \xi(k) . \tag{20}
\end{align*}
$$

Similarly, we have

$$
\begin{align*}
& \sum_{l=k-\tau_{\max }}^{k-\tau_{\min }-1}\left(\tau_{\max }-\tau_{\min }\right) \delta^{T}(l) H^{T} R_{2} H \delta(l) \\
& \quad=\left(\sum_{l=k-\tau_{\max }}^{k-\tau_{k}-1}+\sum_{l=k-\tau_{k}}^{k-\tau_{\min }-1}\right)  \tag{21}\\
& \quad \times\left(\tau_{\max }-\tau_{k}+\tau_{k}-\tau_{\min }\right) \delta^{T}(l) H^{T} R_{2} H \delta(l) \\
& \quad \geq \xi(k)^{T}\left\{\left[\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & 2 R_{2} & -R_{2} & -R_{2} \\
0 & -R_{2} & R_{2} & 0 \\
0 & -R_{2} & 0 & R_{2}
\end{array}\right]\right\} \xi(k)
\end{align*}
$$

By combining (19), (20), and (21), we have

$$
\begin{align*}
& \mathbb{E}\left[\Delta V_{4}\right] \leq \xi(k)^{T}\left\{\left[\begin{array}{c}
\tau_{\max }\left(H\left(\bar{A}_{i}-I\right)\right)^{T} \\
\tau_{\max }\left(H \bar{B}_{i}\right)^{T} \\
0 \\
0
\end{array}\right]\right. \\
& \times R_{1}\left[\tau_{\max } H\left(\bar{A}_{i}-I\right) \quad \tau_{\max } H \bar{B}_{i} \quad 0 \quad 0\right] \\
& +\left[\begin{array}{c}
\tilde{\tau}\left(H\left(\bar{A}_{i}-I\right)\right)^{T} \\
\tilde{\tau}\left(H \bar{B}_{i}\right)^{T} \\
0 \\
0
\end{array}\right] \\
& \times R_{2}\left[\begin{array}{llll}
\tilde{\tau} H\left(\bar{A}_{i}-I\right) & \tilde{\tau} H \bar{B}_{i} & 0 & 0
\end{array}\right] \\
& +\left[\begin{array}{cccc}
-H^{T} R_{1} H & H^{T} R_{1} & 0 & 0 \\
R_{1} H & -2 R_{1} & 0 & R_{1} \\
0 & 0 & 0 & 0 \\
0 & R_{1} & 0 & -R_{1}
\end{array}\right] \\
& \left.+\left[\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & -2 R_{2} & R_{2} & R_{2} \\
0 & R_{2} & -R_{2} & 0 \\
0 & R_{2} & 0 & -R_{2}
\end{array}\right]\right\} \xi(k), \tag{22}
\end{align*}
$$

where $\tilde{\tau}$ is defined in Theorem 3 .
By combining (14), (15), (18), and (22), we have

$$
\begin{align*}
& \mathbb{E}[\Delta V] \leq \xi^{T}(k)\left\{\begin{array}{cccc}
\Pi_{i, m} & H^{T} R_{1} & 0 & 0 \\
R_{1} H & -Q_{3}-2 R_{1}-2 R_{2} & R_{2} & R_{1}+R_{2} \\
0 & R_{2} & -Q_{2}-R_{2} & 0 \\
0 & R_{1}+R_{2} & 0 & -Q_{1}-R_{1}-R_{2}
\end{array}\right] \\
& +\left[\begin{array}{c}
\bar{A}_{i}^{T} \\
\bar{B}_{i}^{T} \\
0 \\
0
\end{array}\right] \bar{P}_{i, m}\left[\begin{array}{llll}
\bar{A}_{i} & \bar{B}_{i} & 0 & 0
\end{array}\right]+\left[\begin{array}{c}
\tau_{\max }\left(H\left(\bar{A}_{i}-I\right)\right)^{T} \\
\tau_{\max }\left(H \bar{B}_{i}\right)^{T} \\
0 \\
0
\end{array}\right] R_{1}\left[\tau_{\max } H\left(\bar{A}_{i}-I\right) \tau_{\max } H \bar{B}_{i} 00\right]  \tag{23}\\
& \left.\left.+\left[\begin{array}{c}
\widetilde{\tau}\left(H\left(\bar{A}_{i}-I\right)\right)^{T} \\
\widetilde{\tau}\left(H \bar{B}_{i}\right)^{T} \\
0 \\
0
\end{array}\right] R_{2}\left[\widetilde{\tau} H\left(\bar{A}_{i}-I\right) \widetilde{\tau} H \bar{B}_{i} \quad 0 \quad 0\right]\right]\right\} \xi(k) \\
& =\xi^{T}(k) \Theta_{i, m} \xi(k),
\end{align*}
$$

where $\Pi_{i, m}$ and $\tilde{\tau}$ are defined in Theorem 3.
By Schur complement and from (9), we have $\Theta_{i, m}<0$. Therefore,

$$
\begin{equation*}
\mathbb{E}[\Delta V] \leq-\boldsymbol{\lambda}_{\min }\left(-\Theta_{i, m}\right) \xi(k)^{T} \xi(k) \leq-\boldsymbol{\eta} x(k)^{T} x(k) \tag{24}
\end{equation*}
$$

where $\boldsymbol{\lambda}_{\text {min }}\left(-\Theta_{i, m}\right)$ denotes the minimal eigenvalue of $-\Theta_{i, m}$
and $\boldsymbol{\eta}=\inf \left\{\lambda_{\text {min }}\left(-\Theta_{i, m}\right)\right\}$. From (24), it is seen that for any $t>0$

$$
\begin{align*}
\mathbb{E} & {[V(x(k+1), \alpha(k+1), \tau(k+1))] } \\
& -\mathbb{E}[V(\varphi, \alpha(0), \tau(0))] \leq-\eta \sum_{k=0}^{t} E\left[x(k)^{T} x(k)\right] . \tag{25}
\end{align*}
$$

Furthermore

$$
\begin{equation*}
\sum_{k=0}^{t} \mathbb{E}\left[x(k)^{T} x(k)\right] \leq \frac{1}{\boldsymbol{\eta}} \mathbb{E}[V(\varphi, \alpha(0), \tau(0))] . \tag{26}
\end{equation*}
$$

By taking limit as $t \rightarrow \infty$, we have

$$
\begin{equation*}
\sum_{k=0}^{\infty} \mathbb{E}\left[x(k)^{T} x(k)\right] \leq \frac{1}{\boldsymbol{\eta}} \mathbb{E}[V(\varphi, \alpha(0), \tau(0))]<\infty . \tag{27}
\end{equation*}
$$

According to Definition 2, the closed-loop system (5) is stochastically stable. This completes the proof.

Theorem 3 gives the sufficient conditions for the stochastic stability of system (5). However, it should be noted that the conditions (9) are no more LMI conditions. To handle this, the equivalent LMI conditions are given in Theorem 4 by Cone Complementarity Linearization (CCL) algorithm.

Theorem 4. Consider system (5) with random but bounded $\operatorname{scalar} \tau(k) \in\left[\begin{array}{ll}\tau_{\min } & \tau_{\max }\end{array}\right]$. There exists an output feedback controller (4) such the resulting closed-loop system is stochastically stable if for each mode $i \in \bar{S}_{1}, m \in \bar{S}_{2}$, there exist matrices $P_{i, m}>0, X_{i, m}>0, Q_{1}>0, Q_{2}>0, Q_{3}>0, R_{1}>0, R_{2}>0$, $U_{1}>0, U_{2}>0$, and $K_{i, m}$ such that

$$
\begin{gather*}
{\left[\begin{array}{cccc}
\Xi_{1} & \widehat{\Xi}_{2} & \widehat{\Xi}_{3} & \widehat{\Xi}_{4} \\
* & -X_{j, n} & 0 & 0 \\
* & * & -U_{1} & 0 \\
* & * & * & -U_{2}
\end{array}\right]<0}  \tag{28}\\
P_{i, m} X_{i, m}=I, \quad R_{1} U_{1}=I, \quad R_{2} U_{2}=I, \tag{29}
\end{gather*}
$$

where

$$
\begin{gather*}
\widehat{\Xi}_{2}=\left[\begin{array}{llll}
\mathscr{L}_{i, m} \bar{A}_{i} & \mathscr{L}_{i, m} \bar{B}_{i} & 0 & 0
\end{array}\right]^{T}, \\
\widehat{\Xi}_{3}=\left[\begin{array}{llll}
\tau_{\max } H\left(\bar{A}_{i}-I\right) & \tau_{\max } H \bar{B}_{i} & 0 & 0
\end{array}\right]^{T}, \\
\widehat{\Xi}_{4}=\left[\begin{array}{lll}
\widetilde{\tau} H\left(\bar{A}_{i}-I\right) & \tilde{\tau} H \bar{B}_{i} & 0
\end{array} 0^{T},\right. \\
X_{j, n}=\operatorname{diag}\left\{\begin{array}{lll}
X_{0,0}, X_{0,1}, \ldots, X_{0, s_{2}}, X_{1,0}, X_{1,1}, \ldots, X_{1, s_{2}}
\end{array}\right\}, \\
\left(\mathscr{L}_{i, m} \bar{A}_{i}\right)^{T}=\left[\begin{array}{llll}
\sqrt{\pi_{i 0} \lambda_{m 0}} \bar{A}_{i}^{T} & \sqrt{\pi_{i 0} \lambda_{m 1}} \bar{A}_{i}^{T} & \cdots & \sqrt{\pi_{i 0} \lambda_{m s_{2}}} \bar{A}_{i}^{T} \\
\sqrt{\pi_{i 1} \lambda_{m 0}} \bar{A}_{i}^{T} & \cdots & \sqrt{\pi_{i 1} \lambda_{m s_{2}}} \bar{A}_{i}^{T}
\end{array}\right], \\
\left(\mathscr{L}_{i, m} \bar{B}_{i}\right)^{T}=\left[\begin{array}{llll}
\sqrt{\pi_{i 0} \lambda_{m 0}} \bar{B}_{i}^{T} & \sqrt{\pi_{i 0} \lambda_{m 1}} \bar{B}_{i}^{T} & \cdots & \sqrt{\pi_{i 0} \lambda_{m s}} \bar{B}_{i}^{T} \\
\sqrt{\pi_{i 1} \lambda_{m 0}} \bar{B}_{i}^{T} & \cdots & \sqrt{\pi_{i 1} \lambda_{m s}} \bar{B}_{i}^{T}
\end{array}\right],  \tag{30}\\
\bar{A}_{i}=\left[\begin{array}{cc}
A & 0 \\
0 & i I
\end{array}\right]+\left[\begin{array}{c}
i B \\
0
\end{array}\right] K_{i, m}\left[\begin{array}{ll}
0 & I
\end{array}\right], \\
\bar{A}_{i}-I=\left[\begin{array}{cc}
A-I & 0 \\
0 & (i-1) I
\end{array}\right]+\left[\begin{array}{cc}
i B \\
0
\end{array}\right] K_{i, m}\left[\begin{array}{ll}
0 & I
\end{array}\right], \\
\bar{B}_{i}=\left[\begin{array}{cc}
0 & \\
(1-i) C
\end{array}\right]+\left[\begin{array}{cc}
(1-i) B \\
0
\end{array}\right] K_{i, m} C, \\
H\left(\bar{A}_{i}-I\right)=\left[\begin{array}{ll}
A-I & 0
\end{array}\right]+i B K_{i, m}\left[\begin{array}{ll}
0 & I
\end{array}\right], \\
H \bar{B}_{i}=(1-i) B K_{i, m} C .
\end{gather*}
$$

$\Xi_{1}$ and $\tilde{\tau}$ are defined in Theorem 3. Moreover, if (28) (29) have solutions, the controller gain is given by $K_{i, m}$.

Proof. By Schur complement, (28) is equivalent to

$$
\left[\begin{array}{cccc}
\Xi_{1} & \widehat{\Xi}_{2} & \widehat{\Xi}_{3} & \widehat{\Xi}_{4}  \tag{31}\\
* & -P_{j, n}^{-1} & 0 & 0 \\
* & * & -R_{1}^{-1} & 0 \\
* & * & * & -R_{2}^{-1}
\end{array}\right]<0 .
$$

Let $P_{j, n}^{-1}=X_{j, n}, R_{1}^{-1}=U_{1}$, and $R_{2}^{-1}=U_{2}$; we can obtain (28), (29). This completes the proof.

The conditions state in Theorem 4 are a set of LMIs with some matrix inverse constraints. Although they are
nonconvex, which prevents us from solving them using the existing convex optimization tool, we can use the con complementary linearization to algorithm transform this problem into the nonlinear minimization problem with LMI constraints as follows:

$$
\begin{gather*}
\min \operatorname{Trace}\left(\sum_{s=1}^{2} R_{s} U_{s}+\sum_{i=0}^{1} \sum_{m=0}^{s_{2}} P_{i, m} X_{i, m}\right) \\
\mathrm{s} \cdot \mathrm{t}\left\{\begin{array}{l}
\text { (i) } \mathrm{LMI}(28) \\
\text { (ii) }\left[\begin{array}{cc}
R_{s} & I \\
I & U_{s}
\end{array}\right]>0, \quad s \in\{1,2\}, \\
\text { (iii) }\left[\begin{array}{cc}
P_{i, m} & I \\
I & X_{i, m}
\end{array}\right]>0, \quad i \in \bar{S}_{1}, m \in \bar{S}_{2}
\end{array}\right. \tag{32}
\end{gather*}
$$

The above nonlinear minimization problem can be solved by an iterative algorithm presented in the following.

Algorithm 5. Step 1. Find a feasible solution satisfying LMIs (i), (ii), and (iii) in (32); set as ( $\left.R_{1}^{0}, U_{1}^{0}, R_{2}^{0}, U_{2}^{0}, P_{i, m}^{0}, X_{i, m}^{0}, K^{0}\right)$ and $k=0$.

Step 2. Solve the following LMI optimization problem for variables $\left(R_{1}, U_{1}, R_{2}, U_{2}, P_{i, m}, X_{i, m}, K\right)$. Minimize trace $\left\{\sum_{s=1}^{2}\left(R_{s}^{k} U_{s}+R_{s} U_{s}^{k}\right)+\sum_{i=0}^{1} \sum_{m=0}^{s_{2}}\left(P_{i, m}^{k} X_{i, m}+P_{i, m} X_{i, m}^{k}\right)\right\}$, subject to LMIs (32). Set $R_{1}^{k+1}=R_{1}, U_{1}^{k+1}=U_{1}, R_{2}^{k+1}=R_{2}, U_{2}^{k+1}=U_{2}$, $P_{i, m}^{k+1}=P_{i, m}, X_{i, m}^{k+1}=X_{i, m}$, and $K^{k+1}=K$.

Step 3. If (31) is satisfied, then exit the iteration. If (31) is not satisfied, let $k=k+1$, and then return to Step 2.

## 4. Numerical Example

To illustrate the effectiveness of the proposed method, we apply the results in Section 3 to a classical angular positioning system [43] in Figure 2, where $\theta$ is the angular position of the antenna, $\theta_{r}$ is the angular position of the moving object, and the angular velocity of the antenna $\dot{\theta}$ is measurable. The control problem is to use the input voltage to the motor to rotate the antenna so that it always points in the direction of a moving object in the plant. The output feedback controller is designed for the following values of the matrices $A, B$, and $C$ :

$$
\begin{gather*}
A=\left[\begin{array}{cc}
1 & 0.0995 \\
0 & 0.99
\end{array}\right], \\
B=\left[\begin{array}{c}
0.0039 \\
0.0783
\end{array}\right],  \tag{33}\\
C=\left[\begin{array}{cc}
1.4 & 0.8 \\
-0.2 & 0.4
\end{array}\right] .
\end{gather*}
$$

The stochastic jumping parameter $\alpha(k) \in\{0,1\}$ and the random delays involved in system (5) are $\tau(k) \in\{0,1,2\}$; the transition probability matrices $\pi$ and $\lambda$ are taken by

$$
\pi=\left[\begin{array}{cc}
0.4 & 0.6  \tag{34}\\
0.55 & 0.45
\end{array}\right], \quad \lambda=\left[\begin{array}{ccc}
0.36 & 0.54 & 0.1 \\
0.26 & 0.52 & 0.22 \\
0.18 & 0.62 & 0.2
\end{array}\right]
$$

Figures 3 and 4 show part of the simulation of the stochastic jumping parameter $\alpha(k)$ and S-C delay $\tau(k)$ governed by their corresponding transition probability matrices, respectively.

The initial value $x(0)=\left[\begin{array}{ll}-0.4 & 0.6\end{array}\right]^{T}$. By Theorem 4, we can obtain the gain matrices $K_{i, m}$ of controller (4) which are constructed as

$$
\begin{aligned}
& K_{0,0}=\left[\begin{array}{ll}
-1.1356 & -1.6672
\end{array}\right], \\
& K_{0,1}=\left[\begin{array}{ll}
-0.2870 & -0.4926
\end{array}\right], \\
& K_{0,2}=\left[\begin{array}{ll}
-0.2898 & -0.4909
\end{array}\right],
\end{aligned}
$$



Figure 2: The angular positioning system.


Figure 3: Values of $\alpha(k)$.

$$
\begin{align*}
& K_{1,0}=\left[\begin{array}{ll}
-0.2886 & -0.4900
\end{array}\right], \\
& K_{1,1}=\left[\begin{array}{ll}
-0.2857 & -0.4909
\end{array}\right], \\
& K_{1,2}=\left[\begin{array}{ll}
-0.2829 & -0.4805
\end{array}\right] . \tag{35}
\end{align*}
$$

The state trajectories and the delay output trajectories are shown in Figures 5 and 6, where four curves represent state trajectories and the delay output trajectories under the controller gains $K_{i, m}$. Figures 5 and 6 indicate that system (5) is stochastically stable. In contrast with the proposed method, the controller gain $K_{\text {dlqr }}$ of a standard linear-quadratic regulator for nominal discrete-time systems designed by MATLAB command dlqr is

$$
K_{\mathrm{dlqr}}=\left[\begin{array}{ll}
0.9332 & 1.6804 \tag{36}
\end{array}\right] .
$$

The eigenvalues of $A+B K^{\mathrm{lqr}}$ are 1.1686 and 0.9567 . Hence, $K^{\mathrm{lqr}}$ cannot stabilize the system in this case. The proposed controller works much better for networked control system than the contrastive dlqr method.


Figure 4: S-C random delays $\tau(k)$.


Figure 5: State trajectories under $K_{i, m}$.

## 5. Conclusion

The stability analysis problem for NCSs with random time delays and packet dropouts is investigated in this paper. The random time delays and packet dropouts existed in feedback communication link are modeled by two independent Markov chains. Then the resulting closed-loop system is modeled as a MJLS with Markov delays. Sufficient conditions on stochastic stability and stabilization are obtained by the Lyapunov stability theory and LMI method. The CCL algorithm is employed to obtain the mode-dependent output feedback controller. Finally, an example is presented to illustrate the effectiveness of the approach. Although the NCSs with random time delays and packet dropouts on only sensor to controller link are considered in this paper, the method of unified modeling and the Lyapunov functional constructing can be extended to the NCSs with the random


Figure 6: The trajectories of delay output $\bar{y}(k)$.
time delays and packet dropouts existing in both the sensor to controller and controller to actuator.

## Acknowledgments

The research is supported by the National Natural Science Foundation of China (Grant no. 61203184) and the Natural Science Foundation of SZU (Grant no. 201207).

## References

[1] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," Proceedings of the IEEE, vol. 95, no. 1, pp. 138-172, 2007.
[2] M. C. F. Donkers, W. P. M. H. Heemels, N. van de Wouw, and L. Hetel, "Stability analysis of networked control systems using a switched linear systems approach," IEEE Transactions on Automatic Control, vol. 56, no. 9, pp. 2101-2115, 2011.
[3] L. X. Zhang, H. J. Gao, and O. Kaynak, "Network-induced constraints in networked control systems-a survey," IEEE Transactions on Automatic Control, vol. 9, no. 1, pp. 403-416, 2013.
[4] J. Xiong and J. Lam, "Stabilization of linear systems over networks with bounded packet loss," Automatica, vol. 43, no. 1, pp. 80-87, 2007.
[5] Z. Wang, F. Yang, D. W. C. Ho, and X. Liu, "Robust $H_{\infty}$ control for networked systems with random packet losses," IEEE Transactions on Systems, Man, and Cybernetics B, vol. 37, no. 4, pp. 916-924, 2007.
[6] W.-A. Zhang and L. Yu, "Output feedback stabilization of networked control systems with packet dropouts," IEEE Transactions on Automatic Control, vol. 52, no. 9, pp. 1705-1710, 2007.
[7] T. Jia, Y. Niu, and X. Wang, " $H_{\infty}$ control for networked systems with data packet dropout," International Journal of Control, Automation and Systems, vol. 8, no. 2, pp. 198-203, 2010.
[8] J. Wu and T. Chen, "Design of networked control systems with packet dropouts," IEEE Transactions on Automatic Control, vol. 52, no. 7, pp. 1314-1319, 2007.
[9] B. Xue, S. Li, and Q. Zhu, "Moving horizon state estimation for networked control system with multiple packet dropouts," IEEE Transactions on Automatic Control, vol. 57, no. 9, pp. 2360-2366, 2012.
[10] Q. Lu, L. Zhang, M. Basin, and H. Tian, "Analysis and synthesis for networked control systems with uncertain rate of packet losses," Journal of the Franklin Institute, vol. 349, no. 7, pp. 25002514, 2012.
[11] B. Ding, "Stabilization of linear systems over networks with bounded packet loss and its use in model predictive control," Automatica, vol. 47, no. 11, pp. 2526-2533, 2011.
[12] D. E. Quevedo and D. Nešić, "Robust stability of packetized predictive control of nonlinear systems with disturbances and Markovian packet losses," Automatica, vol. 48, no. 8, pp. 18031811, 2012.
[13] B. Ding and L. Xie, "Stabilization of linear systems over networks with double-sided, finite bounded, arbitrary packet loss," in Proceedings of the IEEE International Conference on Control and Automation (ICCA '09), pp. 920-925, December 2009.
[14] E. Tian and D. Yue, "A new state feedback $H_{\infty}$ control of networked control systems with time-varying network conditions," Journal of the Franklin Institute, vol. 349, no. 3, pp. 891-914, 2012.
[15] D. Yue, E. Tian, and Q.-L. Han, "A delay system method for designing event-triggered controllers of networked control systems," IEEE Transactions on Automatic Control, vol. 58, no. 2, pp. 475-481, 2013.
[16] M. Azadegan, M. T. H. Beheshti, and B. Tavassoli, "Robust output feedback control of networked control systems with random delay modeled by Markov chain," in Proceedings of the 20th Iranian Conference on Electrical Engineering, pp. 587-592, Tehran, Iran, 2012.
[17] J. Nilsson, Real-Time Control Systems with Delays, Lund Institute of Technology, Lund, Sweden, 1998.
[18] L. Xiao, A. Hassibi, and J. P. How, "Control with random communication delays via a discrete-time jump system approach," in Proceedings of the 2000 American Control Conference, pp. 3545-3550, June 2000.
[19] Y. Shi and B. Yu, "Output feedback stabilization of networked control systems with random delays modeled by Markov chains," IEEE Transactions on Automatic Control, vol. 54, no. 7, pp. 1668-1674, 2009.
[20] M. B. G. Cloosterman, N. van de Wouw, W. P. M. H. Heemels, and H. Nijmeijer, "Stability of networked control systems with uncertain time-varying delays," IEEE Transactions on Automatic Control, vol. 54, no. 7, pp. 1575-1579, 2009.
[21] L. Qiu, B. Xu, and S. Li, "Guaranteed cost control for discretetime networked control systems with random Markov delays," Journal of Systems Engineering and Electronics, vol. 22, no. 4, pp. 661-671, 2011.
[22] E. C. Martins and F. G. Jota, "Design of networked control systems with explicit compensation for time-delay variations," IEEE Transactions on Systems, Man and Cybernetics C, vol. 40, no. 3, pp. 308-318, 2010.
[23] I. Pan, S. Das, and A. Gupta, "Tuning of an optimal fuzzy PID controller with stochastic algorithms for networked control systems with random time delay," ISA Transactions, vol. 50, no. 1, pp. 28-36, 2011.
[24] L. Qiu, B. Xu, and S. Li, " $H_{2} / H_{\infty}$ control of networked control system with random time delays," Science China. Information Sciences, vol. 54, no. 12, pp. 2615-2630, 2011.
[25] Y. Ishido, K. Takaba, and D. E. Quevedo, "Stability analysis of networked control systems subject to packet-dropouts and finite-level quantization," Systems \& Control Letters, vol. 60, no. 5, pp. 325-332, 2011.
[26] Y. Tang, H. J. Gao, W. Zou, and J. Kurths, "Distributed synchronization in networks of agent systems with nonlinearities and random switchings," IEEE Transactions on Systems, Man, and Cybernetics B, vol. 43, pp. 358-370, 2013.
[27] Z.-H. Guan, C.-Y. Chen, G. Feng, and T. Li, "Optimal tracking performance limitation of networked control systems with limited bandwidth and additive colored white Gaussian noise," IEEE Transactions on Circuits and Systems I, vol. 60, no. 1, pp. 189-198, 2013.
[28] J. Yu, L. Wang, and M. Yu, "Switched system approach to stabilization of networked control systems," International Journal of Robust and Nonlinear Control, vol. 21, no. 17, pp. 1925-1946, 2011.
[29] T. Suzuki, M. Kono, N. Takahashi, and O. Sato, "Controllability and stabilizability of a networked control system with periodic communication constraints," Systems \& Control Letters, vol. 60, no. 12, pp. 977-984, 2011.
[30] D. Wu, J. Wu, S. Chen, and J. Chu, "Stability of networked control systems with polytopic uncertainty and buffer constraint," IEEE Transactions on Automatic Control, vol. 55, no. 5, pp. 12021208, 2010.
[31] B. Zhu and K. Xia, "Design of subsidence monitoring system based on wireless sensor networks," in Proceedings of the 6th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '10), pp. 1-4, September 2010.
[32] Y. Xia, L. Li, G.-P. Liu, and P. Shi, " $H_{\infty}$ predictive control of networked control systems," International Journal of Control, vol. 84, no. 6, pp. 1080-1097, 2011.
[33] P. Wang, C. Han, and B. Ding, "Stability of discrete-time networked control systems and its extension for robust $H_{\infty}$ control," International Journal of Systems Science, vol. 44, no. 2, pp. 275-288, 2013.
[34] X. Tang and B. Ding, "Model predictive control of linear systems over networks with data quantizations and packet losses," Automatica, vol. 49, no. 5, pp. 1333-1339, 2013.
[35] N. Xie and B. Xia, "Robust fault tolerant controller design for uncertain networked control systems with fast-varying delay and packet-dropout," in Proceedings of the 6th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '10), pp. 1-4, September 2010.
[36] W.-A. Zhang and L. Yu, "Modelling and control of networked control systems with both network-induced delay and packetdropout," Automatica, vol. 44, no. 12, pp. 3206-3210, 2008.
[37] B. Yu and Y. Shi, "State feedback stabilization of networked control systems with random time delays and packet dropout," in Proceedings of the ASME Dynamic Systems and Control Conference (DSCC '08), pp. 1-7, October 2008.
[38] H. Dong, Z. Wang, and H. Gao, "Robust $H_{\infty}$ filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts," IEEE Transactions on Signal Processing, vol. 58, no. 4, pp. 1957-1966, 2010.
[39] C. Jiang, D. X. Zou, and Q. L. Zhang, "Observer-based control for networked control systems with network induced delay and
packet dropout," in Proceedings of the 8th World Congress on Intelligent Control and Automation (WCICA '10), pp. 4376-4379, July 2010.
[40] X. Li, X. Wu, Z. Xu, and C. Huang, "Fault detection observer design for networked control system with long time-delays and data packet dropout," Journal of Systems Engineering and Electronics, vol. 21, no. 5, pp. 877-882, 2010.
[41] S. B. Li, Y. Q. Wang, F. Xia et al., "Guaranteed cost control of networked guaranteed cost control of networked control systems with time-delays and packet losses," International Journal of Wavelets, Multiresolution and Information Processing, vol. 4, no. 4, pp. 691-7066, 2006.
[42] Y.-L. Wang and G.-H. Yang, "Time delay and packet dropout compensation for networked control systems: a linear estimation method," International Journal of Control, vol. 83, no. 1, pp. 115-124, 2010.
[43] A. Liu, L. Yu, and W. Zhang, "One-step receding horizon $H_{\infty}$ control for networked control systems with random delay and packet disordering," ISA Transactions, vol. 50, no. 1, pp. 44-52, 2011.
[44] L. Qiu, Q. Luo, F. Gong, S. B. Li, and B. G. Xu, "Stability and stabilization of networked control systems with random time delays and packet dropouts," Journal of the Franklin Institute, vol. 350, no. 7, pp. 1886-1907, 2013.
[45] Y. Sun and Y. Huo, "Robust static output feedback control for networked control system with random time delay," in Proceedings of the 2nd International Conference on Advanced Computer Control (ICACC '10), pp. 547-551, March 2010.
[46] C. Peng, Y. Tian, and D. Yue, "Output feedback control of discrete-time systems in networked environments," IEEE Transactions on Systems, Man, and Cybernetics A, vol. 41, no. 1, pp. 185-190, 2011.

# Nonlinear Adaptive Equivalent Control Based on Interconnection Subsystems for Air-Breathing Hypersonic Vehicles 

Chaofang Hu and Yanwen Liu<br>School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China<br>Correspondence should be addressed to Chaofang Hu; cfhu@tju.edu.cn

Received 14 May 2013; Accepted 22 July 2013
Academic Editor: Tao Zou
Copyright © 2013 C. Hu and Y. Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

For the nonminimum phase behavior of the air-breathing hypersonic vehicle model caused by elevator-to-lift coupling, a nonlinear adaptive equivalent control method based on interconnection subsystems is proposed. In the altitude loop, the backstepping strategy is applied, where the virtual control inputs about flight-path angle and attack angle are designed step by step. In order to avoid the inaccurately direct cancelation of elevator-to-lift coupling when aerodynamic parameters are uncertain, the real control inputs, that is, elevator deflection and canard deflection, are linearly converted into the equivalent control inputs which are designed independently. The reformulation of the altitude-flight-path angle dynamics and the attack angle-pitch rate dynamics is constructed into interconnection subsystems with input-to-state stability via small-gain theorem. For the velocity loop, the dynamic inversion controller is designed. The adaptive approach is used to identify the uncertain aerodynamic parameters. Simulation of the flexible hypersonic vehicle demonstrates effectiveness of the proposed method.


## 1. Introduction

Hypersonic vehicles have a promising prospect in both military and commercial applications as its flight speed can be more than 5 times of the speed of sound. However, since the model of hypersonic vehicle is nonlinear, multivariable, uncertain, and coupling [1], it is unstable and extremely sensitive to changes in flight condition and parameters. This brings a great challenge to controller design [2]. At present, most researches focus on dealing with nonlinearity and uncertainty of hypersonic vehicles. For example, linear control methods are attempted according to linearized hypersonic vehicle models, such as pole placement techniques [3], LQR method [4], linear output feedback control [5], and LPV control [6]. In addition, nonlinear control strategies are widely used as well, such as feedback linearization approach [7], sliding control $[8,9]$, and backstepping technique [10]. For uncertainty of hypersonic vehicles, besides adaptive approaches [11], robust strategies are common tools, for example, $\mu$ synthesis, $H_{\infty}$ control [12], stochastic robustness control [13], and nonlinear disturbance observer-based robust control
[14]. Although these methods are proven to be effective, they do not usually consider the coupling problems existing in hypersonic vehicles. These problems lead to more difficulties in the flight controller design. In an air-breathing hypersonic vehicle, it is known that there are structural dynamics, flexible effect, elevator-to-lift coupling, and the coupling between thrust and pitch moment, where elevator-to-lift coupling is not neglectable, and it will generate unstable zero dynamics exponentially, that is, the nonminimum phase behavior in pitch rate model, if the controller is designed directly by the inversion.

With regard to the elevator-to-lift coupling problem, some strategies have been tried. The basic method usually ignores this coupling, and then the nonminimum phase can be removed from the model during the controller design [2], where this coupling is only regarded as unmodeled dynamics. However, this manner cannot ensure the stability of the control system. The other common approach is to offset the influence of the coupling. For example, a canard is adopted to cancel the influence of elevator on lift, and an adaptive robust controller based on nonlinear sequential loop-closure
approach is developed [15, 16]. Nevertheless, the changes of the uncertain parameters are not considered. This may result in the inaccurate cancellation, which means elevator and lift are not decoupled completely. Simultaneously, this approach has an adverse influence on the pitch rate dynamics since its inputs also consist of elevator and canard. In addition, for thermal protection problem resulted from the canard, only the elevator is taken as aerodynamic control surface in reference [17]. The system model is transformed into the interconnection of systems in feedback and feedforward forms to eliminate the nonminimum phase. But the robustness with regard to uncertainty of the hypersonic vehicle model is not addressed totally.

From the analysis, we know that adding canard control surface is an effective and simple way to suppress the nonminimum phase behavior, even though the strict cancellation of the elevator-to-lift coupling cannot be realized actually. In this paper, the flexible air-breathing hypersonic vehicle model is considered. For the tracking requirement of altitude and velocity, a nonlinear adaptive equivalent control method based on interconnection subsystems is proposed by incorporating canard. Firstly, in the altitude loop, the virtual control inputs about flight-path angle and attack angle are designed step by step according to the backstepping strategy. Secondly, the terms about the real control inputs, that is, the elevator and canard deflection in the flight-path angle dynamics and the pitch rate dynamics, are linearly converted into the equivalent control inputs instead of direct cancelation of the elevator-to-lift coupling. By designing the new inputs independently, the altitude control loop is reformulated. And the adaptive technique is used to identify the uncertain aerodynamic parameters. Then the interconnection subsystems including the altitude-flight-path angle dynamics and the attack angle-pitch rate dynamics are constructed. Via the small-gain method, the system is proven to be input-to-state stable. In the velocity loop, the adaptive dynamic inversion controller is designed. Simulation results show the power of our approach.

In Section 2, the air-breathing hypersonic vehicle model is presented. The nonlinear adaptive equivalent control based on interconnection subsystems is introduced in Section 3. Section 4 presents the simulation. The conclusion is drawn in Section 5.

## 2. Air-Breathing Hypersonic Vehicle Model

In this study, the flexible air-breathing hypersonic vehicle model [18] is considered. This model is composed of five rigid-body states, that is, velocity $V$, altitude $h$, flight-path angle $\gamma$, attack angel $\alpha$, pitch rate $q$, and six flexible states, that is, $\eta_{1}, \dot{\eta}_{1}, \eta_{2}, \dot{\eta}_{2}, \dot{\eta}_{3}$, and $\dot{\eta}_{3}$. The equations of motion are written as

$$
\begin{gathered}
\dot{V}=\frac{T \cos \alpha-D}{m}-g \sin \gamma, \\
\dot{h}=V \sin \gamma, \\
\dot{\gamma}=\frac{T \sin \alpha+L}{m V}-\frac{g \cos \gamma}{V},
\end{gathered}
$$

$$
\begin{gather*}
\dot{\alpha}=Q-\dot{\gamma}, \\
\dot{Q}=\frac{M}{I_{y y}}, \\
\ddot{\eta}_{i}=-2 \varsigma_{m} \omega_{m, i} \dot{\eta}_{i}-\omega_{m, i}^{2} \eta_{i}+N_{i} ; \quad i=1,2,3, \tag{1}
\end{gather*}
$$

where $m, I_{y y}, g$ represent mass of the aircraft, moment of inertia, gravitational acceleration; damping ratio and natural frequency of the flexible motion are denoted by $\varsigma_{m}$ and $\omega_{m, i}$, respectively; $T, D, L$, and $N_{i}$ and $M$ are thrust, drag, lift, generalized forces and moment

$$
\begin{gather*}
L=\bar{q} S C_{L}, \\
T=\bar{q}\left(C_{T, \phi} \phi+C_{T}\right), \\
D=\bar{q} S C_{D},  \tag{2}\\
M=z_{T} T+\bar{q} S \bar{c} C_{M}, \\
N_{i}=\bar{q} C_{N_{i}} .
\end{gather*}
$$

The aerodynamic parameters in the above formulation are described as follows:

$$
\begin{align*}
C_{L}= & C_{L}^{\alpha} \alpha+C_{L}^{\delta_{e}} \delta_{e}+C_{L}^{\delta_{c}} \delta_{c}+C_{L}^{0}+C_{L}^{\Delta \tau_{1}} \Delta \tau_{1}+C_{L}^{\Delta \tau_{2}} \Delta \tau_{2}, \\
C_{M}= & C_{M}^{\alpha} \alpha+C_{M}^{\delta_{e}} \delta_{e}+C_{M}^{\delta c} \delta_{c}+C_{M}^{0}+C_{M}^{\Delta \tau_{1}} \Delta \tau_{1}+C_{M}^{\Delta \tau_{2}} \Delta \tau_{2}, \\
C_{N_{i}}= & C_{N_{i}}^{\alpha} \alpha+C_{N_{i}}^{\delta_{e}} \delta_{e}+C_{N_{i}}^{\delta_{c}} \delta_{c}+C_{N_{i}}^{0}+C_{N_{i}}^{\Delta \tau_{1}} \Delta \tau_{1}+C_{N_{i}}^{\Delta \tau_{2}} \Delta \tau_{2}, \\
C_{D}= & C_{D}^{\left(\alpha+\Delta \tau_{1}\right)^{2}}\left(\alpha+\Delta \tau_{1}\right)^{2}+C_{D}^{\left(\alpha+\Delta \tau_{1}\right)}\left(\alpha+\Delta \tau_{1}\right)+C_{D}^{\Delta \tau_{2}} \Delta \tau_{2} \\
& +C_{D}^{\delta_{e}^{2}} \delta_{e}^{2}+C_{D}^{\delta_{e}} \delta_{e}+C_{D}^{\alpha \delta_{e}} \alpha \delta_{e}+C_{D}^{\delta c^{2}} \delta_{c}^{2} \\
& +C_{D}^{\delta_{c}} \delta_{c}+C_{D}^{\alpha \delta_{c}} \alpha \delta_{c}+C_{D}^{0}, \\
C_{T, \phi}= & C_{T, \phi}^{\alpha} \alpha+C_{T, \phi}^{\alpha M_{\infty}^{-2}} \alpha M_{\infty}^{-2}+C_{T, \phi}^{M_{\infty}^{-2}} M_{\infty}^{-2}+C_{T, \phi}^{0} \\
& +C_{T, \phi}^{\alpha \Delta \tau_{1}} \alpha \Delta \tau_{1}+C_{T, \phi}^{\Delta \tau_{1}^{2}} \Delta \tau_{1}^{2}+C_{T, \phi}^{\Delta \tau_{1}} \Delta \tau_{1}, \\
C_{T}= & C_{T}^{\alpha} \alpha+C_{T}^{M_{\infty}^{-2}} M_{\infty}^{-2}+C_{T}^{A_{d}} A_{d}+C_{T}^{\Delta \tau_{1}} \Delta \tau_{1}+C_{T}^{0}, \tag{3}
\end{align*}
$$

where the control inputs are fuel-to-air ratio $\phi$, elevator deflection $\delta_{e}$, and canard deflection $\delta_{c} ; \bar{q}, S, z_{T}, \bar{c}, M_{\infty}$ denote dynamic pressure, reference area, thrust moment arm, mean aerodynamic chord, and Mach number; $\Delta \tau_{1}$ and $\Delta \tau_{2}$ are the forebody turn angle and the aftbody vertex angle which are linear mapping of elastic states $\eta_{i}$.

In (3), the elevator-to-lift coupling orients from that $C_{L}$ includes the term of $\delta_{e}$, which leads to the nonminimum phase behavior. If $\delta_{e}$ is designed by the dynamic inversion directly, the pitch rate dynamics will become a hyperbolic saddle equilibrium. This unstable zero dynamic brings great difficulties to the controller design.

## 3. Nonlinear Adaptive Equivalent Controllers Design

In order to track the altitude and velocity command signals $h_{\text {ref }}$ and $V_{\text {ref }}$, two controllers will be designed independently for the altitude loop and the velocity loop. During the controller design, the flexible motion is viewed as external perturbation, and its influence on aerodynamic model (3) is neglected.
3.1. Altitude Controller. In the altitude loop, the controller is designed according to the backstepping approach. Then the virtual control inputs about flight-path angle and attack angle are determined, respectively.

For the altitude dynamics, let $\widetilde{h}=h-h_{\text {ref }}$; then its error dynamics is written in the following:

$$
\begin{equation*}
\dot{\tilde{h}}=V \sin \gamma-\dot{h}_{\mathrm{ref}} \approx V \gamma-\dot{h}_{\mathrm{ref}} . \tag{4}
\end{equation*}
$$

So the flight-path angle command $\gamma_{d}$ is designed into the following equation:

$$
\begin{equation*}
\gamma_{d}=\frac{-k_{\tilde{h}} \tilde{h}+\dot{h}_{\mathrm{ref}}}{V} \tag{5}
\end{equation*}
$$

where $k_{\widetilde{h}}>0$ is the design parameter for $\widetilde{h}$.
Let $\tilde{\gamma}=\gamma-\gamma_{d}$; the error dynamic of flight-path angle is presented as follows:

$$
\begin{equation*}
\dot{\tilde{\gamma}}=\frac{T \sin \alpha+L}{m V}-\frac{g \cos \gamma}{V}-\dot{\gamma}_{d} . \tag{6}
\end{equation*}
$$

Here, the thrust is described as the function about the attack angle. Define $T=\nabla T \alpha+T_{0}$, where $\nabla T=\bar{q}\left(C_{T, \phi}^{\alpha M_{\infty}^{-2}} M_{\infty}^{-2} \phi+\right.$ $\left.C_{T, \phi}^{\alpha} \phi+C_{T}^{\alpha}\right)$ and $T_{0}=\bar{q}\left(C_{T, \phi}^{M_{\infty}^{-2}} \phi M_{\infty}^{-2}+C_{T, \phi}^{0} \phi+C_{T}^{M_{\infty}^{-2}} M_{\infty}^{-2}+\right.$ $C_{T}^{A_{d}} A_{d}+C_{T}^{0}$ ). As the variation range of the attack angle is small, (6) will be expanded around the final expectation $\alpha^{*}$.

To handle the nonminimum phase problem, the MIMO equivalent method is applied in this paper, which is different from the previous research results [17]. The terms about the elevator and canard deflection are linearly equivalent to the control input vector $\mathbf{U}=\left[U_{1}, U_{2}\right]$. The error model of the flight-path angle (6) can be rewritten as

$$
\begin{align*}
& \begin{aligned}
\dot{\tilde{\gamma}}= & \frac{\nabla T \alpha \sin \alpha+T_{0} \sin \alpha+\bar{q} S C_{L}^{\alpha} \alpha}{m V}+\bar{q} S \frac{C_{L}^{\delta_{e}} \delta_{e}}{m V}+\bar{q} S \frac{C_{L}^{\delta_{C}} \delta_{C}}{m V} \\
& +\frac{\bar{q} S C_{L}^{0}-m g \cos \gamma-m V \dot{\gamma}_{d}}{m V}
\end{aligned} \\
& =\underbrace{\frac{\nabla T \sin \alpha^{*}+\nabla T \alpha^{*} \cos \alpha^{*}+T_{0} \cos \alpha^{*}+\bar{q} S C_{L}^{\alpha}}{m V}}_{C_{1}} \alpha+\underbrace{\bar{q} S \frac{C_{L}^{\delta_{e}} \delta_{e}+C_{L}^{\delta_{c}} \delta_{c}}{m V}}_{U_{1}} \\
& =\underbrace{\frac{\bar{q} S C_{L}^{0}-m g \sin \gamma-m V \dot{\gamma}_{d}}{m V}+\frac{T_{0} \sin \alpha^{*}-\left(\nabla T \alpha^{*} \cos \alpha^{*}+T_{0} \cos \alpha^{*}\right) \alpha^{*}}{m V}}_{\beta_{1} \alpha+U_{1}+\beta_{1},} \\
& =C_{1} \alpha+U_{1}+\beta_{1} \text {, } \tag{7}
\end{align*}
$$

where $C_{1}$ and $\beta_{1}$ are the terms containing the uncertain aerodynamic parameters. They can be expressed as the following equations:

$$
\begin{gather*}
C_{1}=\theta_{1}^{T} \xi_{1} \\
\beta_{1}=\theta_{1}^{T} \xi_{2}-\frac{g \cos \gamma}{V}-\dot{\gamma}_{d}  \tag{8}\\
U_{1}=\theta_{2}^{T} \xi_{3} .
\end{gather*}
$$

$\theta_{1}, \theta_{2}$ are vectors of the uncertain parameters

$$
\begin{align*}
\theta_{1} & =\left[C_{T, \phi}^{\alpha} ; C_{T}^{\alpha} ; C_{T, \phi}^{\alpha M_{\infty}^{-2}} ; C_{T, \phi}^{M_{\infty}^{-2}} ; C_{T, \phi}^{0} ; C_{T}^{M_{\infty}^{-2}} ; C_{T}^{A_{d}} ; C_{T}^{0} ; C_{L}^{\alpha} ; C_{L}^{0}\right], \\
\theta_{2} & =\left[C_{L}^{\delta_{e}} ; C_{L}^{\delta_{c}}\right], \tag{9}
\end{align*}
$$

and $\xi_{1}, i=1 \ldots 3$ are regressors

$$
\begin{aligned}
& \xi_{1}=\frac{\bar{q}}{m V}[ \left(\sin \alpha^{*}+\alpha^{*} \cos \alpha^{*}\right) \phi ;\left(\sin \alpha^{*}+\alpha^{*} \cos \alpha^{*}\right) ; \\
&\left(\sin \alpha^{*}+\alpha^{*} \cos \alpha^{*}\right) M_{\infty}^{-2} \phi ; \cos \alpha^{*} M_{\infty}^{-2} \phi ; \\
&\left.\cos \alpha^{*} \phi ; \cos \alpha^{*} M_{\infty}^{-2} ; \cos \alpha^{*} A_{d} ; \cos \alpha^{*} ; S ; 0\right] \\
& \xi_{2} \\
&=\frac{\bar{q}}{m V}\left[-\alpha^{* 2} \cos \alpha^{*} \phi ;-\alpha^{* 2} \cos \alpha^{*} ;-\alpha^{* 2} \cos \alpha^{*} M_{\infty}^{-2} \phi ;\right. \\
&\left(\sin \alpha^{*}-\alpha^{*} \cos \alpha^{*}\right) M_{\infty}^{-2} \phi ;\left(\sin \alpha^{*}-\alpha^{*} \cos \alpha^{*}\right) \phi ; \\
&\left(\sin \alpha^{*}-\alpha^{*} \cos \alpha^{*}\right) M_{\infty}^{-2} ;\left(\sin \alpha^{*}-\alpha^{*} \cos \alpha^{*}\right) A_{d ;} ; \\
&\left.\left(\sin \alpha^{*}-\alpha^{*} \cos \alpha^{*}\right) ; 0 ; S\right]
\end{aligned}
$$

$$
\begin{equation*}
\xi_{3}=\frac{\bar{q} S}{m V}\left[\delta_{e} ; \delta_{c}\right] \tag{10}
\end{equation*}
$$

Therefore the dynamics (7) is reformulated as

$$
\begin{equation*}
\dot{\tilde{\gamma}}=\theta_{1}^{T} \xi_{1} \alpha+\theta_{2}^{T} \xi_{3}+\theta_{1}^{T} \xi_{2}-\frac{g \cos \gamma}{V}-\dot{\gamma}_{d} \tag{11}
\end{equation*}
$$

Then the virtual command of the attack angle is chosen as $\alpha_{d}=\alpha^{*}-\tilde{\gamma}$.

Let $\widetilde{\alpha}=\alpha-\alpha_{d}$; the error dynamic of the attack angle is formulated as

$$
\begin{equation*}
\dot{\tilde{\alpha}}=Q-\dot{\gamma}-\dot{\alpha}_{d}=Q-\dot{\gamma}_{d} \tag{12}
\end{equation*}
$$

A new variable $Z$ is defined as $Z=Q-\dot{\gamma}_{d}+k_{\tilde{\alpha}} \widetilde{\alpha}$, where $k_{\tilde{\alpha}}>0$ is a design parameter for $\widetilde{\alpha}$. Then (12) is rewritten as

$$
\begin{equation*}
\dot{\tilde{\alpha}}=Z-k_{\tilde{\alpha}} \widetilde{\alpha} . \tag{13}
\end{equation*}
$$

Using the equivalent control method, the time derivative of $Z$ can be formulated with the new input $U_{2}$. It includes the pitch rate dynamics

$$
\begin{align*}
\dot{Z}= & \frac{z_{T} T+\bar{q} S \bar{c} C_{M}}{I_{y y}}+k_{\tilde{\alpha}} \dot{\tilde{\alpha}} \\
= & \underbrace{\bar{q} S \bar{c} \frac{C_{M}^{\delta_{e}} \delta_{e}+C_{M}^{\delta c} \delta_{c}}{I_{y y}}}_{U_{2}}+\underbrace{\frac{z_{T} \nabla T+\bar{q} S \bar{c} C_{M}^{\alpha}}{I_{y y}} \alpha}_{C_{2}}  \tag{14}\\
& +\underbrace{\frac{z_{T} T_{0}+\bar{q} S \bar{c} C_{M}^{0}}{I_{y y}}+k_{\tilde{\alpha}} \dot{\tilde{\alpha}}-\ddot{\gamma}_{d}}_{\beta_{2}} \\
= & C_{2} \alpha+U_{2}+\beta_{2},
\end{align*}
$$

where $C_{2}$ and $\beta_{2}$ are similar terms containing the uncertain parameters. They can also be presented by the vectors of the uncertain parameters and the regressors

$$
\begin{gather*}
C_{2}=\theta_{3}^{T} \xi_{4} \\
\beta_{2}=\theta_{3}^{T} \xi_{5}+k_{\tilde{\alpha}} \dot{\tilde{\alpha}}-\ddot{\gamma}_{d}  \tag{15}\\
U_{2}=\theta_{4}^{T} \xi_{6}
\end{gather*}
$$

where
$\theta_{3}=\left[C_{T, \phi}^{\alpha} ; C_{T}^{\alpha} ; C_{T, \phi}^{\alpha M_{\infty}^{2}} ; C_{T, \phi}^{M_{\infty}^{2}} ; C_{T, \phi}^{0} ; C_{T}^{M_{\infty}^{2}} ; C_{T}^{A_{d}} ; C_{T}^{0} ; C_{M}^{\alpha} ; C_{M}^{0}\right]$,
$\theta_{4}=\left[C_{M}^{\delta_{e}} ; C_{M}^{\delta_{c}}\right]$,
$\xi_{4}=\frac{\bar{q}}{I_{y y}}\left[z_{T} \phi ; z_{T} ; z_{T} M_{\infty}^{-2} \phi ; 0 ; 0 ; 0 ; 0 ; 0 ; S \bar{c} ; 0\right]$,
$\xi_{5}=\frac{\bar{q}}{I_{y y}}\left[0 ; 0 ; 0 ; z_{T} M_{\infty}^{-2} \phi ; z_{T} \phi ; z_{T} M_{\infty}^{-2} ; z_{T} A_{d} ; z_{T} ; 0 ; S \bar{c}\right]$,
$\xi_{6}=\frac{\bar{q} S \bar{c}}{I_{y y}}\left[\delta_{e} ; \delta_{c}\right]$.

So (14) can be reformulated as

$$
\begin{equation*}
\dot{Z}=\theta_{3}^{T} \xi_{4} \alpha+\theta_{4}^{T} \xi_{6}+\theta_{3}^{T} \xi_{5}+k_{\tilde{\alpha}} \dot{\tilde{\alpha}}-\ddot{\gamma}_{d} \tag{17}
\end{equation*}
$$

Due to the uncertainty of the aerodynamic parameters, $\theta_{i}$, $i=1, \ldots, 4$ will change with flight of hypersonic vehicles. Therefore it is necessary to estimate their values by the adaptive technique. Let $\widehat{\theta}_{i}, \widetilde{\theta}_{i}$ be the estimate vector and the estimate error vector of $\theta_{i}$, where $\widetilde{\theta}_{i}=\theta_{i}-\widehat{\theta}_{i}, i=1, \ldots, 4$.

Assumption 1. The aerodynamic parameters $\theta_{i}, i=1, \ldots, 4$ are bounded; they lie in a compact convex set.

In order to guarantee tracking performance of hypersonic vehicles, the equivalent control inputs $U_{1}$ and $U_{2}$ are
designed, respectively, by replacing the uncertain parameter vector $\theta_{i}$ with its estimate vector and estimate error vector

$$
\begin{align*}
\widehat{U}_{1}= & \widehat{\theta}_{2}^{T} \xi_{3}=-\widehat{\theta}_{1}^{T} \xi_{1} \alpha^{*}-\left(\widehat{\theta}_{1}^{T} \xi_{2}-\frac{g \cos \gamma}{V}-\dot{\gamma}_{d}\right) \\
& +\left(\widehat{\theta}_{1}^{T} \xi_{1}-k_{\tilde{\gamma}}\right) \tilde{\gamma}-V \widetilde{h},  \tag{18}\\
\widehat{U}_{2}= & \widehat{\theta}_{4}^{T} \xi_{6}=\ddot{\gamma}_{d}-\left(\widehat{\theta}_{3}^{T} \xi_{5}+k_{\tilde{\alpha}} \dot{\tilde{\alpha}}\right)-\widehat{\theta}_{3}^{T} \xi_{4} \alpha^{*} \\
& -k_{Z} Z-\left(\widehat{\theta}_{3}^{T} \xi_{4}+1\right) \widetilde{\alpha},
\end{align*}
$$

where $k_{\tilde{\gamma}}>0, k_{Z}>0$ are the design parameters for $\tilde{\gamma}$ and $Z$.
Let $\delta=\left[\delta_{e}, \delta_{c}\right]$. There is $\mathbf{U}=B \delta$ according to (7) and (14). $B$ is a coefficient matrix and is equal to $\left[(\bar{q} S / m V) \widehat{\theta}_{2}^{T} ;\left(\bar{q} S \bar{c} / I_{y y}\right) \widehat{\theta}_{4}^{T}\right]$. The real inputs of the altitude loop can be obtained as follows:

$$
\left[\begin{array}{l}
\delta_{e}  \tag{19}\\
\delta_{c}
\end{array}\right]=B^{-1}\left[\begin{array}{l}
\widehat{U}_{1} \\
\widehat{U}_{2}
\end{array}\right]
$$

Combining (18), the state error dynamics about the altitude loop is transformed into the following equations:

$$
\begin{gather*}
\dot{\tilde{h}} \approx-k_{\tilde{h}} \widetilde{h}+V \widetilde{\gamma}, \\
\dot{\tilde{\gamma}}=-k_{\tilde{\gamma}} \widetilde{\gamma}-V \widetilde{h}+y_{\tilde{\alpha}}+\widetilde{\theta}_{1}^{T} \xi_{1} \alpha+\widetilde{\theta}_{1}^{T} \xi_{2}+\widetilde{\theta}_{2}^{T} \xi_{3},  \tag{20}\\
\dot{\tilde{\alpha}}=Z-k_{\widetilde{\alpha}} \widetilde{\alpha}, \\
\dot{Z}=-k_{Z} Z-\widetilde{\alpha}+y_{\tilde{\gamma}}+\widetilde{\theta}_{3}^{T} \xi_{4} \alpha+\widetilde{\theta}_{3}^{T} \xi_{5}+\widetilde{\theta}_{4}^{T} \xi_{6},
\end{gather*}
$$

where $y_{\widetilde{\alpha}}=\widehat{\theta}_{1}^{T} \xi_{1} \widetilde{\alpha}, y_{\widetilde{\gamma}}=-\widehat{\theta}_{3}^{T} \xi_{4} \widetilde{\gamma}$.
For ensuring the stability of the altitude loop, the new formulation (20) is divided into the altitude-flight-path angle subsystem and the attack angle-pitch rate subsystem. As illustrated in Figure 1, these two subsystems constitute a structure of interconnection. It is seen that $y_{\tilde{\alpha}}$ and $y_{\tilde{\gamma}}$ act as the input and output of the altitude-flight-path angle subsystem and $y_{\tilde{\gamma}}, y_{\tilde{\alpha}}$ are the input and output of the attack angle-pitch rate subsystem, respectively.

For the above interconnection subsystems, input-to-state stability will be analyzed via small gain theorem. Firstly, the definition of the asymptotic $L_{\infty}$ norm $\|\cdot\|_{a}$ is given [19]

$$
\begin{equation*}
\|\lambda\|_{a}:=\lim _{t \rightarrow \infty} \sup |\lambda| \tag{21}
\end{equation*}
$$

Then, define $\psi_{1}=\sqrt{\widetilde{h}^{2}+\widetilde{\gamma}^{2}}$, and choose the Lyapunov function candidate of the altitude-flight-path angle subsystem as

$$
\begin{equation*}
W_{1}=\frac{1}{2}\left(\widetilde{h}^{2}+\widetilde{\gamma}^{2}\right)+\frac{1}{2} \widetilde{\theta}_{1}^{T} \tau_{1}^{-1} \widetilde{\theta}_{1}+\frac{1}{2} \widetilde{\theta}_{2}^{T} \tau_{2}^{-1} \widetilde{\theta}_{2} \tag{22}
\end{equation*}
$$

Its time derivative is

$$
\begin{align*}
\dot{W}_{1}= & \widetilde{h} \dot{\tilde{h}}+\widetilde{\gamma} \dot{\tilde{\gamma}}-\widetilde{\theta}_{1}^{T} \tau_{1}^{-1} \dot{\hat{\theta}}_{1}-\widetilde{\theta}_{2}^{T} \tau_{2}^{-1} \dot{\hat{\theta}}_{2} \\
= & -k_{\tilde{h}} \widetilde{h}^{2}-k_{\tilde{\gamma}} \widetilde{\gamma}^{2}+\widetilde{\gamma} y_{\widetilde{\alpha}}+\widetilde{\theta}_{1}^{T} \tau_{1}^{-1}\left\{\tau_{1} \widetilde{\gamma}\left(\xi_{1} \alpha+\xi_{2}\right)-\dot{\hat{\theta}}_{1}\right\} \\
& +\widetilde{\theta}_{2}^{T} \tau_{2}^{-1}\left(\tau_{2} \widetilde{\gamma} \xi_{3}-\dot{\hat{\theta}}_{2}\right) . \tag{23}
\end{align*}
$$



Figure 1: Interconnection subsystem structure.

The adaptive laws of $\widehat{\theta}_{1}, \widehat{\theta}_{2}$ are designed as

$$
\begin{gather*}
\dot{\hat{\theta}}_{1}=\tau_{1} \tilde{\gamma}\left(\xi_{1} \alpha+\xi_{2}\right),  \tag{24}\\
\dot{\hat{\theta}}_{2}=\tau_{2} \widetilde{\gamma} \xi_{3}
\end{gather*}
$$

where $\tau_{1}$ and $\tau_{2}$ are the adaptive parameters.
By (24), (23) becomes

$$
\begin{equation*}
\dot{W}_{1} \leq-\min \left\{k_{\widetilde{h}}, k_{\tilde{\gamma}}\right\}\left|\psi_{1}\right|^{2}+\left|\psi_{1}\right|\left|y_{\tilde{\alpha}}\right| . \tag{25}
\end{equation*}
$$

As a consequence, it satisfies $W_{1}$ which is negative definite when $\left|\psi_{1}\right| \geq\left|y_{\tilde{\alpha}}\right| / \min \left\{k_{\tilde{h}}, k_{\tilde{\gamma}}\right\}$. Then $W_{1}$ is a input-to-state stable Lyapunov function. According to the lemma in [19], we know that $\left\|\psi_{1}\right\|_{a} \leq\left\|y_{\tilde{\alpha}}\right\|_{a} / \min \left\{k_{\widetilde{h}}, k_{\tilde{\gamma}}\right\}$. As $y_{\tilde{\gamma}}=-\widehat{\theta}_{3}^{T} \xi_{4} \widetilde{\gamma}$, the following formulation is obtained:

$$
\begin{equation*}
\left\|y_{\tilde{\gamma}}\right\|_{a}=\left\|-\widehat{\theta}_{3}^{T} \xi_{4} \tilde{\gamma}\right\|_{a} \leq \widehat{\theta}_{3}^{T} \xi_{4}\left\|\psi_{1}\right\|_{a} \leq \frac{\widehat{\theta}_{3}^{T} \xi_{4}}{\min \left\{k_{\tilde{h}}, k_{\tilde{\gamma}}\right\}}\left\|y_{\tilde{\alpha}}\right\|_{a} . \tag{26}
\end{equation*}
$$

For the attack angle-pitch rate subsystem, $\psi_{2}=\sqrt{\widetilde{\alpha}^{2}+Z^{2}}$ is defined, and the following Lyapunov function candidate is chosen:

$$
\begin{equation*}
W_{2}=\frac{1}{2}\left(\widetilde{\alpha}^{2}+Z^{2}\right)+\frac{1}{2} \widetilde{\theta}_{3}^{T} \tau_{3}^{-1} \widetilde{\theta}_{3}+\frac{1}{2} \widetilde{\theta}_{4}^{T} \tau_{4}^{-1} \widetilde{\theta}_{4} . \tag{27}
\end{equation*}
$$

Its time derivative is

$$
\begin{align*}
\dot{W}_{2}= & \widetilde{\alpha} \dot{\tilde{\alpha}}+Z \dot{Z}-\widetilde{\theta}_{3}^{T} \tau_{3}^{-1} \dot{\hat{\theta}}_{3}-\widetilde{\theta}_{4}^{T} \tau_{4}^{-1} \dot{\hat{\theta}}_{4} \\
= & -k_{\tilde{\alpha}} \widetilde{\alpha}^{2}+k_{Z} Z^{2}+Z y_{\widetilde{\gamma}}+\widetilde{\theta}_{3}^{T} \tau_{3}^{-1}\left\{\tau_{3} Z\left(\xi_{4} \alpha+\xi_{5}\right)-\dot{\hat{\theta}}_{3}\right\} \\
& +\widetilde{\theta}_{4}^{T} \tau_{4}^{-1}\left(\tau_{4} Z \xi_{6}-\dot{\hat{\theta}}_{4}\right) \tag{28}
\end{align*}
$$

where the adaptive laws of $\hat{\theta}_{3}, \widehat{\theta}_{4}$ are determined as

$$
\begin{gather*}
\dot{\hat{\theta}}_{3}=\tau_{3} Z\left(\xi_{4} \alpha+\xi_{5}\right),  \tag{29}\\
\dot{\hat{\theta}}_{4}=\tau_{4} Z \xi_{6} .
\end{gather*}
$$

Substituting (29) in (28), we can acquire

$$
\begin{equation*}
\dot{W}_{2} \leq-\min \left\{k_{\tilde{\alpha}}, k_{Z}\right\}\left|\psi_{2}\right|^{2}+\left|\psi_{2}\right|\left|y_{\tilde{\gamma}}\right| . \tag{30}
\end{equation*}
$$

When $\left|\psi_{2}\right| \geq\left|y_{\tilde{\gamma}}\right| / \min \left\{k_{\tilde{\alpha}}, k_{Z}\right\}, \dot{W}_{2} \leq 0$. Similarly, $\left\|\psi_{2}\right\|_{a} \leq$ $\left\|y_{\tilde{\gamma}}\right\|_{a} / \min \left\{k_{\tilde{\alpha}}, k_{Z}\right\}$ can be obtained, and $W_{2}$ is input-to-state stable as well. Because $y_{\tilde{\alpha}}=\widehat{\theta}_{1}^{T} \xi_{1} \tilde{\alpha}$, there is

$$
\begin{equation*}
\left\|y_{\tilde{\alpha}}\right\|_{a}=\left\|\widehat{\theta}_{1}^{T} \xi_{1} \tilde{\alpha}\right\|_{a} \leq \widehat{\theta}_{1}^{T} \xi_{1}\left\|\psi_{2}\right\|_{a} \leq \frac{\widehat{\theta}_{1}^{T} \xi_{1}}{\min \left\{k_{\tilde{\alpha}}, k_{Z}\right\}}\left\|y_{\tilde{\gamma}}\right\|_{a} . \tag{31}
\end{equation*}
$$

The interconnection formulation (20) is input-to-state stable according to small-gain theorem if we choose proper design parameters to make the following equation holds

$$
\begin{equation*}
\frac{\hat{\theta}_{3}^{T} \xi_{4}}{\min \left\{k_{\widetilde{h}}, k_{\tilde{\gamma}}\right\}} \cdot \frac{\hat{\theta}_{1}^{T} \xi_{1}}{\min \left\{k_{\widetilde{\alpha}}, k_{Z}\right\}}<1 \tag{32}
\end{equation*}
$$

Therefore the tracking errors and estimate errors of the altitude loop can converge to a small neighborhood of origin.
3.2. Velocity Controller. Since velocity is controlled by $\phi$ directly, the adaptive dynamic inversion method is used. Let $\widetilde{V}=V-V_{\text {ref }}$; the error dynamics of velocity is written as

$$
\begin{align*}
\dot{\tilde{V}} & =\frac{T \cos \alpha-D}{m}-g \sin \gamma-\dot{V}_{\text {ref }} \\
& =\frac{\bar{q}\left(C_{T, \phi} \phi+C_{T}\right) \cos \alpha-\bar{q} S C_{D}}{m}-g \sin \gamma-\dot{V}_{\text {ref }} \tag{33}
\end{align*}
$$

For existence of uncertain parameters, the following vectors and repressors are defined

$$
\begin{align*}
\theta_{5}= & {\left[C_{D}^{\left(\alpha+\Delta \tau_{1}\right)} ; C_{D}^{\left(\alpha+\Delta \tau_{1}\right)^{2}} ; C_{D}^{\delta_{e}^{2}} ; C_{D}^{\delta_{e}} ; C_{D}^{\delta c^{2}} ; C_{D}^{\delta_{c}} ;\right.} \\
& \left.C_{D}^{\alpha \delta_{e}} ; C_{D}^{\alpha \delta_{c}} ; C_{D}^{0} ; C_{T}^{A_{d}} ; C_{T}^{\alpha} ; C_{T}^{M_{\infty}^{-2}} ; C_{T}^{0}\right] \\
\theta_{6}= & {\left[C_{T, \phi}^{\alpha} ; C_{T, \phi}^{\alpha M_{\infty}^{-2}} ; C_{T, \phi}^{M_{\infty}^{-2}} ; C_{T, \phi}^{0}\right] }  \tag{34}\\
\xi_{7}=\bar{q}[ & {\left[\alpha \alpha ; S \alpha^{2} ; S \delta_{e}^{2} ; S \delta_{e} ; S \delta_{c}^{2} ; S \delta_{c} ; S \alpha \delta_{e} ; S \alpha \delta_{c} ; S ;\right.} \\
& \left.\quad-A_{d} \cos \alpha ;-\alpha \cos \alpha ;-M_{\infty}^{-2} \cos \alpha ;-\cos \alpha\right], \\
\xi_{8}= & \bar{q} \cos \alpha\left[\alpha ; \alpha M_{\infty}^{-2} ; M_{\infty}^{-2} ; 1\right] .
\end{align*}
$$

Consequently,

$$
\begin{equation*}
\dot{\tilde{V}}=\frac{\theta_{6}^{T} \xi_{8} \phi-\theta_{5}^{T} \xi_{7}}{m}-g \sin \gamma-\dot{V}_{\mathrm{ref}} . \tag{35}
\end{equation*}
$$

Then the control input $\phi$ is designed as

$$
\begin{equation*}
\phi=\frac{-m k_{\tilde{v}} \widetilde{V}+m \dot{V}_{\mathrm{ref}}+m g \sin \gamma+\widehat{\theta}_{5}^{T} \xi_{7}}{\hat{\theta}_{6}^{T} \xi_{8}} \tag{36}
\end{equation*}
$$

where $k_{\tilde{v}}>0$ is a design parameter.


Figure 2: Continued.


Figure 2: Climbing maneuver with longitudinal acceleration for case one.

Determine the Lyapunov function candidate as

$$
\begin{equation*}
W_{3}=\frac{1}{2} m \widetilde{V}^{2}+\frac{1}{2} \widetilde{\theta}_{5}^{T} \tau_{5}^{-1} \widetilde{\theta}_{5}+\frac{1}{2} \widetilde{\theta}_{6}^{T} \tau_{6}^{-1} \widetilde{\theta}_{6} \tag{37}
\end{equation*}
$$

Its time derivative is

$$
\begin{align*}
\dot{W}_{3}= & \widetilde{V}\left(\theta_{6}^{T} \xi_{8} \phi-\theta_{5}^{T} \xi_{7}-m g \sin \gamma-m \dot{V}_{\mathrm{ref}}\right) \\
& -\widetilde{\theta}_{5}^{T} \tau_{5}^{-1} \dot{\hat{\theta}}_{5}-\widetilde{\theta}_{6}^{T} \tau_{6}^{-1} \dot{\hat{\theta}}_{6} \\
= & \widetilde{V}\left(\widetilde{\theta}_{6}^{T} \xi_{8} \phi-m k_{\widetilde{v}} \widetilde{V}-\widetilde{\theta}_{5}^{T} \xi_{7}\right)-\widetilde{\theta}_{5}^{T} \tau_{5}^{-1} \dot{\hat{\theta}}_{5}-\widetilde{\theta}_{6}^{T} \tau_{6}^{-1} \dot{\hat{\theta}}_{6} \\
= & -m k_{\widetilde{v}} \widetilde{V}^{2}-\widetilde{\theta}_{5}^{T} \tau_{5}^{-1}\left(\tau_{5} \widetilde{V} \xi_{7}+\dot{\hat{\theta}}_{5}\right)+\widetilde{\theta}_{6}^{T} \tau_{6}^{-1}\left(\tau_{6} \phi \xi_{8}-\dot{\hat{\theta}}_{6}\right) . \tag{38}
\end{align*}
$$

The adaptive laws of $\widehat{\theta}_{5}, \widehat{\theta}_{6}$ are obtained in the following:

$$
\begin{gather*}
\dot{\hat{\theta}}_{5}=-\tau_{5} \widetilde{V} \xi_{7} \\
\dot{\hat{\theta}}_{6}=\tau_{6} \phi \xi_{8} \tag{39}
\end{gather*}
$$

Thus (38) becomes $\dot{W}_{3}=-m k_{\tilde{v}} \widetilde{V}^{2}<0$; that is, when $t \rightarrow$ $\infty$, the tracking errors and estimate errors of the velocity subsystem can converge to zero finally.

Therefore the accurate tracking performance and the stability of altitude and velocity can be guarranteed by the proposed method.

## 4. Numerical Simulation

The feasibility of the proposed method is verified based on a flexible model (1)-(3). The initial trim conditions are $h=$ $85000 \mathrm{ft}, V=7846 \mathrm{ft} / \mathrm{s}, \alpha=0.0174 \mathrm{rad}, \gamma=0 \mathrm{rad}, q=$ $0 \mathrm{rad} / \mathrm{s}, \eta_{1}=0.4588 \mathrm{ft} \cdot \mathrm{sulg}, \eta_{2}=-0.08726 \mathrm{ft} \cdot \mathrm{sulg}$, and $\eta_{3}=$ -0.03671 ft -sulg. Two cases are studied here.

Case one is a climbing maneuver with longitudinal acceleration, and the expected equilibrium is $\alpha^{*}=0.0219 \mathrm{rad}$.

The increments of altitude and velocity are 2000 ft , and $100 \mathrm{ft} / \mathrm{s}$ respectively. Case two is a descending maneuver with velocity reducing gradually, and its corresponding equilibrium is $\alpha^{*}=0.0158 \mathrm{rad}$. The decreasing of altitude is 1000 ft and that of velocity is $100 \mathrm{ft} / \mathrm{s}$.

For these two cases, The corresponding reference commands are generated by filtering step reference commands with a second-order profiler with $\omega=0.1 \mathrm{rad} / \mathrm{s}$ and $\xi=0.9$.

The simulation results of case one are shown in Figure 2.
The simulation results of case two are shown in Figure 3.
From Figures 2(a), 2(b), 3(a), and 3(b), it is seen that the controller can provide stable and accurate tracking of the reference trajectories for the two cases, and the tracking errors of altitude and velocity remain remarkably small. Figures 2(c), 2(d), 3(c), and 3(d) show that both the signals of the flight-path angle and the attack angle can also follow the change of virtual control commands closely.

For the flexible dynamics, its effect on aerodynamic model is neglected, and its motion is taken as external perturbation during the control design. That means that the forebody turn angle $\Delta \tau_{1}$ and the aftbody vertex angle $\Delta \tau_{2}$ are equal to zero in model (3) when the controller is designed. Simultaneously, the second-order equation about flexible states $\eta_{i}$ and $\dot{\eta}_{i}$ is not considered. From Figures 2(e) and 3(e), we can know that the flexible states can converge to stable states ultimately although the flexible dynamics is not taken into account directly. This denotes that our controller has the strong robustness, and it is suitable to control the flexible hypersonic vehicle.

Moreover, the variation ranges of the control inputs that is, fuel-to-air ratio, elevator deflection, and canard deflection are bounded according to Figures 2(f), 2(g), 2(h), 3(f), 3(g), and 3(h).

In summary, the nonminimum phase behavior is suppressed successfully, and the excellent closed-loop behavior of air-breathing hypersonic vehicle can be obtained by the proposed controller for the cases of maneuver of altitude and velocity.



Figure 3: Descending maneuver with velocity reducing gradually for case two.

## 5. Conclusion

For the flexible air-breathing hypersonic vehicle with canard control surface, the controller is designed based on the nonlinear adaptive equivalent control strategy under interconnected structure. The equivalent control inputs are introduced and designed to replace the terms about elevator and canard in the flight-path angle dynamics and the pitchrate dynamics for eliminating the nonminimum phase. The uncertain aerodynamic parameters are identified online by the adaptive method. And input-to-state stability of the interconnection subsystems is guaranteed by small-gain theorem. Similarly, the adaptive dynamic inversion approach is adopted in the velocity loop. With our approach, the stable and accurate tracking of the hypersonic vehicle model with nonminimum phase can be realized.

## Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant 61074064 and Natural Science Foundation of Tianjin 12JCZDJC30300. The authors are grateful to the anonymous reviewers for their helpful comments and constructive suggestions with regard to this paper.

## References

[1] M. A. Bolender and D. B. Doman, "Nonlinear longitudinal dynamical model of an air-breathing hypersonic vehicle," Journal of Spacecraft and Rockets, vol. 44, no. 2, pp. 374-387, 2007.
[2] J. T. Parker, A. Serrani, S. Yurkovich, M. A. Bolender, and D. B. Doman, "Control-oriented modeling of an air-breathing hypersonic vehicle," Journal of Guidance, Control, and Dynamics, vol. 30, no. 3, pp. 856-869, 2007.
[3] B. Fidan, M. Kuipers, P. A. Ioannou, and M. Mirmirani, "Longitudinal motion control of air-breathing hypersonic vehicles based on time-varying models," in Proceedings of the 14th AIAA/AHI International Space Planes and Hypersonics Systems

Technologies Conference, pp. 1705-1717, AIAA, Canberra, Australia, November 2006.
[4] K. P. Groves, D. O. Sigthorsson, A. Serrani, S. Yurkovich, M. A. Bolender, and D. B. Doman, "Reference command tracking for a linearized model of an air-breathing hypersonic vehicle," in Proceedings of the AIAA Guidance, Navigation, and Control Conference, pp. 2901-2914, AIAA, San Francisco, Calif, USA, August 2005.
[5] D. O. Sigthorsson, P. Jankovsky, A. Serrani, S. Yurkovich, M. A. Bolender, and D. B. Doman, "Robust linear output feedback control of an airbreathing hypersonic vehicle," Journal of Guidance, Control, and Dynamics, vol. 31, no. 4, pp. 10521066, 2008.
[6] L. G. Wu, X. B. Yang, and F. B. Li, "Nonfragile output tracking control of hypersonic air-breathing vehicles with LPV model," IEEE/ASME Transactions on Mechatronics, vol. 18, no. 4, pp. 1280-1288, 2013.
[7] H. P. Lee, S. E. Reiman, C. H. Dillon, and H. M. Youssef, "Robust nonlinear dynamic inversion control for a hypersonic cruise vehicle," in Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, pp. 3380-3388, Hilton Head, SC, USA, August 2007.
[8] H. J. Xu, M. D. Mirmirani, and P. A. Ioannou, "Adaptive sliding mode control design for a hypersonic flight vehicle," Journal of Guidance, Control, and Dynamics, vol. 27, no. 5, pp. 829-838, 2004.
[9] B. L. Tian, Q. Zong, J. Wang, and F. Wang, "Quasi-continuous high-order sliding mode controller design for reusable launch vehicles in reentry phase," Aerospace Science and Technology, vol. 28, no. 1, pp. 198-207, 2013.
[10] B. Xu, F. Sun, H. Liu, and J. Ren, "Adaptive Kriging controller design for hypersonic flight vehicle via back-stepping," IET Control Theory and Applications, vol. 6, no. 4, pp. 487-497, 2012.
[11] T. E. Gibson, L. G. Crespo, and A. M. Annaswamy, "Adaptive control of hypersonic vehicles in the presence of modeling uncertainties," in Proceedings of the 2009 American Control Conference (ACC '09), pp. 3178-3183, St. Louis, Mo, USA, June 2009.
[12] H. Buschek and A. J. Calise, "Uncertainty modeling and fixedorder controller design for a hypersonic vehicle model," Journal
of Guidance, Control, and Dynamics, vol. 20, no. 1, pp. 42-48, 1997.
[13] Q. Wang and R. F. Stengel, "Robust nonlinear control of a hypersonic aircraft," Journal of Guidance, Control, and Dynamics, vol. 23, no. 4, pp. 577-585, 2000.
[14] J. Yang, S. H. Li, C. Y. Sun, and L. Guo, "Nonlinear-disturbance-observer-based robust flight control for airbreathing hypersonic vehicles," IEEE Transactions on Aerospace and Electronic Systems, vol. 49, no. 2, pp. 1263-1275, 2013.
[15] L. Fiorentini, A. Serrani, M. A. Bolender, and D. B. Doman, "Nonlinear robust/adaptive controller design for an airbreathing hypersonic vehicle model," in Proceedings of the AIAA Guidance, Navigation, and Control Conferenc and Exhibit, pp. 269-284, AIAA, Hilton Head, SC, USA, August 2007.
[16] L. Fiorentini, A. Serrani, M. A. Bolender, and D. B. Doman, "Nonlinear robust adaptive control of flexible air-breathing hypersonic vehicles," Journal of Guidance, Control, and Dynamics, vol. 32, no. 2, pp. 401-416, 2009.
[17] L. Fiorentini, A. Serrani, M. A. Bolender, and D. B. Doman, "Nonlinear control of nonminimum phase hypersonic vehicle models," in Proceedings of the 2009 American Control Conference (ACC '09), pp. 3160-3165, St. Louis, Mo, USA, June 2009.
[18] D. O. Sigthorsson and A. Serrani, "Development of linear parameter-varying models of hypersonic air-breathing vehicles," in Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, pp. 2009-6282, AIAA, Chicago, Ill, USA, August 2009.
[19] A. R. Teel, "A nonlinear small gain theorem for the analysis of control systems with saturation," IEEE Transactions on Automatic Control, vol. 41, no. 9, pp. 1256-1270, 1996.

## Research Article

# Stability of Nonlinear Stochastic Discrete-Time Systems 

Yan Li, ${ }^{1}$ Weihai Zhang, ${ }^{2}$ and Xikui Liu ${ }^{1}$<br>${ }^{1}$ College of Information Science \& Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China<br>${ }^{2}$ College of Information and Electrical Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China

Correspondence should be addressed to Weihai Zhang; whzhang@163.com
Received 6 May 2013; Accepted 12 July 2013
Academic Editor: Baocang Ding
Copyright © 2013 Yan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studies the stability for nonlinear stochastic discrete-time systems. First of all, several definitions on stability are introduced, such as stability, asymptotical stability, and $p$ th moment exponential stability. Moreover, using the method of the Lyapunov functionals, some efficient criteria for stochastic stability are obtained. Some examples are presented to illustrate the effectiveness of the proposed theoretical results.

## 1. Introduction

Stability is the first of all the considered problems in the system analysis and synthesis of modern control theory, which plays an essential role in dealing with infinite-horizon linear-quadratic regulator, $H_{2} / H_{\infty}$ robust optimal control, and other control problems; see [1-5]. In 1892, Lyapunov introduced the concept of stability of dynamic systems and created a very powerful tool known as the Lyapunov method in the study of stability. It can be found that the Lyapunov method has been developed and applied to investigate stochastic stability of the Itô-type systems, and many important classical results on deterministic differential equations have been generalized to the stochastic Itô systems; we refer the reader to Arnold [6], Friedman [7], Has'minskii [8], Kushner [9], Kolmanovskii and Myshkis [10], Ladde and Lakshmikantham [11], Mohammed [12], and Mao [13].

Compared with the plenty of fruits of the continuoustime Itô systems, few results have been obtained on the stability of discrete-time nonlinear stochastic systems:

$$
\begin{equation*}
x(t+1)=f(x(t), w(t), t) . \tag{1}
\end{equation*}
$$

In [14], the mean square stability of the discrete-time time-varying Markov jump system

$$
\begin{equation*}
x(t+1)=\left[A_{0}\left(t, \eta_{t}\right)+\sum_{k=1}^{r} A_{k}\left(t, \eta_{t}\right) w_{k}(t)\right] x(t) \tag{2}
\end{equation*}
$$

was studied. Nevertheless [15], based on the exact observability assumption, extensively researched the mean square stability of the following linear discrete-time time-invariant system with multiplicative noise:

$$
\begin{equation*}
x(t+1)=A x(t)+B x(t) w(t) \tag{3}
\end{equation*}
$$

where the Classical Lyapunov Theorem was extended. Reference [16] considered the $p$ th mean stability of the following difference equations

$$
\begin{gather*}
x(t+1)=A(t, \omega) x(t), \quad t=0,1,2, \ldots,  \tag{4}\\
x(t+1)=A(t, \omega) x(t)+f(t, x(t)), \quad t=0,1,2, \ldots,
\end{gather*}
$$

with random coefficients. For the nonlinear stochastic difference equation

$$
\begin{align*}
x(t+1)= & \sum_{j=0}^{t+h} a_{j} x(t-j)+\sum_{j=0}^{t+h} \sigma_{j} x(t-j) \xi(t)  \tag{5}\\
& +g(t, x(-h), \ldots, x(t))
\end{align*}
$$

its stability in probability was investigated in [17]. It is not difficult to find that, different from the continuous-time Itô systems, up to now, there lacks the systematic theory on stability of nonlinear discrete-time stochastic systems. The aim of this paper is to develop a parallel theory for stability of general nonlinear stochastic discrete-time systems, and some sufficient criteria for various stabilities are given.

Different from the Itô systems, most sufficient criteria are presented via $\mathscr{L} V(x) \leq 0$ or $\mathscr{L} V(x)<0$ together with other assumptions on the Lyapunov function $V(x)$, where $\mathscr{L}$ is the so-called infinitesimal generator associated with the given Itô system. In discrete-time stochastic systems, most stability criteria are given via $E[\Delta V(x(t))] \leq 0$ or $E[\Delta V(x(t))]<0$, where $E$ represents the mathematical expectation. So general discrete stochastic stability is more difficult to be tested due to the appearance of the mathematical expectation $E$.

The organization of this paper is as follows. Section 2 presents some stability definitions. Section 3 is devoted to developing some efficient criteria for various stabilities. Section 4 contains three examples provided to show the efficiency of the proposed results. Finally, we end this paper by Section 5 with a brief conclusion.

For convenience, we adopt the following notations:
$A^{T}$ : the transpose of the matrix $A$;
$A \geq 0(A>0): A$ is a positive semidefinite (positive definite) matrix;
$D_{r}:=\left\{x \in R^{n}:|x|<r\right\}$ for $r>0$;
$C^{2}(U)$ : the class of functions $V(x)$ twice continuously differential with respect to $x \in U$;
$Z^{+}:=\{0,1,2, \ldots\} ;$
$P\{B\}$ : the probability of event $B ;$
a.s.: almost surely, or with probability 1 ;
$I_{B}$ : the indicator function of a set $B$; that is, $I_{B}(x)=1$ if $x \in B$ or otherwise 0 ;
$a \wedge b$ : the minimum of $a$ and $b$.

## 2. Definitions of Stability

We will investigate various types of stabilities in probability for the $n$-dimensional stochastic discrete-time system

$$
\begin{equation*}
x(t+1)=f(x(t), w(t), t), \quad x\left(t_{0}\right)=x_{0} \tag{6}
\end{equation*}
$$

where $x_{0} \in R^{n}$ is a constant vector. For any given initial value $x\left(t_{0}\right)=x_{0} \in R^{n}$, (6) has a unique solution that is denoted by $x\left(t ; t_{0}, x_{0}\right)$ or $x\left(t ; t_{0}, x_{0}\right)=x(t)$ simply. $w(t)$ is a one-dimensional stochastic process defined on the complete probability space $(\Omega, F, P)$. We assume that $f(0, w(t), t) \equiv 0$ for all $t \in I:=\left\{t_{0}+k: k \in Z^{+}\right\}$, so (6) has the solution $x(t) \equiv$ 0 corresponding to the initial value $x\left(t_{0}\right)=0$. This solution is called the trivial solution or the equilibrium position.

Definition 1. The trivial solution of (6) is said to be stochastically stable or stable in probability if, for every $\varepsilon>0$ and $h>0$, there exists $\delta=\delta\left(\varepsilon, h, t_{0}\right)>0$, such that

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad t \geq t_{0} \tag{7}
\end{equation*}
$$

when $\left|x_{0}\right|<\delta$. Otherwise, it is said to be stochastically unstable.

If the previous $\delta$ is independent of $t_{0}$, that is, $\delta=\delta(\varepsilon, h)>$ 0 , then the trivial solution of (6) is said to be stochastically uniformly stable in probability.

Definition 2. The trivial solution of (6) is said to be stochastically asymptotically stable in probability if it is stochastically stable, and for every $\varepsilon>0$, there exists $\delta=\delta\left(\varepsilon, t_{0}\right)>0$, such that

$$
\begin{equation*}
P\left\{\lim _{t \rightarrow \infty} x(t)=0\right\} \geq 1-\varepsilon \tag{8}
\end{equation*}
$$

when $\left|x_{0}\right|<\delta$.
Definition 3. The trivial solution of (6) is said to be stochastically uniformly asymptotically stable in probability if it is stochastically uniformly stable in probability, and for every $\varepsilon>0, h>0$, there exist $\delta=\delta(\varepsilon, h)>0$ and a $T(\varepsilon)>0$, such that

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad \forall t \geq t_{0}+T(\varepsilon), \text { when }\left|x_{0}\right|<\delta . \tag{9}
\end{equation*}
$$

Definition 4. The trivial solution of (6) is said to be stochastically asymptotically stable in the large in probability if it is stochastically stable, and for all $x_{0} \in R^{n}$,

$$
\begin{equation*}
P\left\{\lim _{t \rightarrow \infty} x(t)=0\right\}=1 \tag{10}
\end{equation*}
$$

Definition 5. The trivial solution of (6) is said to be uniformly bounded if, for every $\alpha>0$ and $t_{0} \in Z^{+}$, there exists $\beta=$ $\beta(\alpha)>0$, such that

$$
\begin{equation*}
|x(t)|<\beta, \quad \text { a.s. }, \tag{11}
\end{equation*}
$$

when $\left|x_{0}\right|<\alpha$ and $t \geq t_{0}$.
Definition 6. The trivial solution of (6) is said to be stochastically uniformly asymptotically stable in the large in probability if the following are satisfied:
(i) it is stochastically uniformly stable;
(ii) it is uniformly bounded;
(iii) for any $\alpha>0, h>0$, and $\varepsilon>0$, there exists $T(\varepsilon, \alpha)>$ 0 , such that

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad \forall t \geq t_{0}+T(\varepsilon, \alpha),\left|x_{0}\right|<\alpha . \tag{12}
\end{equation*}
$$

Definition 7. The trivial solution of (6) is said to be $p$ th moment exponentially stable if there exist positive constants $\lambda$ and $C$, such that

$$
\begin{equation*}
E|x(t)|^{p} \leq C\left|x_{0}\right|^{p} e^{-\lambda\left(t-t_{0}\right)} \tag{13}
\end{equation*}
$$

where $p>0, t \geq t_{0}$, and $x_{0} \in R^{n}$. When $p=2$, it is usually said to be exponentially stable in mean square.

Below, we consider such a continuous function

$$
\begin{equation*}
V(x): R^{n} \longrightarrow R \tag{14}
\end{equation*}
$$

with $V(0)=0$, and write

$$
\begin{equation*}
\Delta V(x(t))=V(x(t+1))-V(x(t)) \tag{15}
\end{equation*}
$$

Definition 8 (see [13]). A continuous function $\varphi:[0,+\infty) \rightarrow$ $[0,+\infty)$ is said to belong to class $K$ if it is strictly increasing and $\varphi(0)=0$.

Definition 9 (see [13]). A continuous function $V(x)$ defined on $D_{r}$ is said to be positive definite (in the sense of Lyapunov) if $V(0)=0$ and, for some $\varphi \in K$,

$$
\begin{equation*}
V(x) \geq \varphi(|x|) \tag{16}
\end{equation*}
$$

A continuous function $V(x)$ defined on $D_{r}$ is said to be negative definite (in the sense of Lyapunov) if $-V(x)$ is positive definite.

Definition 10 (see [13]). A function $V(x)$ defined on $D_{r}$ is said to be radially unbounded if

$$
\begin{equation*}
\lim _{|x| \rightarrow \infty} \inf V(x)=\infty \tag{17}
\end{equation*}
$$

Definition 11. A function $V(x)$ defined on $D_{r}$ is said to have infinite small upper bound if there exists $\varphi \in K$ such that

$$
\begin{equation*}
|V(x)| \leq \varphi(|x|) \tag{18}
\end{equation*}
$$

## 3. Main Results

In this section, we state our main results in this paper. By using the method of the Lyapunov functionals, some efficient criteria for the stability are obtained.

Theorem 12. If there exists a positive definite function $V(x) \in$ $C^{2}\left(D_{r}\right)$, such that

$$
\begin{equation*}
E[\Delta V(x(t))] \leq 0, \tag{19}
\end{equation*}
$$

for all $x(t) \in D_{r}$, then the trivial solution of (6) is stochastically stable in probability.

Proof. By the definition of $V(x)$, we obtain that $V(0)=0$ and that there exists a function $\varphi \in K$, such that

$$
\begin{equation*}
V(x) \geq \varphi(|x|), \quad \forall x \in D_{r} \tag{20}
\end{equation*}
$$

For any $\varepsilon \in(0,1)$ and $h>0$, without loss of generality, we assume that $h<r$. Because $V(x)$ is continuous, we can find that $\delta=\delta\left(\varepsilon, h, t_{0}\right)>0$, such that

$$
\begin{equation*}
V(x) \leq \varepsilon \varphi(h), \quad \forall x \in D_{\delta} \tag{21}
\end{equation*}
$$

It is obvious that $\delta<h$. We fix the initial value $x_{0} \in D_{\delta}$ arbitrarily. Let $\mu$ be the first exit time of $x(t)$ from $D_{h}$; that is,

$$
\begin{equation*}
\mu=\inf \left\{t \geq t_{0}: x(t) \notin D_{h}\right\} \tag{22}
\end{equation*}
$$

Let $\tau=\mu \wedge t$, for any $t \geq t_{0}$, we have

$$
\begin{align*}
V(x(\mu \wedge t))-V\left(x_{0}\right)= & V(x(\tau))-V(x(\tau-1)) \\
& +V(x(\tau-1))-V(x(\tau-2)) \\
& +\cdots+V\left(x\left(t_{0}+1\right)\right)-V\left(x_{0}\right) \\
= & \sum_{t=t_{0}}^{\tau-1} \Delta V(x(t)) . \tag{23}
\end{align*}
$$

Taking the expectation on both sides, it is easy to see that

$$
\begin{equation*}
E V(x(\mu \wedge t)) \leq V\left(x_{0}\right) \tag{24}
\end{equation*}
$$

If $\mu \leq t$ and we note that $|x(\mu \wedge t)|=|x(\mu)|=h$, then

$$
\begin{equation*}
\varphi(h) P\{\mu \leq t\} \leq E\left[I_{\{\mu \leq t\}} V(x(\mu))\right] \leq E V(x(\mu \wedge t)) . \tag{25}
\end{equation*}
$$

From (21) and (24), we achieve that

$$
\begin{equation*}
P\{\mu \leq t\} \leq \varepsilon \tag{26}
\end{equation*}
$$

Letting $t \rightarrow+\infty$, then $P\{\mu<\infty\} \leq \varepsilon$; that is,

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad t \geq t_{0} . \tag{27}
\end{equation*}
$$

Therefore, the trivial solution of (6) is stochastically stable.

Theorem 13. If there exists a positive definite and infinite small upper bounded function $V(x)$, such that

$$
\begin{equation*}
E[\Delta V(x(t))] \leq 0 \tag{28}
\end{equation*}
$$

then the trivial solution of (6) is stochastically uniformly stable in probability.

Proof. By the assumptions, there exist $\varphi \in K$ and $\psi \in K$, such that

$$
\begin{equation*}
\varphi(|x|) \leq V(x) \leq \psi(|x|), \quad \forall x \in D_{r} \tag{29}
\end{equation*}
$$

Let $\varepsilon \in(0,1)$ and $h>0$ be arbitrary. Without loss of generality, we may assume that $h<r$. We define

$$
\begin{equation*}
\psi(\delta)=\varphi_{h}(\varepsilon) . \tag{30}
\end{equation*}
$$

Because of $\varphi \in K$, we can obtain $\delta=\psi^{-1}\left(\varphi_{h}(\varepsilon)\right)$, and it has nothing to do with $t_{0}$.

Similar to the proof of Theorem 12, Theorem 13 is established.

Remark 14. We note that $E[\Delta V(x(t))] \leq 0$ in Theorems 12-13 corresponds to $\mathscr{L} V(x) \leq 0$ in the Itô systems. In Theorem 13, $V(\cdot)$ is not only a positive function, but it is also an infinite small upper bounded function; this is because Theorem 13 is stronger than Theorem 12.

Theorem 15. If there exist a function $\varphi \in K$ and a positive definite function $V(x) \in C^{2}\left(D_{r}\right)$, such that $E[\Delta V(x(t)] \leq$ $-E \varphi(|x(t)|)$ for all $x(t) \in D_{r}$, then the trivial solution of (6) is stochastically asymptotically stable in probability.

Proof. From Theorem 12, we have that the trivial solution of (6) is stochastically stable. Fix $\varepsilon \in(0,1)$ arbitrarily; then there is $\delta_{0}=\delta_{0}\left(\varepsilon, t_{0}\right)>0$, such that

$$
\begin{equation*}
P\left\{|x(t)|<\frac{r}{2}\right\} \geq 1-\frac{\varepsilon}{4} \tag{31}
\end{equation*}
$$

when $x_{0} \in D_{\delta_{0}}$.

Fix $x_{0} \in D_{\delta_{0}}$ arbitrarily. By the assumptions on function $V(x)$, we know that $V(0)=0$ and that there exist two functions $\varphi_{1}, \varphi \in K$, such that

$$
\begin{gather*}
\varphi_{1}(|x|) \leq V(x),  \tag{32}\\
E[\Delta V(x(t))] \leq-E \varphi(|x(t)|), \quad \forall x \in D_{r} .
\end{gather*}
$$

Let $0<\beta<\left|x_{0}\right|$ arbitrarily, and choose $0<\alpha<\beta, 0<$ $\eta<\alpha$ sufficiently small; because of $V(x)$ being continuous, we can find that $0<\delta=\delta\left(\varepsilon, t_{0}\right)<\delta_{0}$, such that

$$
\begin{equation*}
V(x) \leq \frac{\varepsilon}{4} \varphi(\eta), \quad \forall x \in D_{\delta} . \tag{33}
\end{equation*}
$$

Define the stopping times

$$
\begin{align*}
& \mu_{\alpha}=\inf \left\{t \geq t_{0}:|x(t)| \leq \alpha\right\} \\
& \mu_{r}=\inf \left\{t \geq t_{0}:|x(t)| \geq \frac{r}{2}\right\} \tag{34}
\end{align*}
$$

Choose $\theta$ sufficiently large, such that

$$
\begin{equation*}
P\left\{\mu_{\alpha}<\theta\right\} \geq 1-\frac{\varepsilon}{2} \tag{35}
\end{equation*}
$$

Let $\tau=\mu_{\alpha} \wedge \mu_{r} \wedge t$, for any $t \geq t_{0}$, we have

$$
\begin{align*}
0 \leq V(x(\tau))-V\left(x_{0}\right)= & V(x(\tau))-V(x(\tau-1)) \\
& +V(x(\tau-1))-V(x(\tau-2)) \\
& +\cdots+V\left(x\left(t_{0}+1\right)\right)-V\left(x_{0}\right) \\
= & \sum_{t=t_{0}}^{\tau-1} \Delta V(x(t)) . \tag{36}
\end{align*}
$$

Taking the expectation on both sides, we can derive that

$$
\begin{equation*}
0 \leq E V(x(\tau)) \leq V\left(x_{0}\right)-\varphi(\alpha)\left(\tau-t_{0}\right) \tag{37}
\end{equation*}
$$

Hence,

$$
\begin{align*}
\frac{V\left(x_{0}\right)}{\varphi(\alpha)} & \geq E\left(\mu_{\alpha} \wedge \mu_{r} \wedge t-t_{0}\right)  \tag{38}\\
& =E\left(\tau-t_{0}\right) \geq\left(t-t_{0}\right) P\left\{\mu_{\alpha} \wedge \mu_{r} \geq t\right\}
\end{align*}
$$

This means that

$$
\begin{equation*}
P\left\{\mu_{\alpha} \wedge \mu_{r}<\infty\right\}=1 \tag{39}
\end{equation*}
$$

By (31), $P\left\{\mu_{r}<\infty\right\} \leq \varepsilon / 4$. So

$$
\begin{align*}
& P\left\{\mu_{\alpha}<\infty\right\}+\frac{\varepsilon}{4} \geq P\left\{\mu_{\alpha}<\infty\right\}+P\left\{\mu_{r}<\infty\right\}  \tag{40}\\
& \leq P\left\{\mu_{\alpha} \wedge \mu_{r}<\infty\right\}=1
\end{align*}
$$

which implies that

$$
\begin{equation*}
1-\frac{\varepsilon}{4} \leq P\left\{\mu_{\alpha}<\infty\right\} \tag{41}
\end{equation*}
$$

Hence,

$$
\begin{align*}
P\left\{\mu_{\alpha}<\mu_{r} \wedge \theta\right\} & \geq P\left(\left\{\mu_{\alpha}<\theta\right\} \cap\left\{\mu_{r}=\infty\right\}\right) \\
& \geq P\left\{\mu_{\alpha}<\theta\right\}-P\left\{\mu_{r}<\infty\right\} \geq 1-\frac{3}{4} \varepsilon . \tag{42}
\end{align*}
$$

Define the two stopping times

$$
\begin{align*}
\sigma & = \begin{cases}\mu_{\alpha} & \text { if } \mu_{\alpha}<\mu_{r} \wedge \theta, \\
\infty & \text { otherwise },\end{cases}  \tag{43}\\
\mu_{\beta} & =\inf \{t>\sigma:|x(t)| \geq \beta\} .
\end{align*}
$$

Similar to the proof of (24), we can show that, for $t \geq \theta$,

$$
\begin{equation*}
E V(x(\sigma \wedge t)) \geq E V\left(x\left(\mu_{\beta} \wedge t\right)\right) \tag{44}
\end{equation*}
$$

If $\mu_{\alpha} \geq \mu_{r} \wedge \theta$, then we note that $\left|x\left(\mu_{\beta} \wedge t\right)\right|=|x(\sigma \wedge t)|=$ $|x(t)|=\eta$, then

$$
\begin{equation*}
E\left[I_{\left\{\mu_{\alpha}<\mu_{r} \wedge \theta\right\}} V\left(x\left(\mu_{\alpha}\right)\right)\right] \geq E\left[I_{\left\{\mu_{\alpha}<\mu_{r} \wedge \theta\right\}} V\left(x\left(\mu_{\beta} \wedge t\right)\right)\right] . \tag{45}
\end{equation*}
$$

By (31) and $\left\{\mu_{\alpha}<\mu_{r} \wedge \theta\right\} \supset\left\{\mu_{\beta} \leq t\right\}$, we have

$$
\begin{align*}
\varphi_{1}(\eta) P\left\{\mu_{\beta} \leq t\right\} & \leq E\left[I_{\left\{\mu_{\beta} \wedge t\right\}} V\left(x\left(\mu_{\beta} \wedge t\right)\right)\right]  \tag{46}\\
& \leq E V\left(x\left(\mu_{\beta} \wedge t\right)\right)
\end{align*}
$$

Together with (33), we get

$$
\begin{equation*}
\frac{\varepsilon}{4} \geq P\left\{\mu_{\beta} \leq t\right\} \tag{47}
\end{equation*}
$$

Letting $t \rightarrow \infty$, we obtain

$$
\begin{equation*}
\frac{\varepsilon}{4} \geq P\left\{\mu_{\beta}<\infty\right\} \tag{48}
\end{equation*}
$$

By (42), it follows that

$$
\begin{align*}
1-\varepsilon & \leq P\left\{\mu_{\alpha}<\mu_{h} \wedge \theta\right\}-P\left\{\mu_{\beta}<\infty\right\}  \tag{49}\\
& \leq P\left\{\sigma<\infty, \mu_{\beta}=\infty\right\} .
\end{align*}
$$

This means that

$$
\begin{equation*}
P\left\{\lim _{t \rightarrow \infty} \sup |x(t)| \leq \beta\right\} \geq 1-\varepsilon \tag{50}
\end{equation*}
$$

Since $\beta$ is arbitrary, then we have

$$
\begin{equation*}
P\left\{\lim _{t \rightarrow+\infty} x(t)=0\right\} \geq 1-\varepsilon \tag{51}
\end{equation*}
$$

Theorem 16. If there exist a function $\varphi \in K$ and a positive definite, infinite small upper bounded function $V(x)$, such that $E\left[\Delta V(x(t)] \leq-E \varphi(|x(t)|)\right.$, for all $x(t) \in D_{r}$, then the trivial solution of (6) is stochastically uniformly asymptotically stable in probability.

Proof. By the assumptions, there exist $\varphi_{1}, \varphi_{2}$, and $\varphi \in K$, such that

$$
\begin{gather*}
\varphi_{1}(|x|) \leq V(x) \leq \varphi_{2}(|x|), \\
E[\Delta V(x(t))] \leq-E \varphi(|x(t)|) . \tag{52}
\end{gather*}
$$

From Theorem 13, we know that the trivial solution of (6) is stochastically uniformly stable. Therefore, for every $\varepsilon>0$ and $h>0$, there exists $\delta=\delta(\varepsilon, h)>0$, such that

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad \forall t \geq t_{0},\left|x_{0}\right|<\delta . \tag{53}
\end{equation*}
$$

According to Definition 3 we only need to show that, for every $\varepsilon>0$ and $h>0$, there exist $\delta=\delta(\varepsilon, h)>0$ and $T(\varepsilon)>0$, such that
$P\{|x(t)|<h\} \geq 1-\varepsilon, \quad \forall t \geq t_{0}+T(\varepsilon)$, whenever $\left|x_{0}\right|<\delta$.

We use a contradiction argument; take $T(\varepsilon)=\varphi_{2}\left(\delta_{0}\right) /$ $\varphi(\delta(\varepsilon, h))$; suppose, for any $\tau$, that $t_{0} \leq \tau \leq t_{0}+T(\varepsilon)$, such that $|x(\tau)| \geq \delta(\varepsilon, h)$. By $E[\Delta V(x(t)] \leq-\varphi(|x(t)|)$, we can show that

$$
\begin{equation*}
E[\Delta V(x(t))] \leq-\varphi(|x(t)|) \leq-\varphi(\delta(\varepsilon, h)) \tag{55}
\end{equation*}
$$

So

$$
\begin{align*}
\sum_{t=t_{0}}^{\tau-1} E[\Delta V(x(t))] & \leq \sum_{t=t_{0}}^{\tau-1}-\varphi(\delta(\varepsilon, h))  \tag{56}\\
& =-\varphi(\delta(\varepsilon, r))\left(\tau-t_{0}\right)
\end{align*}
$$

That is,

$$
\begin{equation*}
E V(x(\tau))-V\left(x_{0}\right) \leq-\varphi(\delta(\varepsilon, h))\left(\tau-t_{0}\right) . \tag{57}
\end{equation*}
$$

Thus,

$$
\begin{align*}
E V(x(\tau)) & \leq V\left(x_{0}\right)-\varphi(\delta(\varepsilon, h))\left(\tau-t_{0}\right) \\
& \leq \varphi_{2}\left(\left|x_{0}\right|\right)-\varphi(\delta(\varepsilon, h))\left(\tau-t_{0}\right)  \tag{58}\\
& <\varphi_{2}\left(\delta_{0}\right)-\varphi(\delta(\varepsilon, h))\left(\tau-t_{0}\right),
\end{align*}
$$

whenever $t_{0} \leq \tau \leq t_{0}+T(\varepsilon)$.
Especially, if $\tau=t_{0}+T(\varepsilon)$, it follows that

$$
\begin{equation*}
E V\left(x\left(t_{0}+T(\varepsilon)\right)\right)<\varphi_{2}\left(\delta_{0}\right)-\varphi(\delta(\varepsilon, h)) T(\varepsilon)=0 . \tag{59}
\end{equation*}
$$

This contradicts the positive definite property of $V(x)$. Then, we can prove that there exists $t_{1} \in\left[t_{0}, t_{0}+T(\varepsilon)\right]$, such that

$$
\begin{equation*}
\left|x\left(t_{1}\right)\right|<\delta(\varepsilon, h) \tag{60}
\end{equation*}
$$

According to Definition 3, we have

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad t \geq t_{1} . \tag{61}
\end{equation*}
$$

Therefore,

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad t \geq t_{0}+T(\varepsilon) . \tag{62}
\end{equation*}
$$

The proof is complete.

Remark 17. By comparing Theorems 12-15, we know that $E[\Delta V(x(t))] \leq-E \varphi(|x(t)|)$ guarantees the system to be stochastically asymptotically stable. The difference between Theorems 15 and 16 is that $V(x)$ is additionally required to have an infinite small upper bound in Theorem 16, which ensures the trivial solution of (6) to be stochastically uniformly asymptotically stable in probability.

Theorem 18. If there exist a function $\varphi \in K$ and a positive definite radially unbounded function $V(x) \in C^{2}\left(D_{r}\right)$, such that $E\left[\Delta V(x(t)] \leq-E \varphi(|x(t)|)\right.$, for all $x(t) \in D_{r}$, then the trivial solution of (6) is stochastically asymptotically stable in the large.

Proof. By Theorem 12, we know that the trivial solution of (6) is stochastically stable.

Let $\varepsilon \in(0,1)$ be arbitrary, and fix any $x_{0}$. Since $V(x)$ is radially unbounded, then we can choose $r>\left|x_{0}\right|$ sufficiently large, such that

$$
\begin{equation*}
\inf _{|x| \geq r, t \geq t_{0}} V(x) \geq \frac{4 V\left(x_{0}\right)}{\varepsilon} \tag{63}
\end{equation*}
$$

Define the stopping time

$$
\begin{equation*}
\mu_{r}=\inf \left\{t \geq t_{0}:|x(t)| \geq r\right\} \tag{64}
\end{equation*}
$$

Similar to the proof of (24), we can obtain that, for any $t \geq t_{0}$,

$$
\begin{equation*}
V\left(x_{0}\right) \geq E V\left(x\left(\mu_{r} \wedge t\right)\right) \tag{65}
\end{equation*}
$$

From (63), we have

$$
\begin{equation*}
V\left(x_{0}\right) \geq E V\left(x\left(\mu_{r} \wedge t\right)\right) \geq \frac{4 V\left(x_{0}\right)}{\varepsilon} P\left\{\mu_{r} \leq t\right\} \tag{66}
\end{equation*}
$$

Together with (65), it yields that

$$
\begin{equation*}
P\left\{\mu_{r} \leq t\right\} \leq \frac{\varepsilon}{4} \tag{67}
\end{equation*}
$$

Let $t \rightarrow \infty$; we have $P\left\{\mu_{r}<\infty\right\} \leq \varepsilon / 4$. That is to say,

$$
\begin{equation*}
P\{|x(t)| \leq r\} \geq 1-\frac{\varepsilon}{4}, \quad \forall t \geq t_{0} . \tag{68}
\end{equation*}
$$

In the same way as that of the proof of Theorem 15, we can show that

$$
\begin{equation*}
P\left\{\lim _{t \rightarrow+\infty} x(t)=0\right\} \geq 1-\varepsilon \tag{69}
\end{equation*}
$$

This immediately implies that $P\left\{\lim _{t \rightarrow+\infty} x(t)=0\right\}=1$. The proof is complete.

Theorem 19. If there exist a function $\varphi \in K$ and a positive definite, infinite small upper bound and radially unbounded function $V(x)$, such that $E[\Delta V(x(t)] \leq-E \varphi(|x(t)|)$, for all $x(t) \in D_{r}$, then the trivial solution of (6) is stochastically uniformly asymptotically stable in the large in probability.

Proof. Under the conditions of Theorem 19, there exist $\varphi_{1}, \varphi_{2}$, and $\varphi \in K$, such that

$$
\begin{gather*}
\varphi_{1}(|x|) \leq V(x) \leq \varphi_{2}(|x|), \\
E[\Delta V(x(t))] \leq-E \varphi(|x(t)|) . \tag{70}
\end{gather*}
$$

By Theorem 13, we know that the trivial solution of (6) is stochastically uniformly stable.

In the following, we first verify that the trivial solution of (6) is uniformly bounded. Actually, for any $\alpha>0, t_{0} \geq 0$, due to $\varphi_{1}, \varphi_{2} \in K$, there exists $\beta=\beta(\alpha)$, such that

$$
\begin{equation*}
\varphi_{1}(\beta)=\varphi_{2}(\alpha), \quad \text { that is, } \beta=\varphi_{1}^{-1}\left(\varphi_{2}(\alpha)\right) \tag{71}
\end{equation*}
$$

It is easy to show that

$$
\begin{equation*}
E V(x(t))-V\left(x_{0}\right)=\sum_{t=t_{0}}^{t-1} E[\Delta V(x(t))] \leq 0, \quad t \geq t_{0} \tag{72}
\end{equation*}
$$

When $\left|x_{0}\right|<\alpha$, we have

$$
\begin{align*}
\varphi_{1}(|x(t)|) & \leq E V(x(t)) \leq V\left(x_{0}\right) \\
& \leq \varphi_{2}\left(\left|x_{0}\right|\right)<\varphi_{2}(\alpha)  \tag{73}\\
& =\varphi_{1}(\beta) .
\end{align*}
$$

Because of $\varphi_{1}$ being strictly increasing, so $|x(t)|<\beta$, a.s., $t \geq t_{0}$. This implies that the trivial solution of (6) is uniformly bounded.

We further show that, for every $\alpha>0, \varepsilon>0$, and $h>0$, there exists $T(\varepsilon, \alpha)>0$, such that

$$
\begin{array}{r}
P\{|x(t)|<h\} \geq 1-\varepsilon,  \tag{74}\\
\forall t \geq t_{0}+T(\varepsilon, \alpha), \text { whenever }\left|x_{0}\right|<\alpha .
\end{array}
$$

As previously stated the trivial solution of (6) is stochastically uniformly stable. Therefore, for every $\varepsilon>0$ and $h>0$, there exists $\delta=\delta(\varepsilon, h)>0$, such that

$$
\begin{equation*}
P\{|x(t)|<h\} \geq 1-\varepsilon, \quad \forall t \geq t_{0},\left|x_{0}\right|<\delta . \tag{75}
\end{equation*}
$$

The rest is similar to the proof of Theorem 16 and is thus omitted.

Remark 20. Theorems 18 and 19 are stronger versions of Theorems 15 and 16 , respectively, where $V(x)$ is additionally required to be a radially unbounded function that is used to prove the stability in the large.

In what follows, we will discuss the $p$ th moment exponential stability for (6).

Theorem 21. Suppose that there exist a function $V(x) \in$ $C^{2}\left(D_{r}\right)$ and positive constants $c_{1}, c_{2}$, and $c_{3}$, such that

$$
\begin{gather*}
c_{1}|x|^{p} \leq V(x) \leq c_{2}|x|^{p}, \\
E[\Delta V(x(t))] \leq-c_{3} E V(x(t)) \tag{76}
\end{gather*}
$$

Then

$$
\begin{equation*}
E|x(t)|^{p} \leq \frac{c_{2}}{c_{1}}\left|x_{0}\right|^{p} e^{-c_{3}\left(t-t_{0}\right)}, \quad t \geq t_{0}, \quad \forall x_{0} \in R^{n} \tag{77}
\end{equation*}
$$

That is, the trivial solution of (6) is pth moment exponentially stable.

Proof. Define the stopping time

$$
\begin{equation*}
\mu_{n}=\inf \left\{t \geq t_{0}:|x(t)| \geq n\right\}, \quad n \geq\left|x_{0}\right| \tag{78}
\end{equation*}
$$

It is easy to see that $\mu_{n} \rightarrow \infty$ as $n \rightarrow \infty$ almost surely. By $E[\Delta V(x(t))] \leq-c_{3} E V(x(t))$, we can derive that

$$
\begin{align*}
& E\left[e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}\right)\right)\right]-V\left(x_{0}\right) \\
&=E\left[e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}\right)\right)\right. \\
&-e^{c_{3}\left(t \wedge-1-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}-1\right)\right) \\
&+e^{c_{3}\left(t \wedge \mu_{n}-1-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}-1\right)\right) \\
& \quad-e^{c_{3}\left(t \wedge \tau_{n}-2-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}-2\right)\right) \\
&\left.+\cdots+e^{c_{3}\left(t_{0}+1-t_{0}\right)} V\left(x\left(t_{0}+1\right)\right)\right]-V\left(x_{0}\right) \\
&=E\left[e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}\right)\right)\right. \\
& \quad-V\left(x\left(t \wedge \mu_{n}-1\right)\right) \\
&+e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}-1\right)\right) \\
& \quad-e^{c_{3}\left(t \wedge \mu_{n}-1-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}-1\right)\right) \\
&\left.\quad+\cdots+e^{c_{3}} V\left(x\left(t_{0}+1\right)\right)\right]-V\left(x_{0}\right) \\
&=e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} E\left[\Delta V\left(x\left(t \wedge \mu_{n}-1\right)\right)\right] \\
& \quad+e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}-1\right)\right)\left(1-\frac{1}{e^{c_{3}}}\right) \\
& \quad+\cdots+e^{c_{3}} E\left[\Delta V\left(x\left(t_{0}\right)\right)\right]+e^{c_{3}} V\left(x_{0}\right)\left(1-\frac{1}{e^{c_{3}}}\right) \\
& \leq e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)}\left(-c_{3}+1-\frac{1}{e^{c_{3}}}\right) E V\left(x\left(t \wedge \mu_{n}-1\right)\right) \\
& \quad+\cdots+e^{c_{3}}\left(-c_{3}+1-\frac{1}{\left.e^{c_{3}}\right) V(x) \leq 0}\right. \tag{79}
\end{align*}
$$

By $c_{1}|x|^{p} \leq V(x) \leq c_{2}|x|^{p}$, we have that

$$
\begin{align*}
& c_{1} e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} E\left|x\left(t \wedge \mu_{n}\right)\right|^{p} \\
& \quad \leq E\left[e^{c_{3}\left(t \wedge \mu_{n}-t_{0}\right)} V\left(x\left(t \wedge \mu_{n}\right)\right)\right]  \tag{80}\\
& \quad \leq V\left(x_{0}\right) \leq c_{2}\left|x_{0}\right|^{p} .
\end{align*}
$$

Letting $n \rightarrow \infty$, then

$$
\begin{equation*}
c_{1} e^{c_{3}\left(t-t_{0}\right)} E|x(t)|^{p} \leq c_{2}\left|x_{0}\right|^{p} \tag{81}
\end{equation*}
$$

which implies (77).

As a corollary, Theorem 21 yields a sufficient criterion for the exponential stability in mean square sense.

Corollary 22. Suppose that there exist a function $V(x) \in$ $C^{2}\left(D_{r}\right)$ and positive constants $c_{1}, c_{2}$, and $c_{3}$, such that

$$
\begin{gather*}
c_{1}|x|^{2} \leq V(x) \leq c_{2}|x|^{2}, \\
E[\Delta V(x(t))] \leq-c_{3} E|x(t)|^{2} . \tag{82}
\end{gather*}
$$

Then the trivial solution of (6) is exponentially stable in mean square.

## 4. Illustrative Examples

In this section, we present three simple examples to illustrate applications of the stability results developed in this paper. We will let $w(t)$ be a one-dimensional stochastic process defined on the complete probability space $(\Omega, F, P)$, such that $E w(t)=$ 0 and $E[w(t) w(s)]=\delta_{s t}$, where $\delta_{s t}$ is the Kronecker delta.

Example 1. Consider the following equation:

$$
\begin{equation*}
x(t+1)=A x(t)+B x(t) w(t), \quad x\left(t_{0}\right)=x_{0} \tag{83}
\end{equation*}
$$

where $A$ and $B$ are $n \times n$ matrices. Assume that there is a symmetric positive definite matrix $P$, such that

$$
\begin{equation*}
A^{T} P A+B^{T} P B-P \leq 0 \tag{84}
\end{equation*}
$$

Now, define the stochastic Lyapunov function $V(x)=x^{T} P x$. It is obvious that

$$
\begin{equation*}
E[\Delta V(x(t))]=E\left[x^{T}(t)\left(A^{T} P A+B^{T} P B-P\right) x(t)\right] \leq 0 . \tag{85}
\end{equation*}
$$

By Theorem 12, we conclude that the trivial solution $x \equiv 0$ of (83) is stochastically stable in probability.

Example 2. Consider the following stochastic difference equation:

$$
\begin{align*}
x(t+1) & =A(t) x(t)+B(t) x(t) w(t) \\
& =[A(t)+B(t) w(t)] x(t)  \tag{86}\\
& =H(t, w(t)) x(t),
\end{align*}
$$

where $A(t), B(t)$, and $H(t, w(t))=A(t)+B(t) w(t)=\left(h_{i, j}(t\right.$, $w(t))$ ) are all $2 \times 2$ matrix-valued functions defined on $t=$ $t_{0}, t_{0}+1$, and $t_{0}+2, \ldots, x\left(t_{0}\right)=x_{0} \in R^{n}$. Assume that

$$
\begin{equation*}
\max _{i=1,2} E\left\{\sum_{j=1}^{2}\left|h_{i j}(t, w(t))\right|^{2}\right\}<\frac{1}{2} \tag{87}
\end{equation*}
$$

for all $x(t) \in R^{2}$.

We define the Lyapunov function $V(x)=\max _{i=1,2}\left\{\left|x_{i}\right|^{2}\right\}$. It is positive definite and radially unbounded. Moreover,

$$
\begin{align*}
E V(x(t+1))= & \max _{i=1,2} E\left\{\left|\sum_{j=1}^{2} h_{i j}(t, w(t)) x_{j}(t)\right|^{2}\right\} \\
\leq & \max _{i=1,2} E\left\{\sum_{j=1}^{2} \mid h_{i j}\left(t,\left.w(t)\right|^{2} \sum_{j=1}^{2}\left|x_{j}(t)\right|^{2}\right\}\right. \\
\leq & \max _{i=1,2} E\left\{\sum_{j=1}^{2} \mid h_{i j}\left(t,\left.w(t)\right|^{2}\right\}\right. \\
& \times \max _{j=1,2} E\left\{2\left|x_{j}(t)\right|^{2}\right\} \\
< & \max _{j=1,2} E\left\{\left|x_{j}(t)\right|^{2}\right\}=E V(x(t)) . \tag{88}
\end{align*}
$$

That is, $E[\Delta V(x(t))]<0$. By Theorem 18, the trivial solution is stochastically asymptotically stable in the large.

Example 3. Consider a one-dimensional linear stochastic difference equation

$$
\begin{equation*}
x(t+1)=a x(t)+b x(t) w(t) \tag{89}
\end{equation*}
$$

where $a, b$ are all constants, and $a^{2}<b^{2} / 4$. We assume that there exist positive constants $p$ and $c<1$, such that $5 b^{2}<$ $4(1-c) p$.

We define the Lyapunov function $V(x)=p x^{2}$; then

$$
\begin{equation*}
E[\Delta V(x(t))]=\left(\frac{5 b^{2}}{4}-p\right) E x^{2}(t)<-c p E x^{2}(t) \tag{90}
\end{equation*}
$$

By Corollary 22, the trivial solution is exponentially stable in mean square.

## 5. Conclusions

This paper has discussed the stability in probability for stochastic discrete-time systems. Using the method of Lyapunov functionals, some efficient criteria for the stability are obtained. Some results of the stability [13] for stochastic differential equations are generalized to stochastic discretetime systems. There are some interesting problems such as the almost sure exponential stability and the stochastic nonlinear $H_{\infty}$ control that merit further study.

## Acknowledgments

This work is supported by NSF of China (Grants nos. 61174078 and 61170054), Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20103718110006), the Research Fund for the Taishan Scholar Project of Shandong Province of China, the SDUST Research Fund (Grant no. 2011KYTD105), and the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (Grant no. LAPS13018).

## References

[1] D. J. N. Limebeer, B. D. O. Anderson, and B. Hendel, "A Nash game approach to mixed $H_{2} / H_{\infty}$ control," IEEE Transactions on Automatic Control, vol. 39, no. 1, pp. 69-82, 1994.
[2] D. Hinrichsen and A. J. Pritchard, "Stochastic $H^{\infty}$," SIAM Journal on Control and Optimization, vol. 36, no. 5, pp. 15041538, 1998.
[3] W. Zhang and B.-S. Chen, "On stabilizability and exact observability of stochastic systems with their applications," Automatica, vol. 40, no. 1, pp. 87-94, 2004.
[4] B.-S. Chen and W. Zhang, "Stochastic $H_{2} / H_{\infty}$ control with state-dependent noise," IEEE Transactions on Automatic Control, vol. 49, no. 1, pp. 45-57, 2004.
[5] V. Dragan, T. Morozan, and A.-M. Stoica, Mathematical Methods in Robust Control of Discrete-Time Linear Stochastic Systems, Springer, New York, NY, USA, 2010.
[6] L. Arnold, Stochastic Differential Equations: Theory and Applications, Wiley-Interscience, New York, NY, USA, 1972.
[7] A. Friedman, Stochastic Differential Equations and Their Applications, vol. 2, Academic Press, San Diego, Calif, USA, 1976.
[8] R. Z. Has'minskii, Stochastic Stability of Differential Equations, vol. 7 of Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis, Sijthoff \& Noordhoff, Rockville, Md, USA, 1980.
[9] H. J. Kushner, Stochastic Stability and Control, Academic Press, New York, NY, USA, 1967.
[10] V. B. Kolmanovskii and A. Myshkis, Applied Theory of Functional Differential Equations, Kluwer Academic Publishers, Norwell, Mass, USA, 1992.
[11] G. S. Ladde and V. Lakshmikantham, Random Differential Inequalities, vol. 150 of Mathematics in Science and Engineering, Academic Press, New York, NY, USA, 1980.
[12] S. E. A. Mohammed, Stochastic Functional Differential Equations, Longman, New York, NY, USA, 1986.
[13] X. R. Mao, Stochastic Differential Equations and Applications, Horwood, Chichester, UK, 1997.
[14] V. Dragan and T. Morozan, "Mean square exponential stability for some stochastic linear discrete time systems," European Journal of Control, vol. 12, no. 4, pp. 373-399, 2006.
[15] Y. Huang, W. Zhang, and H. Zhang, "Infinite horizon linear quadratic optimal control for discrete-time stochastic systems," Asian Journal of Control, vol. 10, no. 5, pp. 608-615, 2008.
[16] T. Taniguchi, "Stability theorems of stochastic difference equations," Journal of Mathematical Analysis and Applications, vol. 147, no. 1, pp. 81-96, 1990.
[17] B. Paternoster and L. Shaikhet, "About stability of nonlinear stochastic difference equations," Applied Mathematics Letters, vol. 13, no. 5, pp. 27-32, 2000.

# Output-Feedback and Inverse Optimal Control of a Class of Stochastic Nonlinear Systems with More General Growth Conditions 

Liu Jianwei, Guo Longchuan, Zuo Xin, and Liang Huaqing<br>Department of Automation, China University of Petroleum, Beijing 102249, China<br>Correspondence should be addressed to Liu Jianwei; liujw@cup.edu.cn

Received 8 May 2013; Accepted 3 June 2013
Academic Editor: Baocang Ding
Copyright © 2013 Liu Jianwei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

This paper investigates the problem of output-feedback stabilization for a class of stochastic nonlinear systems in which the nonlinear terms depend on unmeasurable states besides measurable output. We extend linear growth conditions to power growth conditions and reduce the control effort. By using backstepping technique, choosing a high-gain parameter, an output-feedback controller is designed to ensure the closed-loop system to be globally asymptotically stable in probability, and the inverse optimal stabilization in probability is achieved. The efficiency of the output-feedback controller is demonstrated by a simulation example.


## 1. Introduction

The design of output-feedback controller for stochastic nonlinear systems has achieved remarkable research development, because output feedback control is more suitable for practical engineering systems; for example, see [1-12] and references therein. In recent years such research hotspot has mainly focused on a class of special nonlinear stochastic systems in which the nonlinear vector terms depend on the unmeasurable states besides the measurable output; for example, see [13-17] and references therein. The work of [13] discussed the output-feedback controller design by introducing a stability concept named globally asymptotically stable in probability. Based on the purpose of reducing the amount of control, [14] considered the output-feedback stabilization problem.

However, in [13-17], the nonlinear vector terms satisfy the linear growth conditions strictly, which greatly narrows the scope of application of the research results. Naturally, one may ask about an interesting and challenging problem: can we further relax the linear growth conditions? To our knowledge, the existing research results on this problem are as in [18-20]. In [18], authors discussed the outputfeedback stabilization problem by introducing a rescaling transformation under more relaxed growth conditions. On
the basis of [18], the work of [19] and [20] further considered the output-feedback controller design problem for highorder stochastic nonlinear systems. However, for [18-20], the observer gain $K$ is usually larger than 1 , and the choice can lead to a controller design which needs larger control effort. So another challenging problem is proposed that is whether the assumption $K>1$ can be removed.

In this paper, we investigate the output-feedback stabilization problem for a class of stochastic nonlinear systems satisfying power growth conditions. Inspired by [13, 14], we find the maximum value interval of observer gain for the desired controller by using backstepping technique. For this interval, the designed output-feedback controller ensures that the equilibrium at the origin is globally asymptotically stable in probability and the inverse optimal stabilization in probability is achieved. The main contributions of this paper are characterized as follows. (i) We extend the linear growth conditions to the power growth conditions. (ii) The assumption of $K>1$ in [18-20] is removed so that we can get less control effort.

The paper is organized as follows. Section 2 provides some preliminary results. In Section 3, the problem to be investigated is presented. In Sections 4 and 5, an outputfeedback controller is designed and analysed. In Section 6, the inverse optimal stabilization in probability is achieved.

Section 7 provides a simulation example. Section 8 concludes this paper.

## 2. Notations and Preliminary Results

Throughout this paper, the following notations will be used. $R$ denotes the set of all real numbers; $R_{+}$denotes the set of all nonnegative real numbers; $R^{n}$ denotes the real $n$-dimensional space; $R^{n \times r}$ denotes the real $n \times r$ matrix space; $\operatorname{Tr}(\cdot)$ denotes the trace for square matrix $X ;|X|$ denotes the Euclidean norm of a vector $X$, and $\|X\|$ is the Frobenius norm of matrix $X$ defined by $\|X\|=\left(\sum_{i=1}^{n} \sum_{j=1}^{r} x_{i j}^{2}\right)^{1 / 2}$; for a given vector $x=\left(x_{1}, \ldots, x_{n}\right)^{T}, x_{[i]}$ denotes $\left(x_{1}, \ldots, x_{i}\right)^{T} ; C^{i}$ denotes the set of all function with continuous $i$ th partial derivatives; $C^{2,1}\left(R^{n} \times R_{+}, R_{+}\right)$is the family of all nonnegative functions $V(x, t)$ on $R^{n} \times R_{+}$, which are $C^{2}$ in $x$ and $C^{1}$ in $t ; K$ denotes the set of all functions: $R^{+} \rightarrow R^{+}$, which are continuous, strictly increasing, and vanish at zero; $K_{\infty}$ denotes the set of all functions which are of class $K$ and unbounded; $K L$ denotes the set of all functions $\beta(s, t): R_{+} \times R_{+} \rightarrow R_{+}$, which are of $K$ for each fixed $t$ and decrease to zero as $t \rightarrow \infty$ for each fixed $s$.

Lemma 1. The inequality $(|x|+|y|)^{p} \leq 2^{p-1}\left(|x|^{p}+|y|^{p}\right)$ is established for any $x \in R, y \in R, p \geq 1, p \in Z$.

Proof. For an assumption of $\boldsymbol{\alpha}=\left(a_{1}, a_{2}\right)=(|x|,|y|), \boldsymbol{\beta}=$ $\left(b_{1}, b_{2}\right)=(1,1)$, inspired by Holder inequality [21], we can get

$$
\begin{equation*}
|x|+|y| \leq\left(|x|^{p}+|y|^{p}\right)^{1 / p} 2^{1-1 / p} \tag{1}
\end{equation*}
$$

and further get

$$
\begin{equation*}
(|x|+|y|)^{p} \leq 2^{p-1}\left(|x|^{p}+|y|^{p}\right) \tag{2}
\end{equation*}
$$

Then the proof is completed.
Consider the following stochastic system:

$$
\begin{equation*}
d x=f(x, u) d t+g(x) d \omega, \quad \forall x_{0} \in R^{n} \tag{3}
\end{equation*}
$$

where $x=\left(x_{1}, \ldots, x_{n}\right) \in R^{n}$ is the state of the system. $u \in R^{m}$ is the control input of the system. $\omega$ is an $r$-dimensional standard Wiener process defined on a probability space $\{\Omega, F, P\}$. The nonlinear functions $f: R^{m+n} \rightarrow R^{n}$ and $g:$ $R^{n} \rightarrow R^{n \times r}$ are locally Lipschitz with $f(0)=0, g(0)=0$. For any given $V \in C^{2}\left(R^{n} ; R\right)$ associated with stochastic system (3), the differential operator $L$ is defined as follows:

$$
\begin{equation*}
L V(x)=\frac{\partial V}{\partial x} f(x)+\frac{1}{2} \operatorname{Tr}\left(g^{T}(x) \frac{\partial^{2} V}{\partial x^{2}} g(x)\right) \tag{4}
\end{equation*}
$$

In order to discuss the stability of stochastic nonlinear systems, we introduce the following stability notion.

Definition 2 (see [22]). For the stochastic nonlinear system (3) with $f(0, u)=0, g(0)=0$, the equilibrium $x(t)=0$ of (3) is said to be globally asymptotically stable (GAS) in probability if, for any $\varepsilon>0$, there exists a class $K L$ function $\beta(\cdot, \cdot)$ such that $P\left\{|x(t)|<\beta\left(\left|x_{0}\right|, t\right)\right\} \geq 1-\varepsilon, \forall t \geq 0$, $x_{0} \in R^{n} \backslash\{0\}$.

The following lemmas give some sufficient conditions ensuring global asymptotical stability in probability.

Lemma 3 (see [23]). For system (3), if there exist $V(x) \in C^{2}$, class $K_{\infty}$ functions $\alpha_{1}, \alpha_{2}$, and a class $K$ function $\alpha_{3}$ such that $\alpha_{1}(|x|) \leq V(x) \leq \alpha_{2}(|x|), L V(x) \leq-\alpha_{3}(|x|)$, then there exists an almost surely unique solution to system (3) on $[0, \infty)$, and the equilibrium $x(t)=0$ is globally asymptotically stable in probability.

Lemma 4 (see [24]). Consider the following control law:

$$
\begin{equation*}
u=\alpha(x)=-R_{2}^{-1}\left(L_{\varphi_{2}} V\right)^{T} \frac{\operatorname{l\gamma }\left(\left|L_{\varphi_{2}} V R_{2}^{-1 / 2}\right|\right)}{\left|L_{\varphi_{2}} V R_{2}^{-1 / 2}\right|^{2}} \tag{5}
\end{equation*}
$$

where $\gamma(\cdot)$ is a class $K_{\infty}$ function, $\ell \gamma(s)=s(\dot{\gamma})^{-1}(s)-$ $\gamma\left((\dot{\gamma})^{-1}(s)\right)$, and $R_{2}(x)$ is a matrix valued function such that $R_{2}(x)=R_{2}^{T}(x)>0$. If the control law (5) to be ensures system (3) globally asymptotically stable in probability, then the control law

$$
\begin{align*}
u^{*} & =\beta^{*}(x) \\
& =-\frac{\theta}{2} R_{2}^{-1}\left(L_{\varphi_{2}} V\right)^{T} \frac{\ell \gamma\left(\left|L_{\varphi_{2}} V R_{2}^{-1 / 2}\right|\right)}{\left|L_{\varphi_{2}} V R_{2}^{-1 / 2}\right|^{2}}, \quad \theta \geq 2 \tag{6}
\end{align*}
$$

solves the problem of inverse optimal stabilization in probability for system (3) by minimizing the cost functional

$$
\begin{equation*}
J(u)=E\left\{\int_{0}^{\infty}\left[l(x)+\theta^{2} \gamma\left(\frac{2}{\theta}\left|R_{2}(x)^{1 / 2} u\right|\right)\right] d \tau\right\} \tag{7}
\end{equation*}
$$

where $l(x)$ is a positive definite radially unbounded function satisfying

$$
\begin{align*}
l(x)= & 2 \theta\left[\ell \gamma\left(\left|L_{\varphi_{2}} V R_{2}^{-1 / 2}\right|\right)-\frac{1}{2} \operatorname{Tr}\left\{\varphi_{1}^{T} \frac{\partial^{2} V}{\partial x^{2}} \varphi_{1}\right\}\right]  \tag{8}\\
& +\theta(\theta-2) \ell \gamma\left(\left|L_{\varphi_{2}} V R_{2}^{-1 / 2}\right|\right)
\end{align*}
$$

## 3. Problem Formulation

Consider the following stochastic nonlinear systems:

$$
\begin{gather*}
d x_{i}=x_{i+1} d t+\varphi_{i}(x) d \omega, \quad i=1, \ldots, n-1 \\
d x_{n}=u d t+\varphi_{n}(x) d \omega  \tag{9}\\
y=x_{1}
\end{gather*}
$$

where $x=\left(x_{1}, \ldots, x_{n}\right) \in R^{n}, u \in R$, and $y \in R$ are the states, the control input, and the measurable output of the system, $\omega \in R^{r}$ is defined as in (3), and $x_{2}, \ldots, x_{n}$ are the unmeasurable states. $\varphi_{i}: R^{n} \rightarrow R^{r}, i=1, \ldots, n$, are locally Lipschitz with $\varphi_{i}(0)=0$ and satisfy the following power growth conditions.

Assumption 5. For each $1 \leq i \leq n$, there exists the known positive constant $d \geq 0$ such that $\left|\varphi_{i}(\mathbf{x})\right| \leq d\left(\left|x_{1}\right|^{p}+\left|x_{2}\right|^{p}+\right.$ $\cdots+\left|x_{i}\right|^{p}$ ), where $p$ is any positive integer.

Remark 6. Assumption 5 can be simplified into linear growth conditions when $p=1$. Therefore, linear growth conditions as a special case are included in Assumption 5. This paper extends previous work and gets a new result.

The objective of this paper is to design a smooth outputfeedback controller for system (9), such that the closedloop system is globally asymptotically stable in probability at the origin and achieves the design of the inverse optimal stabilization in probability.

## 4. Output-Feedback Controller Design

Since $x_{2}, \ldots, x_{n}$ are unmeasured, the following observer is introduced:

$$
\begin{gather*}
\dot{\hat{x}}_{i}=\widehat{x}_{i+1}+K^{i} h_{i}\left(x_{1}-\widehat{x}_{1}\right), \quad i=1, \ldots, n-1,  \tag{10}\\
\dot{\hat{x}}_{n}=u+K^{n} h_{n}\left(x_{1}-\widehat{x}_{1}\right),
\end{gather*}
$$

where $\widehat{x}_{i}$ is the estimated value of $x_{i}, K \in R_{+}$is the observer gain to be determined, and $h_{i}>0, i=1, \ldots, n$, are chosen such that matrix $A=\left(\begin{array}{ccc}-h_{1} & \\ \vdots & I_{n-1} \\ -h_{n} & 0 \cdots 0\end{array}\right)$ is asymptotically stable; thus there exists a positive definite matrix $P$ satisfying $A^{T} P+$ $P A=-I$. Let $\tilde{x}=\left(\widetilde{x}_{1}, \ldots, \widetilde{x}_{n}\right)^{T}$, where $\widetilde{x}_{i}=\left(x_{i}-\widehat{x}_{i}\right) / K^{i-1}$ for each $i=1, \ldots, n$. By (9) and (10), we can get error system

$$
\begin{equation*}
d \tilde{x}=K A \tilde{x} d t+\Phi(x) d \omega \tag{11}
\end{equation*}
$$

where $\Phi(x)=\left(\varphi_{1}(x), \varphi_{2}(x) / K, \ldots, \varphi_{n}(x) / K^{n-1}\right)^{T}$.
Now we give the backstepping controller design procedure.

Step 0 . Choosing the zeroth Lyapunov function $V_{0}(\widetilde{x})=(n+$ 1) $\tilde{x}^{T} P \tilde{x}$, applying $2 a b \leq 2\left(a^{2}+b^{2}\right),(a+b)^{2} \leq 2\left(a^{2}+b^{2}\right)$, $\sum_{i=1}^{n}\left|a_{i}\right|^{2} \leq\left(\sum_{i=1}^{n}\left|a_{i}\right|\right)^{2},\left(\sum_{i=1}^{n} a_{i}\right)^{2} \leq n \sum_{i=1}^{n} a_{i}^{2}$, Lemma 1, Assumption 5, and (4), we can get

$$
\begin{aligned}
L V_{0}= & -(n+1) K|\widetilde{x}|^{2}+(n+1) \operatorname{Tr}\left(\Phi^{T}(x) P \Phi(x)\right) \\
\leq & -(n+1) K|\widetilde{x}|^{2}+(n+1)\|P\|\left(\sum_{i=1}^{n}\left|\frac{\varphi_{i}(x)}{K^{i-1}}\right|^{2}\right) \\
\leq & -(n+1) K|\widetilde{x}|^{2}+(n+1)\|P\|\left(\sum_{i=1}^{n}\left|\frac{\varphi_{i}(x)}{K^{i-1}}\right|\right)^{2} \\
\leq & -(n+1) K|\widetilde{x}|^{2}+(n+1)\|P\| d^{2}\left(\sum_{i=1}^{n} \frac{1}{K^{i-1}}\right)^{2} \\
& \times\left(\left|x_{1}\right|^{p}+\frac{\left|x_{2}\right|^{p}}{K}+\cdots+\frac{\left|x_{n}\right|^{p}}{K^{n-1}}\right)
\end{aligned}
$$

$$
\begin{align*}
\leq & -(n+1) K|\widetilde{x}|^{2}+2^{2 p-1} n d^{*} \\
& \times\left(\sum_{i=1}^{n}\left(\frac{\widehat{x}_{i}^{p}}{K^{i-1}}\right)^{2}+\sum_{i=1}^{n}\left(K^{(i-1) p-(i-1)} \widetilde{x}_{i}^{p}\right)^{2}\right) \\
\leq & -\left((n+1) n K-2^{2 p-1} n d^{*} \sum_{i=1}^{n}\left(K^{(i-1) p-(i-1)}\right)^{2}\right)\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\widehat{x}_{1}^{2 p}+\frac{\widehat{x}_{2}^{2 p}}{K^{2}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right), \tag{12}
\end{align*}
$$

where

$$
\begin{align*}
d^{*} & =(n+1)\|P\| d^{2}\left(\sum_{i=1}^{n} \frac{1}{K^{i-1}}\right)^{2} \\
& =\frac{(n+1)\|P\| d^{2}\left(\sum_{i=0}^{n-1}(i+1) K^{i}+\sum_{i=n}^{2 n-2}(2 n-i-1) K^{i}\right)}{K^{2 n-2}} \tag{13}
\end{align*}
$$

and $\|\tilde{x}\|_{\infty}=\max _{i}\left|x_{i}\right|$.
We introduce a series of coordinate changes as follows:

$$
\begin{gather*}
w_{1}=\widehat{x}_{1}  \tag{14}\\
w_{i}=\widehat{x}_{i}-\beta_{i-1}\left(\widehat{x}_{[i-1]}\right),
\end{gather*}
$$

where $\beta_{i-1}\left(\widehat{x}_{[i-1]}\right)(i=2, \ldots, n)$ is the virtual control law to be designed.

Step 1. Constructing the 1st Lyapunov function

$$
\begin{equation*}
V_{1}\left(\widetilde{\mathbf{x}}, w_{1}\right)=V_{0}(\widetilde{\mathbf{x}})+\frac{1}{p+1} w_{1}^{p+1} \tag{15}
\end{equation*}
$$

using (3), (10), (12)-(15), and Young's inequality [22], we can obtain

$$
\begin{align*}
L V_{1} \leq & -\left((n+1) n K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right)\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{3}^{2 p}}{K^{4}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)+2^{2 p-1} n d^{*} w_{1}^{2 p} \\
& +2^{2 p-1} n d^{*} \frac{\widehat{x}_{2}^{2 p}}{K^{2}}+w_{1}^{p} w_{2}+w_{1}^{p} \beta_{1}+K \widetilde{x}_{1}^{2}+\frac{K}{4} h_{1}^{2} w_{1}^{2 p} . \tag{16}
\end{align*}
$$

Applying (14) and Lemma 1, choosing $K \geq 2^{2 p} n d^{*}$, we can get

$$
\begin{gather*}
2^{2 p-1} n d^{*} w_{1}^{2} \leq \frac{K}{2} w_{1}^{2} \\
2^{2 p-1} n d^{*} \frac{\widehat{x}_{2}^{2 p}}{K^{2}} \leq 4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+4^{2 p-1} n d^{*} \frac{1}{K^{2}} \beta_{1}^{2 p}, \tag{17}
\end{gather*}
$$

which one substitutes in (16) to obtain

$$
\begin{aligned}
& L V_{1} \leq-\left((n+1) n K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{3}^{2 p}}{K^{4}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)+2^{2 p-1} n d^{*} w_{1}^{2} \\
& +4^{2 P-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+4^{2 p-1} n d^{*} \frac{1}{K^{2}} \beta_{1}^{2 p} \\
& +w_{1}^{p} w_{2}+w_{1}^{p} \beta_{1}+K \tilde{x}_{1}^{2}+\frac{K}{4} h_{1}^{2} w_{1}^{2 p} \\
& \leq-\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{3}^{2 p}}{K^{4}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)+4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{1}^{2 p}+w_{1}^{p} w_{2}+w_{1}^{p} \beta_{1} \\
& +K\left(\frac{1}{2}+\frac{h_{1}^{2}}{4}\right) w_{1}^{2 p} \\
& \leq-\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{3}^{2 p}}{K^{4}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)-n K w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} v_{1}^{2 p} \sqrt{\underbrace{w_{1}^{2 p} \cdots w_{1}^{2 p}}_{2 p}} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+w_{1}^{p} w_{2} \\
& \leq-\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{3}^{2 p}}{K^{4}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)-n K w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} v_{1}^{2 p} \sqrt{\left(2 p w_{1}^{2 p}\right)^{2}} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+w_{1}^{p} w_{2}
\end{aligned}
$$

$$
\begin{align*}
= & -\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{3}^{2 p}}{K^{4}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\left(n K-2^{4 p-1} p n d^{*} v_{1}^{2 p}\right) w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+w_{1}^{p} w_{2} \tag{18}
\end{align*}
$$

by choosing the 1st virtual control law

$$
\begin{gather*}
\beta_{1}=-K v_{1} w_{1}^{p} \\
v_{1}=\frac{1}{2}+\frac{h_{1}^{2}}{4}+n \tag{19}
\end{gather*}
$$

Step 2. Using (10), (14), and (18), we can get

$$
\begin{align*}
d w_{2}= & \left(\widehat{x}_{3}+K^{2} h_{2} \widetilde{x}_{1}+K v_{1} p w_{1}^{p-1}\left(\widehat{x}_{2}+K h_{1} \widetilde{x}_{1}\right)\right) d t \\
= & \left(\widehat{x}_{3}+K^{2} h_{2} \widetilde{x}_{1}+K^{2} h_{1} v_{1} p w_{1}^{p-1} \widetilde{x}_{1}\right.  \tag{20}\\
& \left.+K v_{1} p w_{1}^{p-1}\left(w_{2}+\beta_{1}\right)\right) d t
\end{align*}
$$

Constructing the 2nd Lyapunov function

$$
\begin{equation*}
V_{2}\left(\widetilde{x}, w_{[2]}\right)=V_{1}\left(\widetilde{x}, w_{1}\right)+\frac{1}{K^{2}} \cdot \frac{1}{p+1} w_{2}^{p+1} \tag{21}
\end{equation*}
$$

applying (14), (18)-(21), $K \geq 2^{2 p} n d^{*}$, Lemma 1, and Young's inequality [20], we obtain

$$
\begin{aligned}
L V_{2} \leq & -\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\tilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{4}^{2 p}}{K^{6}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\left(n K-2^{4 p-1} p n d^{*} v_{1}^{2 p}\right) w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+w_{1}^{p} w_{2}+2^{2 p-1} n d^{*} \frac{\widehat{x}_{3}^{2 p}}{K^{4}} \\
& +\frac{1}{K^{2}} w_{2}^{p}\left(\widehat{x}_{3}+K^{2} h_{2} \tilde{x}_{1}+K^{2} h_{1} v_{1} p w_{1}^{p-1} \widetilde{x}_{1}\right. \\
\leq & -\left(\left(\left(n v_{1} p w_{1}^{p-1} w_{2}-K^{2} v_{1}^{2} p w_{1}^{2 p-1}\right)\right.\right. \\
\leq & \times\|\widetilde{x}\|_{\infty}^{2 p}
\end{aligned}
$$

$$
\begin{align*}
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{4}^{2 p}}{K^{6}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\left(n K-2^{4 p-1} p n d^{*} v_{1}^{2 p}\right) w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+w_{1}^{p} w_{2}+2^{2 p-1} n d^{*} \frac{\hat{x}_{3}^{2 p}}{K^{4}}+\frac{1}{K^{2}} w_{2}^{p} \widehat{x}_{3} \\
& +w_{2}^{p} h_{2} \widetilde{x}_{1}+w_{2}^{p} h_{1} v_{1} p w_{1}^{p-1} \widetilde{x}_{1}+\frac{1}{K} w_{2}^{p+1} v_{1} p w_{1}^{p-1} \\
& -v_{1}^{2} p w_{1}^{2 p-1} w_{2}^{p} \\
& \leq-\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{4}^{2 p}}{K^{6}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\left(n K-2^{4 p-1} p n d^{*} v_{1}^{2 p}\right) w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{4}} w_{3}^{2 p}+4^{2 p-1} n d^{*} \frac{1}{K^{4}} \beta_{2}^{2 p}+\frac{1}{K^{2}} w_{2}^{p} w_{3} \\
& +\frac{1}{K^{2}} w_{2}^{p} \beta_{2}+\left(w_{2}^{p}\left(h_{2}+h_{1} v_{1} p w_{1}^{p-1}\right)\right) \tilde{x}_{1} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2}} w_{2}^{2 p}+w_{1}^{p} w_{2}+\frac{1}{K} w_{2}^{p+1} v_{1} p w_{1}^{p-1} \\
& -v_{1}^{2} p w_{1}^{2 p-1} w_{2}^{p} \\
& \leq-\left(((n+1) n-1) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{4}^{2 p}}{K^{6}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\left(n K-2^{4 p-1} p n d^{*} v_{1}^{2 p}\right) w_{1}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{4}} w_{3}^{2 p}+4^{2 p-1} n d^{*} \frac{1}{K^{4}} \beta_{2}^{2 p}+\frac{1}{K^{2}} w_{2}^{p} w_{3} \\
& +\frac{1}{K^{2}} w_{2}^{p} \beta_{2}+\left(w_{2}^{p}\left(h_{2}+h_{1} v_{1} p w_{1}^{p-1}\right)\right) \widetilde{x}_{1} \\
& +\left(\frac{1}{4 n d^{*}}-1\right) w_{2}^{2 p}+w_{1}^{2 p}+\frac{1}{4} w_{2}^{2}+\frac{1}{K} w_{2}^{2 p+2} ; \tag{22}
\end{align*}
$$

then we can get the 2 nd virtual control law

$$
\begin{gather*}
\beta_{2}\left(\widehat{x}_{[2]}\right)=-K^{2} v_{2} w_{2}^{p}, \\
v_{2}=\frac{h_{2}^{2}}{4}+\frac{v_{1}^{2} h_{1}^{2}}{4}+\frac{v_{1}^{4}}{4}+v_{1}+1+n-1, \tag{23}
\end{gather*}
$$

which satisfies

$$
\begin{align*}
L V_{2} \leq & -\left(((n+1) n-2) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\tilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{4}^{2 p}}{K^{6}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\left((n-1) K-2^{4 p-1} p n d^{*} v_{1}^{2 p}\right) w_{1}^{2 p} \\
& -\frac{1}{K^{2}}\left(\left(n-4^{p-1}\right) K-2^{4 p-1} p n d^{*} v_{2}^{2 p}\right) w_{2}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{4}} w_{3}^{2 p}+\frac{1}{K^{2}} w_{2}^{p} w_{3} . \tag{24}
\end{align*}
$$

Step $i(i=3, \ldots, n-1)$. Suppose at $(i-1)$ th, that there are a set of virtual control laws $\beta_{1}\left(\widehat{x}_{1}\right), \ldots, \beta_{i-1}\left(\widehat{x}_{[i-1]}\right)$ : as follows

$$
\begin{align*}
& \beta_{1}\left(\widehat{x}_{1}\right)=-K v_{1} w_{1}^{p}, \\
& \beta_{2}\left(\widehat{x}_{[2]}\right)=-K^{2} v_{2} w_{2}^{p}, \\
& \vdots \\
& \beta_{i-1}\left(\widehat{x}_{[i-1]}\right)=-K^{i-1} v_{i-1} w_{i-1}^{p}, \\
& v_{1}= \frac{1}{2}+\frac{h_{1}^{2}}{4}+n, \\
& v_{2}= \frac{h_{2}^{2}}{4}+\frac{v_{1}^{2} h_{1}^{2}}{4}+\frac{v_{1}^{4}}{4}+v_{1}+1+n-1,  \tag{25}\\
& \vdots \\
& v_{i-1}= \frac{1}{4}\left(h_{i-1}+v_{i-2} h_{i-2}+v_{i-2} v_{i-3} h_{i-3}\right. \\
&\left.+\cdots+v_{i-2} v_{i-3} \cdots v_{1} h_{1}\right)^{2} \\
&+\cdots+\frac{1}{4}\left(v_{i-3} v_{i-2}-v_{i-2}^{2}\right)^{2}+v_{i-2} \\
&+1+n-(i-2),
\end{align*}
$$

with $v_{j}>0(j=1, \ldots, i-1)$ being independent of $K$ such that the $i$ th Lyapunov function

$$
\begin{equation*}
V_{i-1}\left(\widetilde{x}, w_{[i-1]}\right)=V_{0}(\widetilde{x})+\frac{1}{p+1} \sum_{j=1}^{i-1} \frac{1}{K^{j}} w_{j}^{p+1} \tag{26}
\end{equation*}
$$

satisfies

$$
\begin{aligned}
L V_{i-1} \leq & -\left(((n+1) n-(i-1)) K-2^{2 p-1} n d^{*}\right. \\
& \left.\times\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right)\|\widetilde{x}\|_{\infty}^{2 p} \\
+ & 2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{i+1}^{2 p}}{K^{2 i}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)
\end{aligned}
$$

$$
\begin{align*}
& -\sum_{j=1}^{i-1} \frac{1}{K^{2 j-2}}\left(\left(n-4^{(p-1)(j-1)}\right) K-2^{4 p-1} p n d^{*} v_{j}^{2 p}\right) w_{j}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2(i-1)}} w_{i}^{2 p}+\frac{1}{K^{2(i-2)}} w_{i-1}^{p} w_{i} \tag{27}
\end{align*}
$$

In the sequel, we will prove that (27) still holds for

$$
\begin{equation*}
V_{i}\left(\widetilde{x}, w_{[i]}\right)=V_{i-1}\left(\widetilde{x}, w_{[i-1]}\right)+\frac{1}{p+1} \cdot \frac{1}{K^{i}} w_{i}^{p+1} \tag{28}
\end{equation*}
$$

Using (14) and (25), a direct calculation leads to

$$
\begin{align*}
d z_{i}=( & \widehat{x}_{i+1}+K^{i} h_{i} \widetilde{x}_{i}+K h_{i-1} p w_{i-1}^{p-1}\left(\widehat{x}_{i}+K^{i-1} h_{i-1} \widetilde{x}_{1}\right) \\
& +K^{2} v_{i-1} v_{i-2} p w_{i-2}^{p-1}\left(\widehat{x}_{i-1}+K^{i-2} h_{i-2} \widetilde{x}_{1}\right) \\
& \left.+\cdots+K^{i-1} v_{i-1} v_{i-2} \cdots v_{1} p w_{1}^{p-1}\left(\widehat{x}_{2}+K h_{1} \widetilde{x}_{1}\right)\right) d t \tag{29}
\end{align*}
$$

Using $K \geq 2^{2 p} n d^{*}$, Lemma 1, Young's inequality [22], (14), (27), (28), and (29), we obtain
$L V_{i} \leq-\left(((n+1) n-(i-1)) K-2^{2 p-1} n d^{*}\right.$ $\left.\times\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right)\|\widetilde{x}\|_{\infty}^{2 p}$ $+2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{i+2}^{2 p}}{K^{2(i+1)}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)$ $-\sum_{j=1}^{i-1} \frac{1}{K^{2 j-2}}\left(\left(n-4^{(p-1)(j-1)}\right) K-2^{4 p-1} p n d^{*} v_{j}^{2 p}\right) w_{j}^{2 p}$ $+2^{2 p-1} n d^{*} \frac{1}{K^{2 i}} \widehat{x}_{i+1}^{2 p}+\frac{1}{K^{2(i-1)}} w_{i}^{2 p} \widehat{x}_{i+1}$ $+\frac{1}{K^{i}} w_{i}^{p}$

$$
\times\left(\widehat{x}_{i+1}+K^{i} h_{i} \widetilde{x}_{i}+K v_{i-1} p w_{i-1}^{p-1}\left(\widehat{x}_{i}+K^{i-1} h_{i-1} \widetilde{x}_{1}\right)\right.
$$

$$
\left.+\cdots+K^{i-1} v_{i-1} v_{i-2} \cdots v_{1} p w_{1}^{p-1}\left(\widehat{x}_{2}+K h_{1} \widetilde{x}_{1}\right)\right)
$$

$$
\leq-\left(((n+1) n-i) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right)
$$

$$
\times\|\tilde{x}\|_{\infty}^{2 p}
$$

$$
+2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{i+2}^{2 p}}{K^{2(i+1)}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right)
$$

$$
-\sum_{j=1}^{i} \frac{1}{K^{2 j-2}}\left(\left(n-4^{(p-1) j}\right) K-2^{4 p-1} p n d^{*} v_{j}^{2 p}\right) w_{j}^{2 p}
$$

$$
+4^{2 p-1} n d^{*} \frac{1}{K^{2 i}} w_{i+1}^{2 p}+4^{2 p-1} n d^{*} \frac{1}{K^{2 i}} \beta_{i}^{2 p}
$$

$$
+\frac{1}{K^{2(i-1)}} w_{i}^{p} w_{i+1}+\frac{1}{K^{2(i-1)}} w_{i}^{p} \beta_{i}
$$

$$
\begin{align*}
+\frac{1}{K^{2 i-3}} w_{i}^{2 p}\left(\frac { 1 } { 4 } \left(h_{i}\right.\right. & +v_{i-1} h_{i-1}+v_{i-1} v_{i-2} h_{i-2} \\
& \left.+\cdots+v_{i-1} v_{i-2} \cdots v_{1} h_{1}\right)^{2} \\
+ & \left.\cdots+\frac{1}{4}\left(v_{i-3} v_{i-2}-v_{i-2}^{2}\right)^{2}+v_{i-1}+1\right) \tag{30}
\end{align*}
$$

then we can choose the $i$ th smooth virtual control law

$$
\begin{gather*}
\beta_{i}\left(\widehat{x}_{[i]}\right)=-K^{i} v_{i} w_{i}^{p}, \\
v_{i}=\frac{1}{4}\left(h_{i}+v_{i-1} h_{i-1}+v_{i-1} v_{i-2} h_{i-2}+\cdots+v_{i-1} v_{i-2} \cdots v_{1} h_{1}\right)^{2} \\
+\cdots+\frac{1}{4}\left(v_{i-3} v_{i-2}-v_{i-2}^{2}\right)^{2}+v_{i-1}+1+n-(i-1) \tag{31}
\end{gather*}
$$

and get

$$
\begin{align*}
L V_{i} \leq & -\left(((n+1) n-i) K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right) \\
& \times\|\widetilde{x}\|_{\infty}^{2 p} \\
& +2^{2 p-1} n d^{*}\left(\frac{\widehat{x}_{i+2}^{2 p}}{K^{2(i+1)}}+\cdots+\frac{\widehat{x}_{n}^{2 p}}{K^{2 n-2}}\right) \\
& -\sum_{j=1}^{i} \frac{1}{K^{2 j-2}}\left(\left(n-4^{(p-1) j}\right) K-2^{4 p-1} p n d^{*} v_{j}^{2 p}\right) w_{j}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2 i}} w_{i+1}^{2 p}+\frac{1}{K^{2(i-1)}} w_{i}^{p} w_{i+1} . \tag{32}
\end{align*}
$$

Step $n$. Using repeatedly the previous arguments, at the $n-1$ th step, we can get

$$
\begin{align*}
L V_{n-1} \leq & -\left(((n+1) n-(n-1)) K-2^{2 p-1} n d^{*}\right. \\
& \left.\times\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right)\|\tilde{x}\|_{\infty}^{2 p} \\
- & \sum_{j=1}^{n-1} \frac{1}{K^{2 j-2}}  \tag{33}\\
& \times\left(\left(n-4^{(p-1)(j-1)}\right) K-2^{4 p-1} p n d^{*} v_{j}^{2 p}\right) w_{j}^{2 p} \\
& +4^{2 p-1} n d^{*} \frac{1}{K^{2(i-1)}} w_{n}^{2 p}+\frac{1}{K^{2(i-2)}} w_{n-1}^{p} w_{n}
\end{align*}
$$

where

$$
\begin{equation*}
V_{n-1}\left(\tilde{x}, w_{[n-1]}\right)=V_{n-2}\left(\tilde{x}, w_{[n-2]}\right)+\frac{1}{K^{n-1}} \cdot \frac{1}{p+1} w_{n}^{2 p} . \tag{34}
\end{equation*}
$$

At the end of the recursive procedure, choosing the controller

$$
\begin{equation*}
u\left(\widehat{x}_{[n]}\right)=-K^{n} v_{n} w_{n}^{p}, \tag{35}
\end{equation*}
$$

where $v_{n}>0$ satisfies (25) and is independent of $K$, we can get

$$
\begin{align*}
L V_{n} \leq & -\left(n^{2} K-2^{2 p-1} n d^{*}\left(\sum_{i=1}^{n} K^{(i-1) p-(i-1)}\right)^{2}\right)\|\widetilde{x}\|_{\infty}^{2 p} \\
& -\sum_{j=1}^{n-1} \frac{1}{K^{2 j-2}} \\
& \times\left(\left(n-4^{(p-1)(j-1)}\right) K-2^{4 p-1} p n d^{*} v_{j}^{2 p}\right) w_{j}^{2 p} \\
& -\frac{1}{K^{2 n-5}} w_{n}^{2 p} \tag{36}
\end{align*}
$$

where

$$
\begin{equation*}
V_{n}\left(\widetilde{x}, w_{[n]}\right)=(n+1) \tilde{x}^{T} P \tilde{x}+\frac{1}{p+1} \sum_{j=1}^{n} \frac{1}{K^{j}} w_{j}^{p+1} . \tag{37}
\end{equation*}
$$

Remark 7. The item $\left(\widehat{x}_{1}^{2 p}+\widehat{x}_{2}^{2 p} / K^{2}+\cdots+\widehat{x}_{n}^{2 p} / K^{2 n-2}\right)$ is canceled at step $n-1$. By the following analysis, we obtain the maximum value interval of $K$ to ensure the system to be globally asymptotically stable in probability at the origin.

## 5. Performance Analysis

Next, we give the main result in this paper.
Theorem 8. If Assumption 5 holds for stochastic nonlinear system (9) under the controllers (10) and (35), then there always exists a constant $K^{*} \geq 0$, such that for any $K>K^{*}$,
(1) the closed-loop system has an almost surely unique solution on $[0, \infty)$ for any $x_{0}$;
(2) the equilibrium at the origin of the closed-loop system is globally asymptotically stable in probability.

Proof. Using $K \geq 0$, (18), (23), and (31), obviously, if

$$
\left.\begin{array}{rl}
K & >\max \left\{\left|\frac{2^{4 p-1} p n d^{*} v_{1}^{2 p}}{n}\right|,\left|\frac{2^{4 p-1} p n d^{*} v_{2}^{2 p}}{n-4^{p-1}}\right|, \ldots,\right. \\
& \left.=\left|\frac{2^{4 p-1} p n d^{*} v_{n}^{2 p}}{n-4^{(p-1)(n-1)}}\right|\right\}  \tag{38}\\
n-4^{(p-1)(n-1)}
\end{array}\right)
$$

holds, then conclusions (1) and (2) follow from (36), (37), and Lemma 3. In the following, we analyze (38). From (13), (25), and (31), it is easy to find that $d^{*}$ depends on $K$. Substituting (13) into (38) leads to

$$
\begin{align*}
K> & \frac{2^{4 p-1} p n v_{n}^{2 p}}{n-4^{(p-1)(n-1)}} \cdot \frac{(n+1)\|P\| d^{2}}{K^{2 n-2}} \\
& \cdot\left(\sum_{i=0}^{n-1}(i+1) K^{i}+\sum_{i=n}^{2 n-2}(2 n-i-1) K^{i}\right) \tag{39}
\end{align*}
$$

equivalently,

$$
\begin{align*}
& K^{2 n-1} \\
& \quad>\frac{n}{n-4^{(p-1)(n-1)}} \cdot 2^{4 p-1} \cdot p \cdot v_{n}^{2 p} \\
& \quad \cdot(n+1)\|P\| d^{2}\left(\sum_{i=0}^{n-1}(i+1) K^{i}+\sum_{i=n}^{2 n-2}(2 n-i-1) K^{i}\right) \tag{40}
\end{align*}
$$

which is equivalent to

$$
\begin{equation*}
K^{2 n-1}+\sum_{i=0}^{2 n-2} a_{i} K^{i}>0 \tag{41}
\end{equation*}
$$

with the real numbers

$$
\begin{align*}
a_{0}= & -\Delta, \\
a_{1}= & -2 \Delta, \\
& \vdots \\
a_{n-1}= & -n \Delta, \\
a_{n}= & -(n-1) \Delta, \\
a_{n+1}= & -(n-2) \Delta, \\
& \vdots \\
a_{2 n-2}= & -\Delta-(n+1)\|P\|^{2}, \quad \Delta=2^{4 p-1} p n(n+1) d^{2} v_{n-1}^{2 p} . \tag{42}
\end{align*}
$$

According to the factorization theorem of real coefficient polynomial, (41) can be further expressed as

$$
\begin{gather*}
\left(K-K_{1}\right)^{m_{1}} \cdots\left(K-K_{r}\right)^{m_{r}}\left(K^{2}+p_{1} K+q_{1}\right)^{n_{1}}  \tag{43}\\
\cdots\left(K^{2}+p_{s} K+q_{s}\right)^{n_{s}}>0,
\end{gather*}
$$

where $m_{i}, n_{j}$ are positive integers with $\sum_{i=1}^{r} m_{i}+2 \sum_{j=1}^{s} n_{j}=$ $2 n-1, K_{i}, i \leq r$, are different real numbers, and $\left(p_{j}, q_{j}\right)$, $j \geq s$, satisfy $p_{j}^{2}-4 q_{j}<0$ for all $j=1, \ldots, s$. Obviously, $K^{2}+p_{j} K+K_{j}>0$ for all $j=1, \ldots, s$. Now, we divide into two cases to discuss the choice of $K_{i}$. (1) If there is at least one positive number for $K_{1}, \ldots, K_{r}$ under the condition of appropriate value of $p$, one chooses $K^{*}=\max _{1 \leq i \leq r}\left\{K_{i}\right\}$. (2) Otherwise, $K^{*}=0$. Thus there always exists $K^{*} \geq 0$, such that for any $K>K^{*}$, (38) holds.

## 6. Inverse Optimal Controller Design

In this section we will design the inverse optimal controller on the basis of Theorem 8 to meet specific performance indicators besides achieving control objectives.

Theorem 9. The control law

$$
\begin{equation*}
u^{*}=-\theta K^{n} v_{n} w_{n}^{p}, \quad \theta \geq 2 \tag{44}
\end{equation*}
$$

solves the problem of inverse optimal stabilization in probability for (9) by minimizing the cost function

$$
\begin{equation*}
J(u)=E\left\{\int_{0}^{\infty}\left[l(\widetilde{x}, \widehat{x})+\frac{1}{K^{2(n-1)}} \frac{v_{n}^{-1}(\widehat{x})}{K^{n}} u^{2}\right] d r\right\} \tag{45}
\end{equation*}
$$

where $\bar{\varphi}_{1}(\widetilde{x}, \widehat{x})=\left(\Phi^{T}(x), 0, \ldots, 0\right)^{T}, \bar{\varphi}_{2}(\widetilde{x}, \widehat{x})=(0, \ldots, 0,1)^{T}$, $V=V_{n}$.

Proof. Equations (10) and (11) can be represented as

$$
\begin{equation*}
\binom{d \tilde{x}}{d \widehat{x}}=\bar{\varphi}_{1}(\widetilde{x}, \widehat{x}) d \omega+\bar{\varphi}_{2}(\widetilde{x}, \widehat{x}) u d t \tag{46}
\end{equation*}
$$

where $\bar{\varphi}_{1}, \bar{\varphi}_{2}$ are identified in Theorem 9. Choosing $\gamma(r)=$ $\left(1 / 2 K^{2(n-1)}\right) r^{2}$, we can get $\left(\gamma^{\prime}\right)^{-1}(r)=K^{2(n-1)} r$ and $\ell \gamma(r)=$ $(1 / 2) K^{2(n-1)} r^{2}$. Applying Lemma 4, we get

$$
\begin{equation*}
u=\beta(\widehat{x})=-R_{2}^{-1}(\widehat{x}) \frac{1}{K^{2(n-1)}} w_{n}^{p} \frac{1}{2} K^{2(n-1)}=-\frac{1}{2} R_{2}^{-1}(\widehat{x}) w_{n}^{p} \tag{47}
\end{equation*}
$$

According to Theorem 8 and Lemma 4, the inverse optimal controller can be designed as follows:

$$
\begin{align*}
u^{*} & =\beta^{*}(\widehat{x})=-\frac{\theta}{2} R_{2}^{-1}(\widehat{x}) \frac{1}{K^{2(n-1)}} w_{n}^{p} K^{2(n-1)} \\
& =-\frac{\theta}{2} R_{2}^{-1}(\widehat{x}) w_{n}^{p}=\theta \beta(\widehat{x})=\theta u, \quad \theta \geq 2 \tag{48}
\end{align*}
$$

where $R_{2}(\widehat{x})=1 / 2 K^{n} v_{n}$.

## 7. Simulation Examples

In this section, for a numerical example, we design the output-feedback controller by using two methods, where one method is introduced in this paper and the other is introducted in [19, 20].

Consider the following stochastic system:

$$
\begin{gather*}
d x_{1}=x_{2} d t+\frac{1}{10} x_{1}^{3} \sin x_{2} d \omega \\
d x_{2}=u d t+\frac{1}{10}\left(x_{1}^{3}+x_{2}^{3}\right) d \omega  \tag{49}\\
y=x_{1}
\end{gather*}
$$

where

$$
\begin{gather*}
\left|\varphi_{1}(\mathbf{x})\right|=\left|\frac{1}{10} x_{1}^{3} \sin x_{2}\right| \leq \frac{1}{10}\left|x_{1}\right|^{3}  \tag{50}\\
\left|\varphi_{2}(\mathbf{x})\right|=\left|\frac{1}{10}\left(x_{1}^{3}+x_{2}^{3}\right)\right| \leq \frac{1}{10}\left(\left|x_{1}\right|^{3}+\left|x_{2}\right|^{3}\right) .
\end{gather*}
$$


(a)


$---x_{2}$
(b)


(c)

Figure 1: The responses of the closed-loop systems (49)~(53).

(a)

$\begin{array}{ll}- & x_{1} \\ --x_{2} & x_{2}\end{array}$
(b)

-_ $x_{1}$ observer
--- $x_{2}$ observer
(c)

Figure 2: The responses of the closed-loop systems (49)~(53) when adopting the method in $[19,20]$.

With the notation of Assumption 5, one can take $p=3, c=$ 1/10.

In line with design method discussed in Section 4, we can design the observer states as follows:

$$
\begin{align*}
& \dot{\hat{x}}_{1}=\widehat{x}_{2}+K\left(y-\widehat{x}_{1}\right), \\
& \dot{\hat{x}}_{2}=u+K^{2}\left(y-\widehat{x}_{1}\right) . \tag{51}
\end{align*}
$$

According to the design procedure in Section 4, we construct the controller as follows:

$$
\begin{gather*}
\beta_{1}\left(\widehat{x}_{1}\right)=-K v_{1} w_{1}^{3}, \quad v_{1}=2.75 \\
u=-K^{2} v_{2} w_{2}^{3}, \quad v_{2}=8.27 \tag{52}
\end{gather*}
$$

where $K$ will be chosen later, $h_{1}=h_{2}=1$, and $d^{*}=$ $(1 / 12) \times 10^{-4}\|P\|(1+1 / K)^{2},\|P\|=((15+5 \sqrt{5}) / 8)^{1 / 2}$. With Theorem 8, one gets $K>42.35$. According to the design procedure in Section 6, we choose $\theta=2$ and construct the inverse optimal controller as follows:

$$
\begin{equation*}
u^{*}=-2 K^{2} v_{2} w_{2}^{3} . \tag{53}
\end{equation*}
$$

In simulation, we choose the initial values $x_{1}(0)=0.02$, $x_{2}(0)=0.01, \widehat{x}_{1}(0)=0.01, \widehat{x}_{2}(0)=-0.01$, and $K=$ 50. Figure 1 shows the responses of the closed-loop system (49)~(53), which demonstrate the effectiveness of the control scheme.

If the method in $[19,20]$ is adopted for the same systems, Figure 2 gives the corresponding responses of the systems (here the controller design theory of $[19,20]$ is not tackled details, and the interested readers can consult the relevant literature).

Remark 10. By comparing the two figures, we can observe that the value of the control of Figure 1 is far less than Figure 2. In other words, our method requires less control effort to ensure the closed-loop system to be globally asymptotically stable in probability, and it demonstrates the advantage of this method clearly.

## 8. Concluding and Outlook

In this paper, we have studied the output-feedback stabilization for a class of stochastic nonlinear systems. We have given a design of the output-feedback controller so as to make the equilibrium at the origin of the closedloop system globally asymptotically stable in probability by using the backstepping design technique and choosing a high-gain parameter, and the inverse optimal stabilization in probability is achieved. Our main contribution is extending the linear growth conditions to the more general power growth conditions so as to enable the result to be more general and to have a broader field of use.

There are two problems to be investigated.
(1) By extending the value of $p$ in Assumption 5 from positive integers to the rationales, it can further weaken the conditions of the system (5). For this system, output-feedback problem deserves further research.
(2) Another is to extend stochastic nonlinear systems in this paper to delay systems and study the design of controller.

## References

[1] H. Deng and M. Krstić, "Output-feedback stochastic nonlinear stabilization," IEEE Transactions on Automatic Control, vol. 44, no. 2, pp. 328-333, 1999.
[2] H. Deng and M. Krstić, "Output-feedback stabilization of stochastic nonlinear systems driven by noise of unknown covariance," Systems \& Control Letters, vol. 39, no. 3, pp. 173182, 2000.
[3] M. Krstić and H. Deng, Stabilization of Nonlinear Uncertain Systems, Communications and Control Engineering Series, Springer, New York, NY, USA, 1998.
[4] Y.-G. Liu and J.-F. Zhang, "Practical output-feedback risksensitive control for stochastic nonlinear systems with stable zero-dynamics," SIAM Journal on Control and Optimization, vol. 45, no. 3, pp. 885-926, 2006.
[5] Z.-J. Wu, X.-J. Xie, and S.-Y. Zhang, "Stochastic adaptive backstepping controller design by introducing dynamic signal and changing supply function," International Journal of Control, vol. 79, no. 12, pp. 1635-1646, 2006.
[6] S.-J. Liu, J.-F. Zhang, and Z.-P. Jiang, "Decentralized adaptive output-feedback stabilization for large-scale stochastic nonlinear systems," Automatica, vol. 43, no. 2, pp. 238-251, 2007.
[7] Z. Pan, Y. Liu, and S. Shi, "Output feedback stabilization for stochastic nonlinear systems in observer canonical form with stable zero-dynamics," Science in China. Series F, vol. 44, no. 4, pp. 292-308, 2001.
[8] Y. Liu, J. Zhang, and Z. Pan, "Design of satisfaction output feedback controls for stochastic nonlinear systems under quadratic tracking risk-sensitive index," Science in China. Series F, vol. 46, no. 2, pp. 126-144, 2003.
[9] Z.-J. Wu, X.-J. Xie, and S.-Y. Zhang, "Adaptive backstepping controller design using stochastic small-gain theorem," Automatica, vol. 43, no. 4, pp. 608-620, 2007.
[10] X. Yu and X.-J. Xie, "Output feedback regulation of stochastic nonlinear systems with stochastic iISS inverse dynamics," IEEE Transactions on Automatic Control, vol. 55, no. 2, pp. 304-320, 2010.
[11] W.-Q. Li and X.-J. Xie, "Inverse optimal stabilization for stochastic nonlinear systems whose linearizations are not stabilizable," Automatica, vol. 45, no. 2, pp. 498-503, 2009.
[12] W. Li, X. Liu, and S. Zhang, "Further results on adaptive statefeedback stabilization for stochastic high-order nonlinear systems," Automatica, vol. 48, no. 8, pp. 1667-1675, 2012.
[13] S.-J. Liu and J.-F. Zhang, "Output-feedback control of a class of stochastic nonlinear systems with linearly bounded unmeasurable states," International Journal of Robust and Nonlinear Control, vol. 18, no. 6, pp. 665-687, 2008.
[14] N. Duan and X.-J. Xie, "Further results on output-feedback stabilization for a class of stochastic nonlinear systems," IEEE Transactions on Automatic Control, vol. 56, no. 5, pp. 1208-1213, 2011.
[15] C. Qian, "A homogeneous domination approach for global output feedback stabilization of a class of nonlinear systems," in Proceedings of the American Control Conference (ACC '05), pp. 4708-4715, June 2005.
[16] C. Qian and W. Lin, "Non-Lipschitz continuous stabilizers for nonlinear systems with uncontrollable unstable linearization," Systems \& Control Letters, vol. 42, no. 3, pp. 185-200, 2001.
[17] J. Polendo and C. Qian, "A generalized framework for global output feedback stabilization of genuinely nonlinear systems," in Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference (CDC-ECC '05), pp. 2646-2651, December 2005.
[18] X.-J. Xie and W.-Q. Li, "Output-feedback control of a class of high-order stochastic nonlinear systems," International Journal of Control, vol. 82, no. 9, pp. 1692-1705, 2009.
[19] W. Li, Y. Jing, and S. Zhang, "Output-feedback stabilization for stochastic nonlinear systems whose linearizations are not stabilizable," Automatica, vol. 46, no. 4, pp. 752-760, 2010.
[20] W. Li, X. Liu, and S. Zhang, "Output-feedback stabilization of high-order stochastic nonlinear systems with more general growth conditions," in Proceedings of the 30th Chinese Control Conference (CCC '11), pp. 5953-5957, July 2011.
[21] V. Kolmanovskii and A. Myshkis, Introduction to the Theory and Applications of Functional Differential Equations, vol. 463 of Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
[22] M. Krstić and H. Deng, Stabilization of Uncertain Nonlinear Systems, Springer, New York, NY, USA, 1980.
[23] R. Z. Khas'minskii, Stochastic Stability of Differential Equations, vol. 7 of Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis, Sijthoff \& Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.
[24] H. Deng and M. Krstić, "Stochastic nonlinear stabilization. II. Inverse optimality", Systems \& Control Letters, vol. 32, no. 3, pp. 151-159, 1997.

# A Multilayer Feed Forward Small-World Neural Network Controller and Its Application on Electrohydraulic Actuation System 

Xiaohu Li, Feng Xu, Jinhua Zhang, and Sunan Wang<br>School of Mechanical Engineering, Xian Jiaotong University, Xian 710049, China<br>Correspondence should be addressed to Xiaohu Li; li.xiaohu@mail.xjtu.edu.cn

Received 26 March 2013; Accepted 21 May 2013
Academic Editor: Tao Zou
Copyright © 2013 Xiaohu Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

Being difficult to attain the precise mathematical models, traditional control methods such as proportional integral (PI) and proportional integral differentiation (PID) cannot meet the demands for real time and robustness when applied in some nonlinear systems. The neural network controller is a good replacement to overcome these shortcomings. However, the performance of neural network controller is directly determined by neural network model. In this paper, a new neural network model is constructed with a structure topology between the regular and random connection modes based on complex network, which simulates the brain neural network as far as possible, to design a better neural network controller. Then, a new controller is designed under smallworld neural network model and is investigated in both linear and nonlinear systems control. The simulation results show that the new controller basing on small-world network model can improve the control precision by $30 \%$ in the case of system with random disturbance. Besides the good performance of the new controller in tracking square wave signals, which is demonstrated by the experiment results of direct drive electro-hydraulic actuation position control system, it works well on anti-interference performance.


## 1. Introduction

As we know, neural network is an adaptive function estimator needless to know the determined math relationship between input and output, and it also has good adaptability and learning ability. These features make it very suitable to be used as intelligent controller for complex systems. Many researches show that the neural network controller is especially suitable for those uncertain or nonlinear control objects, which make it have a wide application prospect in the field of intelligent control. Obviously, the constructed neural network model directly determines the quality of the neural network controller; in short, the architecture design is critical for neural network model. At present, the structures of artificial neural network model are often designed as feedback, feed forward, single neuron, multilayer, and so on. As such, it is an important problem worthy of discussion whether these structures of neural network models are optimal and if they
are able to reflect the real human brain neural network structure or not.

Recent researches have shown that the structure topology and function of human brain neural network are closely related to each other; the connection mode of brain neural network structure topology offers the possibility for different brain areas to mutual collaboration [1-6]. This collaboration is mainly due to the large number of neurons in human brain, which means that the simple structure topology and function of a single neuron are multiplied by the large number of neurons. All those neurons connecting with each other by nerve fibers according to a certain kind of connection mode can make up highly complex human brain neural network. Unfortunately, the existing artificial neural network models are simple simulation of the biological neural network in structure topology and function. The common connection modes of artificial neural network can be divided into feed forward, feedback, single layer, multilayer, and so forth, all
of which can be regarded as regular structure topology. In the past few decades, lots of researches have been carried out regarding artificial neural network with regular structure topology [7-10]. Although the bioneurological studies have shown that neural network inherently has random features on structure topology, the research of artificial neural network with random structure topology is relatively rarely reported. Consequently, it is essential to design a neural network model with stochastic characteristic structure, which will not only help in obtaining an optimal network performance, but also be a more realistic reflection of the structural features of the brain neural network.

At present, there are two typical methods to reflect the random features of neural network: one is to adopt random connection weight between the neurons of neural network; the other is to use random neuron activation function [11]. But neither of the methods mentioned above can really reveal the random features on neural network structure topology. New achievements of complex network have provided a new method for constructing neural network model with random feature structure. Watts and Strogatz researched on the structure topology of many realistic complex networks and defined the intermediate network between completely regular and random networks as small-world network (referred to as W-S model) [12]. Many studies on complex network have shown that the realistic networks, such as disease transmission network, social network, food chain network, metabolic network, and so forth, are all small-world network on structure topology [13-18]. Meanwhile, scholars discovered that biological neural networks are also smallworld network [19-23]. All of these small-world networks have random characteristics in structure. Obviously, if the structure topology of artificial neural network is built as a small-world one, it will really reflect the structure topology of the biological neural network. Erkaymaz et al. and Simard et al. rewired the links of multilayer feed forward neural network to build a small-world neural network model [24, 25]. However, in the construction process, it is apparent that the unchanging number of rewiring means a lack of randomness on structure topology, which is not consistent with W-S model. Therefore, it is necessary to rebuild a new neural network model, which relies on the rewiring probability. As self-learning is an important characteristic of artificial neural network, control application can be used to assess the performance of multilayer feed forward smallworld neural network model. Since the new controller does not require a precise mathematical model, it can be applied to the electrohydraulic actuation system, with the help of which we may probe the features of small-world neural network model as well as its controller.

In this paper, firstly, the multilayer feed forward smallworld neural network model is built up according to the WS model. Secondly, the mathematical model of small-world neural network is briefly described. Finally, a new controller with a small-world neural network model is designed for system control, which will not only explore the control performance of small-world neural network, but also compare control precision and anti-interference of small-world neural network with those of regular neural network. The
performance of the controller is also verified by experiment on electrohydraulic actuation system.

## 2. Model Construction

In the multilayer feed forward neural network model, the $i$ th neuron of the $l$ th layer $v_{i}^{l}$ only connects with its neighbor neurons, which belong to neuron sets $V^{l-1}$ and $V^{l+1}$. All of these links are feed forward, and there are no links between neurons of the same layer. As the links of each neuron are similar, this network structure topology can be regarded as regularity $[26,27]$. Assuming that the number of neurons in each layer is $n_{l}$, and the number of layers of neural network is $L$ (including the input layer and output layer), the regular network structure topology is shown in Figure 1(a) when rewiring probability $p=0$. The operating mode of regular network is that input signals pass through the hidden layers, and transfer forwardly layer by layer, until they reach the output layer.

Referring to the construction process of W-S model, a multilayer feed forward small-world neural network is built up in the literature [24]. In the construction process, the regular links of multilayer feed forward neural network are reconnected, but as a result of the determinate number of links and the network construction different from W-S model, the rewiring probability $p$ is not used to reconstruct the network. Thus it cannot fully reflect the construction ideology of W-S model. Taking into account that the rewiring probability $p$ has a direct impact on the generated network structure and characteristics, this paper proposes an algorithm to construct the multilayer feed forward small-world neural network model according to the rewiring probability $p$, and the process of algorithm is described as follows.
(1) Generate the same number of neurons in each layer, and connect neighbor neurons with feed forward links. Thus the multilayer feed forward regular neural network model is built, as shown in Figure 1(a), and the connection mode of this model is regular.
(2) With rewiring probability $p$, disconnect the links from neuron $i$ of the $l$ th layer to neuron $j$ of the $(l+1)$ th layer in regular network; then randomly select neuron $j^{\prime}$ ahead the $(l+1)$ th layer to rewire; the new long links are not reconnection or self-loop. Clearly, if the links of the $(L-1)$ th layer disconnect, they could not generate new long links. Therefore, the last two layers cannot be rewired.
(3) Repeat (2), until all the links are rewired besides those of the last two layers.

Figure 1(a) shows that when $p=0$, the connection mode of neural network model is completely regular, in which each neuron maintains the same number of links with adjacent neurons and this neural network model is commonly used. When $p=1$, as shown in Figure 1(c), all original links in the regular neural network model (except for the last two layers) are rewired into links with completely random features under disordered structure topology, forming a completely random


Figure 1: Process of construction multilayer feed forward small-world neural network.
neural network model. If $0<p<1$ (e.g., $p=0.05$ ), long cross-layer links will be generated by rewiring probability $p$, and this structure topology is between completely regular and random, as shown in Figure 1(b). Obviously, this connection structure topology is an intermediate form from regular to random, which is in consistent with the ideology of W-S model. Therefore, this model is defined as multilayer feed forward small-world neural network model, and the smaller $p$ is, the fewer there exist the long links in the new network model.

## 3. Network Model

In order to describe multilayer feed forward small-world neural network model distinctly, graph theory is utilized. Each neuron in multilayer feed forward small-world neural network model can be seen as a node in the graph. The links within all neurons are edges of the graph. The node set of small-world neural network model can be defined as

$$
\begin{equation*}
V=\left\{V^{l} \mid l=1,2, \ldots, L\right\} \tag{1}
\end{equation*}
$$

Where $V^{l}=\left\{v_{i}^{l} \mid i=1,2, \ldots, n_{l}\right\}$ is the node subset of the $l$ th layer, $L$ is the total layer number of the neural network model, $v_{i}^{l}$ is the $i$ th node of the $l$ th layer, and $n_{l}$ is the node number of each layer.

Suppose the connection matrix of the neural network model is

$$
\begin{equation*}
\mathbf{W}=\left\{\mathbf{W}^{1}, \ldots, \mathbf{W}^{1}, \ldots, \mathbf{W}^{L-1}\right\} \tag{2}
\end{equation*}
$$

where $\mathbf{W}^{l}(l=1,2, \ldots, L-1)$ is the connection submatrix of the $l$ th layer, and $\mathbf{w}_{i}^{l}$ is the connection vector of $V^{l} \rightarrow v_{i}^{l+1}$ :

$$
\begin{align*}
& \mathbf{w}_{i}^{l}=\left(w_{1 i}^{l}, w_{2 i}^{l}, \ldots, w_{n_{l} i}^{l}\right)^{T}  \tag{3}\\
& \mathbf{W}^{l}=\left(\mathbf{w}_{1}^{l}, \mathbf{w}_{2}^{l}, \ldots, \mathbf{w}_{n_{l}}^{l}\right)
\end{align*}
$$

where $i=1,2, \ldots, n_{l}$, and $w_{i j}^{l} \in \mathbb{R}$ is the connection weight of $v_{i}^{l} \rightarrow v_{j}^{l+1}$; if neuron $i$ of the $l$ th layer connects with neuron $j$
of the $(l+1)$ th layer, $w_{i j}^{l} \neq 0$, contrariwise, $w_{i j}^{l}=0$. Therefore, the regular neural network model connection matrix can be expressed as

$$
\mathbf{W}=\begin{gather*}
 \tag{4}\\
V^{1} \\
V^{2} \\
\vdots \\
V^{L-1} \\
V^{L}
\end{gather*}\left[\begin{array}{cccccc}
0 & \mathbf{W}^{1} & 0 & \cdots & 0 & 0 \\
0 & 0 & \mathbf{W}^{2} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \mathbf{W}^{L-1} \\
0 & 0 & 0 & \cdots & 0 & 0
\end{array}\right] .
$$

As for the multilayer small-world neural network model, due to the rewiring, the connection matrix even becomes

$$
\left.\mathbf{W}=\begin{array}{c}
V^{1} \\
V^{1}  \tag{5}\\
V^{2} \\
V^{2} \\
\vdots \\
V^{L-1} \\
V^{L}
\end{array}\left[\begin{array}{ccccc}
0 & \mathbf{W}^{1^{\prime}} & B_{1}^{1} & \cdots & B_{1}^{L-3} \\
0 & 0 & \mathbf{W}^{2^{\prime}} & \cdots & B_{1}^{L-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hline 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0
\end{array}\right] . \mathbf{W}_{2}^{L-3}{ }^{L-1^{\prime}}\right]
$$

As the reconnections of the last two layers in the neural network model do not exist $\left(\mathbf{W}^{L-1^{\prime}}=\mathbf{W}^{L-1}, \mathbf{W}^{l^{\prime}}=\left(w_{i j}^{l \prime}\right)_{n_{l} \times n_{l}}\right)$ is the reconnection matrix of $\mathbf{W}^{l}$, the elements of this matrix can be deformed into the following equation:

$$
w_{i j}^{l \prime}= \begin{cases}w_{i j}^{l}, & \operatorname{rand}_{i j} \geq p  \tag{6}\\ 0, & \operatorname{rand}_{i j}<p\end{cases}
$$

where $\operatorname{rand}_{i j}$, which is generated at the $i$ th row and the $j$ th column in the matrix, is a random number between 0 and $1 . \mathbf{B}_{l}^{l l}=\left(b_{n_{1} n_{2}}^{l l \prime}\right)_{n_{l} \times n_{l}}$ is the reconnection submatrix of $V^{l} \rightarrow V^{l \prime}, l \in\{1,2, \ldots, L-2\}, l^{\prime} \in\{3,4, \ldots, L\}, n_{1}, n_{2} \in$ $\left\{1,2, \ldots, n_{l}\right\}$.

From the connection matrix $\mathbf{W}$, it can be seen that the neurons of the $l$ th layer connect the $(l+1)$ th layer's neurons with the probability $(1-p)$ and connect the neurons of
the layer following the $(l+1)$ th layer with the probability $p$. If the number of network layers and neurons is large and the rewiring probability $p$ is small, the connection matrix will be a sparse matrix. In short, the neurons of the $l$ th layer connect not only with the neurons of the $(l-1)$ th layer in multilayer feed forward small-world neural network model; they also connect with the neurons of the $1 \sim(l-2)$ th layers in accordance with rewiring probability $p$.

## 4. Control Simulation

Because of the strong ability in adaptive learning, neural network is usually used to design intelligent controller for complex industrial system. In addition, neural network can fully approximate any complex nonlinear system. Its identifier can also distinguish the characteristics of uncertain systems with high precision [28, 29]. In order to establish effective control for linear and nonlinear systems, this paper designs a controller which consists of two kinds of neural network model to obtain good performance. The controller structure is given in Figure 2. As shown in Figure 2, NNI identifies the controlled system online, $e_{1}$ is the error between the reference input and the controlled system output, $e_{2}$ is the error between the identification output and the controlled system output, and $e(t)$ is the error between the system input and the controlled system output. In NNI, the gradient information of $e_{2}$ is used to adjust the weight coefficient.

In Figure 2, NNI is an online identifier of the controlled system with regular structure topology and SNNC is the neural network controller with small-world structure topology. So the principle of small-world neural network control system can be described as follows: NNI identifies the controlled system online, by the use of identification result, then SNNC adjusts the weight coefficients using the identification result and outputs the control variable $u(t)$, and then $u(t)$ is applied to the controlled system and finally make the system output track the setting input to realize adaptive control.

Select the following linear differential system as a controlled system:

$$
\begin{align*}
y(k)= & 0.33 y(k-1)+0.132 y(k-2) \\
& +0.5 u(k-1)+0.038 u(k-2) . \tag{7}
\end{align*}
$$

SNNC network structure topology is selected as 3-6-61 (the number of input neurons is 3 , the first and second hidden neurons are 6 , and the output neuron is 1 ), NNI network structure topology is 3-6-1, the network weights and the thresholds are initialized to the range $[-1,1]$, incremental weight updating strategy is used in Back-Propagation algorithm, and set learning rate $\eta=0.1$, inertia coefficient $\alpha=$ 0.9 , the control period number is 400 , and the error criterion function is defined as follows:

$$
J(k)= \begin{cases}\frac{1}{2}[y(k)-r(k)]^{2}, & \text { SNNC, }  \tag{8}\\ \frac{1}{2}[y(k)-\hat{y}(k)]^{2}, & \text { NNI. }\end{cases}
$$



Figure 2: Controller with small-world neural network model.


Figure 3: Linear system ( $p=0$ ).

Select the input signal $r(k)$ as

$$
\begin{gather*}
r_{1}(k)=0.2 \sin \frac{2 \pi k}{26}+0.3 \sin \frac{\pi k}{15}+0.3 \sin \frac{\pi k}{75}+v_{1}(k) \\
r_{2}(k)=0.3 \sin \frac{2 \pi k}{50}+0.2 \sin \frac{2 \pi k}{100}+v_{2}(k) \tag{9}
\end{gather*}
$$

where,

$$
\begin{gather*}
v_{1}(k)= \begin{cases}0.5, & 50 \leq k \leq 150 \\
-0.5, & 150 \leq k \leq 250, \\
0.1 \times \operatorname{rand}(), & \text { others }\end{cases}  \tag{10}\\
v_{2}(k)=0.05 \times \operatorname{rand}()
\end{gather*}
$$

In the equation, $\operatorname{rand}()$ is a random number in the range of $[-1,1]$. In Figure 3, the input and output curve, the error curve are given when $p=0$ and Figure 4 shows the same curves when $p=0.1$.


From the figures, we can see that when constant disturbance is added ( $50 \leq k \leq 150,150 \leq k \leq 250$ ), neural network control systems all have better ability of selfadapting both $p=0$ and $p=0.1$, regardless of the rewiring probability. The neural network controller can adapt to the impact of constant disturbance through self-learning and adjusting connection weights so quickly that the output of the controlled system catches up with the input signal fast, and it can obtain a smaller error. When adding random disturbance to the input signal, especially within $250 \leq k \leq 400$, the control error of $p=0$ is bigger than that of $p=0.1$ at least by $30 \%$, which means that through rewiring, small-world neural network control system has good anti-interference capability.

Select the following nonlinear system as a controlled system:

$$
\begin{equation*}
y(k)=\frac{1.2 y(k-1)}{\left[2.5+y^{2}(k-1)\right]}+u^{3}(k-1) \tag{11}
\end{equation*}
$$

Figures 5 and 6 are given as the input signal and output curve, the error curve when $p=0, p=0.1$. Those diagrams also show that the error is largest when $p=0$. If the input is added with random disturbance, the error is relatively small when $p=0.1$. So small-world neural network control system has a better ability to suppress random disturbances. When the disturbance is a constant, the result is the same for linear system.

## 5. Experiment on Electrohydraulic Actuation System

Presently, the valve-controlled hydraulic servo systems have been applied widely. This kind of system has the features of large output power, fast response, and high accuracy.


Figure 5: Nonlinear system ( $p=0$ ).


Figure 6: Nonlinear system ( $p=0.1$ ).

However, it also has some shortcomings, such as low reliability, low efficiency, high requirements for oil cleanness, high manufacturing precision for electrohydraulic servo valve, and so on. In order to improve this situation, in recent years, the direct drive electrohydraulic servo system has been widely investigated. A swash plate mechanism is used to control the flow of hydraulic pump in direct drive electrohydraulic servo system. The change of the oblique angle of swash plate, which is regulated by the speed of the direct current motor, may help direct drive variable displacement electrohydraulic position servo control system to change its output flow


Figure 7: Schematic of direct drive variable displacement electrohydraulic actuation position servo control system (including hydraulic loading system).
and consequently serve to control the position of hydraulic actuator. Because of the high efficiency, easy operation, and big output force of this kind of system, it has been applied in many industry fields, such as electrohydraulic actuation system, precision forging machines, marine steering gear, and injection molding machines [30-34]. Due to existing nonlinear factors, it is difficult to establish the accurate mathematical model of this system. Figure 7 shows the schematic of direct drive variable displacement electrohydraulic actuation position control system. As shown in Figure 7, the hydraulic loading system is used to simulate the external load, and the control platform can control the hydraulic loading system to output different force value. The control objective of this electrohydraulic actuation system is to accurately track the given position signal, in which tracking performance, fast response, high accuracy, and good anti-interference must be taken into account. Therefore, the control strategy of this system is very important, and the small-world neural network controller is adopted to control this system given that it does not need accurate mathematical model.

Hardware configuration of electrohydraulic actuation system is listed in Table 1.

Firstly, in the case of no external load applied on direct drive variable displacement electrohydraulic actuation position system, the system is controlled to track square wave signal using small-world neural network. Figure 8 shows the tracking result of 1 Hz square wave signal with peak value of 200 mm , where the curve of input signal is marked A, and output is marked B. The following diagrams are similar. As shown in the figure, after 2 cycles of learning, the system output signal can track the target displacement, the response

TABLE 1: Hardware list of electro-hydraulic actuation system.

| Name | Type | Specification |
| :--- | :---: | :--- |
| Brushless DC <br> motor | BLF | Three phases 300 W |
| Displacement <br> sensor | LVDT | $\pm 400 \mathrm{~mm}$ |
| Angular <br> displacement <br> sensor | RVDT | $\pm 30^{\circ}$ |
| Swash plate <br> hydraulic pump | Displacement <br> pump | Swash plate angle: $\pm 20^{\circ}$ <br> Maximum displacement: <br> $84 \mathrm{~mL} / \mathrm{r}$ |
| Hydraulic cylinder | Single-acting <br> hydraulic cylinder | Stroke: 800 mm <br> Internal diameter: 50 mm |



Figure 8: Track control of 1 Hz square wave (no load).
time is about 0.2 s , and the steady state error is less than 0.2 mm .

This result shows that the use of small-world neural network control strategy for position control obtains satisfactory control precision and response speed in the case of noload. The structure topology of small-world neural network used in control test is the same as the simulation test.

Figure 9 shows the track control result by the use of small-world neural network control strategy under 10000 N external load. The hydraulic cylinder is controlled to track 2 Hz , peak value of 200 mm square wave signal. The result shows that the system has rapid response, about 0.25 s . The steady-state error is slightly larger, at about 2 mm . It indicates that the influence of external load will be further weakened by small-world neural network controller. Since the electrohydraulic actuation control system can also achieve fast tracking square wave signal, it can obtain satisfactory control accuracy under 10000 N external load.

The load under actual situation usually appears quickly and unexpectedly. In order to testify whether the small-world neural network control strategy performs well or not, we simulate the step response control in the case of impact load 5000 N. The result is shown in Figure 10.

As shown in the figure, at 3.7 s , the impact load 5000 N is added, and after about 0.25 s , the electrohydraulic actuation system can correct the error caused by the change of external load. The steady state error can be controlled within 0.2 mm or less. It shows that the small-world neural network


Figure 9: Track control of 2 Hz square wave ( 10000 N load).


Figure 10: Step control curve (impact load 5000 N).
controller has a better ability to suppress the impact of interference. The digital simulation results are basically consistent with the experimental results under external interference.

## 6. Conclusions

In this paper, a multilayer feed forward neural network model is proposed through reconnecting the regular links which relies heavily on the rewiring probability, and then a controller based on small-world neural network model is designed to control linear and nonlinear systems. Simulation results show that the regular and the small-world network all have better control performance under constant disturbance. But when adding random disturbance, the small-world neural network control system is superior to the corresponding regular neural network control system in control accuracy. So the multilayer small-world neural network control system has good anti-interference features. Furthermore, no matter there is load or not, the small-world neural network controller for direct drive electrohydraulic actuation position control system can obtain faster response, better control precision, and an ability of anti-interference, which means that small-world neural network can be used to develop intelligent controller for industrial systems. However, it still needs further study on how to obtain the optimal rewiring probability to make the best of the controller's performance.

## Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant no. 51105297.

## References

[1] D. Batalle, E. Eixarch, F. Figueras et al., "Altered smallworld topology of structural brain networks in infants with intrauterine growth restriction and its association with later neurodevelopmental outcome," NeuroImage, vol. 60, no. 2, pp. 1352-1366, 2012.
[2] M. Bolaños, E. M. Bernat, B. He, and S. Aviyente, "A weighted small world network measure for assessing functional connectivity," Journal of Neuroscience Methods, vol. 212, no. 1, pp. 133142, 2013.
[3] O. Sporns and C. J. Honey, "Small worlds inside big brains," Proceedings of the National Academy of Sciences of the United States of America, vol. 103, no. 51, pp. 19219-19220, 2006.
[4] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," Reviews of Modern Physics, vol. 74, no. 1, pp. 47-97, 2002.
[5] J. Lu, D. W. C. Ho, and Z. Wang, "Pinning stabilization of linearly coupled stochastic neural networks via minimum number of controllers," IEEE Transactions on Neural Networks, vol. 20, no. 10, pp. 1617-1629, 2009.
[6] J. Cao, P. Li, and W. Wang, "Global synchronization in arrays of delayed neural networks with constant and delayed coupling," Physics Letters A, vol. 353, no. 4, pp. 318-325, 2006.
[7] Y. Xia and J. Wang, "A general methodology for designing globally convergent optimization neural networks," IEEE Transactions on Neural Networks, vol. 9, no. 6, pp. 1331-1343, 1998.
[8] S. Hu, X. Liao, and X. Mao, "Stochastic Hopfield neural networks," Journal of Physics A, vol. 36, no. 9, pp. 2235-2249, 2003.
[9] K. O. Stanley, D. B. D'Ambrosio, and J. Gauci, "A hypercubebased encoding for evolving large-scale neural networks," Artificial Life, vol. 15, no. 2, pp. 185-212, 2009.
[10] O. R. de Lautour and P. Omenzetter, "Prediction of seismicinduced structural damage using artificial neural networks," Engineering Structures, vol. 31, no. 2, pp. 600-606, 2009.
[11] E. Gelenbe and J. M. Fourneau, "Random neural networks with multiple classes of signals," Neural Computation, vol. 11, no. 4, pp. 953-963, 1999.
[12] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'smallworld9 networks," Nature, vol. 393, no. 6684, pp. 440-442, 1998.
[13] G. A. Pagani and M. Aiello, "The power grid as a complex network: a survey," Physica $A$, vol. 392, no. 11, pp. 2688-2700, 2013.
[14] F. Gerhard, G. Pipa, B. Lima, S. Neuenschwander, and W. Gerstner, "Extraction of network topology from multi-electrode recordings: is there a small-world effect?" Frontiers in Computational Neuroscience, vol. 5, no. 4, pp. 1-13, 2011.
[15] H. D. Rozenfeld, C. Song, and H. A. Makse, "Small-world to fractal transition in complex networks: a renormalization group approach," Physical Review Letters, vol. 104, no. 2, Article ID 025701, 2010.
[16] A. R. Backes, D. Casanova, and O. M. Bruno, "A complex network-based approach for boundary shape analysis," Pattern Recognition, vol. 42, no. 1, pp. 54-67, 2009.
[17] M. Karsai, M. Kivelä, R. K. Pan et al., "Small but slow world: how network topology and burstiness slow down spreading," Physical Review E, vol. 83, no. 2, Article ID 025102, 2011.
[18] C. Wu and B. Zhou, "Complex network analysis of tag as a social network," Journal of Zhejiang University, vol. 44, no. 11, pp. 21942197, 2010 (Chinese).
[19] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, "Scale-free brain functional networks," Physical Review Letters, vol. 94, no. 1, Article ID 018102, 2005.
[20] D. S. Bassett and E. Bullmore, "Small-world brain networks," Neuroscientist, vol. 12, no. 6, pp. 512-523, 2006.
[21] J. Piersa, F. Piekniewski, and T. Schreiber, "Theoretical model for mesoscopic-level scale-free self-organization of functional brain networks," IEEE Transactions on Neural Networks, vol. 21, no. 11, pp. 1747-1758, 2010.
[22] C. Li, "Memorizing morph patterns in small-world neuronal network," Physica A, vol. 388, no. 2-3, pp. 240-246, 2009.
[23] S. Lu, J. Fang, A. Guo, and Y. Peng, "Impact of network topology on decision-making," Neural Networks, vol. 22, no. 1, pp. 30-40, 2009.
[24] O. Erkaymaz, M. Özer, and N. Yumuşakc, "Performance analysis of a feed forward artifical neural network with small-world topology," Procedia Technology, vol. 1, pp. 291-296, 2012.
[25] D. Simard, L. Nadeau, and H. Kröger, "Fastest learning in smallworld neural networks," Physics Letters A, vol. 336, no. 1, pp. 815, 2005.
[26] G. Trtnik, F. Kavčič, and G. Turk, "Prediction of concrete strength using ultrasonic pulse velocity and artificial neural networks," Ultrasonics, vol. 49, no. 1, pp. 53-60, 2009.
[27] V. Kůrková, P. C. Kainen, and V. Kreinovich, "Estimates of the number of hidden units and variation with respect to halfspaces," Neural Networks, vol. 10, no. 6, pp. 1061-1068, 1997.
[28] Z. Z. Yuan, Z. Q. Chen, and X. Li, "Connectionism intelligent control: a survey," Acta Automatica Sinica, vol. 28, no. 1, pp. 3859, 2002 (Chinese).
[29] H. M. R. Ugalde, J. C. Carmona, V. M. Alvarado, and J. ReyesReyes, "Neural network design and model reduction approach for black box nonlinear system identification with reduced number of parameters," Neurocomputing, vol. 101, pp. 170-180, 2013.
[30] S. Habibi, R. Burton, and E. Sampson, "High precision hydrostatic actuation systems for micro- and nanomanipulation of heavy loads," Transactions of the ASME, vol. 128, no. 4, pp. 778787, 2006.
[31] S. R. Habibi and R. Burton, "Parameter identification for a highperformance hydrostatic actuation system using the variable structure filter concept," Transactions of the ASME, vol. 129, no. 2, pp. 229-235, 2007.
[32] J. L. Liu, J. H. Jiang, T. R. Lu, Q. H. Liu, C. W. Zhang, and H. G. Xu, "Direct drive volume control electro-hydraulic servo active mass driver system," Journal of Harbin Institute of Technology, vol. 43, no. 9, pp. 51-55, 2011 (Chinese).
[33] K. M. Elbayomy, Z. Jiao, and H. Zhang, "PID controller optimization by GA and its performances on the electro-hydraulic servo control system," Chinese Journal of Aeronautics, vol. 21, no. 4, pp. 378-384, 2008.
[34] J. Yao, Z. Jiao, Y. Shang, and C. Huang, "Adaptive nonlinear optimal compensation control for electro-hydraulic load simulator," Chinese Journal of Aeronautics, vol. 23, no. 6, pp. 720-733, 2010.

Research Article

# A Novel Coherence Reduction Method in Compressed Sensing for DOA Estimation 

Jing Liu, ChongZhao Han, XiangHua Yao, and Feng Lian<br>School of Electronics and Information Engineering, Xian Jiaotong University, Xian 710049, China<br>Correspondence should be addressed to Jing Liu; elelj20080730@gmail.com

Received 14 March 2013; Accepted 9 May 2013
Academic Editor: Xianxia Zhang
Copyright © 2013 Jing Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

A novel method named as coherent column replacement method is proposed to reduce the coherence of a partially deterministic sensing matrix, which is comprised of highly coherent columns and random Gaussian columns. The proposed method is to replace the highly coherent columns with random Gaussian columns to obtain a new sensing matrix. The measurement vector is changed accordingly. It is proved that the original sparse signal could be reconstructed well from the newly changed measurement vector based on the new sensing matrix with large probability. This method is then extended to a more practical condition when highly coherent columns and incoherent columns are considered, for example, the direction of arrival (DOA) estimation problem in phased array radar system using compressed sensing. Numerical simulations show that the proposed method succeeds in identifying multiple targets in a sparse radar scene, where the compressed sensing method based on the original sensing matrix fails. The proposed method also obtains more precise estimation of DOA using one snapshot compared with the traditional estimation methods such as Capon, APES, and GLRT, based on hundreds of snapshots.


## 1. Introduction

Compressed sensing has received considerable attention recently and has been applied successfully in diverse fields, for example, image processing [1], video technology [2], wireless communication [3], and radar systems [4-10]. The central goal of compressed sensing is to capture attributes of a signal using very few measurements. In most work to date, this broader objective is exemplified by the important special case in which a $K$-sparse vector $x \in R^{N}$ (with $N$ large) is to be reconstructed from a small number $M$ of linear measurements with $K<M<N$. The two fundamental questions in compressed sensing are how to construct suitable sensing matrices $\Phi$ and how to recover $x$ from $y$ efficiently.

In early work of compressed sensing, the entries of the sensing matrix are generated by an i.i.d Gaussian or Bernoulli process or from random Fourier ensembles [1113]. The role of random measurement provides the worst case performance guarantees in the context of an adversarial signal/error model. Random sensing matrices are
easy to construct and are $2 K$-RIP with high probability [13].

With the application area of compressed sensing extended to wider fields, the random sensing matrix is replaced by more structured sensing matrix. Most of the recent compressed sensing work related to sensing matrix construction focuses on the construction of structured matrices which often exhibit a considerable structure [14]. This largely follows from efforts to model the way the samples are acquired in practice, which leads to sensing matrices that inherent their structure from the real world. However, most of the structured sensing matrices based on the practical acquisition equipment do not satisfy the RIP property, which guarantees the perfect reconstruction of the original sparse signal with large probability.

In this paper, we are considering changing the original sensing matrix into a random Gaussian matrix or a matrix with low coherence via some software-based algorithm in the reconstruction side. Firstly, a novel method, named as coherent column replacement method, is proposed to reduce
the coherence of a partially deterministic sensing matrix. The proposed method is to replace the highly coherent columns in the original sensing matrix $\Phi$ with random Gaussian columns to obtain a new sensing matrix $\Phi^{\prime}$. The measurement vector $y$ is changed accordingly. It is proved that the original sparse signal $x$ could be reconstructed well from the newly changed measurement vector $y^{\prime}$ based on the new sensing matrix $\Phi^{\prime}$ with large probability.

The proposed column replacement method is then extended to a more practical condition when highly coherent columns and incoherent columns are considered, for example, the direction of arrival (DOA) estimation problem in phased array radar system. The applications of compressed sensing to radar systems are investigated in [4-10]. In [4], it is demonstrated that the compressed sensing method could eliminate the need for match filter at the receiver and has the potential to reduce the required sampling rate. In the context of ground penetrating radar (GPR), [5] presents a compressed sensing based data acquisition and imaging algorithm. By exploiting the sparsity of targets in the spatial space, the proposed algorithm could generate sharper target space images with much less compressed sensing measurements than the standard back projection methods. Also the sparsity of targets in the time-frequency plane is exploited for radar in [6, 7]. In [8], compressed sensing is used to identify targets in a passive radar system. There are plenty of work concerning compressed sensing based phased array radar in recent years [ $9,15,16$ ]. In [9], the author puts focus on the generalization of the radar signal model for compressed sensing and does not provide realizable procedures. In [15], the authors address the narrow-band source localization problem for arbitrary arrays with known geometry in the presence of arbitrary noise of unknown spatial spectral density. In [16], the authors present a source localization method based on a sparse representation of sensor measurements with an overcomplete basis composed of samples from the array manifold.

Most of the present work for compressed sensing radar systems concentrates in designing the transmitted wave to implement a sensing matrix $\Phi$ with low coherence [10]. However, in phased array radar system, the steering matrix is deterministic and cannot be changed in practice. Therefore, it is required to develop a novel compressed sensing method which brings a little change to the existing hardware system of the phased array radar.

In the proposed method developed for compressed sensing based phased array radar system, a hybrid system is built with a bottom subsystem and a top subsystem for reconstruction. The bottom subsystem consists of a hardware specific sensing matrix $\Phi$, an acquired measurement vector $y$, and the original signal $x$. The original sensing matrix $\Phi$ and measurement vector $y$ are input to the software-based top subsystem, where a new sensing matrix $\Phi^{\prime}$ and a new measurement vector $y^{\prime}$ are generated. Since the new sensing matrix $\Phi^{\prime}$ is with low coherence, the original signal $x$ could be reconstructed from the new measurement vector $y^{\prime}$ perfectly with large probability. There are three key points to be aware of with this compressed sensing based approach for phased array radar system as follows. (1) There is no requirement
for the transmitted signal; it could be either "incoherent" or "coherent." (2) This approach does not use a matched filter. (3) The beamforming procedure is omitted in the proposed compressed phased array radar system.

The rest of the sections are organized as follows: the proposed coherent column replacement method is introduced in Section 2. In Section 3, the proposed method is extended to solve the DOA estimation problem in phased array radar system. Firstly, the signal model for DOA estimation in phased array radar system is represented in a standard compressed sensing form in Section 3.1, where the sparse radar scene is abstracted as a sparse signal. In Section 3.2, the proposed method is then used to generate a new sensing matrix with low coherence, based on which the original sparse signal could be reconstructed well with large probability. The simulation results are listed in Section 4, and the paper is summarized in Section 5.

## 2. The Proposed Coherent Column Replacement Method

Recent work related with structured sensing matrix construction tries to change the existing hardware system to generate a new sensing matrix satisfying the RIP property. In this paper, we are exploring the possibility of changing the sensing matrix and measurement vector in the reconstruction side while not changing the hardware system.

It is assumed that the original sensing matrix $\Phi$ is a partially deterministic matrix which is comprised of highly coherent columns and random Gaussian columns, and this could be extended to a more practical condition when highly coherent columns and incoherent columns are considered. For the original sensing matrix $\Phi$, its highly coherent columns and random Gaussian columns are denoted by $\varphi_{j}, j=1, \ldots, N_{c}$ and $\psi_{j}, j=1, \ldots, N_{r}$, respectively. $N_{c}$ is the number of highly coherent columns, and $N_{r}$ is the number of random Gaussian columns with $N_{c}+N_{r}=N$. Without loss of generality, the highly coherent columns are put at the leftmost of the sensing matrix, while the random Gaussian columns are put next to them as $\Phi=\left[\varphi_{1}, \varphi_{2}, \ldots, \varphi_{N_{c}}, \psi_{1}, \psi_{2}, \ldots, \psi_{N_{r}}\right]$. Accordingly, the signal $x$ could be divided into two groups, $x=$ $\left[x_{c} ; x_{r}\right]^{T}=\left[x_{c, 1}, x_{c, 2}, \ldots, x_{c, N_{c}}, x_{r, 1}, x_{r, 2}, \ldots, x_{r, N_{r}}\right]^{T}$, where $x_{c}=\left[x_{c, 1}, x_{c, 2}, \ldots, x_{c, N_{c}}\right]^{T}$ and $x_{r}=\left[x_{r, 1}, x_{r, 2}, \ldots, x_{r, N_{r}}\right]^{T}$ correspond to the highly coherent columns and random Gaussian columns respectively. It is assumed that the value of each element of $x_{c},\left\{x_{c, 1}, x_{c, 2}, \ldots, x_{c, N_{c}}\right\}$, is chosen as either one or zero.

The sensing matrix $\Phi$ is changed into a random Gaussian matrix $\Phi^{\prime}$ through replacing the highly coherent columns $\left[\varphi_{1}, \varphi_{2}, \ldots, \varphi_{N_{c}}\right]$ with random Gaussian columns $\left[\varphi_{1}^{\prime}, \varphi_{2}^{\prime}, \ldots, \varphi_{N_{c}}^{\prime}\right]$. The resulting new sensing matrix $\Phi^{\prime}$ could be represented as $\Phi^{\prime}=\left[\varphi_{1}^{\prime}, \varphi_{2}^{\prime}, \ldots, \varphi_{N_{c}}^{\prime}, \psi_{1}, \psi_{2}, \ldots, \psi_{N_{r}}\right]$.

Lemma 1. Given the standard model in compressed sensing $y=\Phi x+e$, where $e$ denotes the measurement noise, and given the sensing matrix $\Phi^{\prime}$ defined above, the $K$-sparse signal $x$ could be reconstructed from the new measurement
vector $y^{\prime}=\Phi^{\prime} x+e$ perfectly with large probability. The new measurement vector $y^{\prime}$ could be calculated via (1) provided that part of the signal $x, x_{c}=\left[x_{c, 1}, x_{c, 2}, \ldots, x_{c, N_{c}}\right]^{T}$ is known:

$$
\begin{align*}
y^{\prime}= & y+x_{c, 1}\left(\varphi_{1}^{\prime}-\varphi_{1}\right)+x_{c, 2}\left(\varphi_{2}^{\prime}-\varphi_{2}\right) \\
& +\cdots+x_{c, N_{c}}\left(\varphi_{N_{c}}^{\prime}-\varphi_{N_{c}}\right) . \tag{1}
\end{align*}
$$

Proof. Obviously, the new sensing matrix $\Phi^{\prime}$ is a Gaussian random matrix and satisfies the RIP property. We could reconstruct $x$ perfectly from the new measurement vector $y^{\prime}=\Phi^{\prime} x+e$ with large probability.

The equation $y^{\prime}=\Phi^{\prime} x+e$ could be expanded in columns:

$$
\begin{align*}
y^{\prime}= & \Phi^{\prime} x \\
= & x_{c, 1} \varphi_{1}^{\prime}+x_{c, 2} \varphi_{2}^{\prime}+\cdots+x_{c, N_{c}} \varphi_{N_{c}}^{\prime}  \tag{2}\\
& +x_{r, 1} \psi_{1}+x_{r, 2} \psi_{2}+\cdots+x_{r, N_{r}} \psi_{N_{r}}+e
\end{align*}
$$

Similarly, the equation $y=\Phi x+e$ could be expanded as in

$$
\begin{align*}
y= & \Phi x \\
= & x_{c, 1} \varphi_{1}+x_{c, 2} \varphi_{2}+\cdots+x_{c, N_{c}} \varphi_{N_{c}}  \tag{3}\\
& +x_{r, 1} \psi_{1}+x_{r, 2} \psi_{2}+\cdots+x_{r, N_{r}} \psi_{N_{r}}+e .
\end{align*}
$$

Equation (3) subtracts (2), resulting in

$$
\begin{align*}
y^{\prime}-y= & x_{c, 1}\left(\varphi_{1}^{\prime}-\varphi_{1}\right)+x_{c, 2}\left(\varphi_{2}^{\prime}-\varphi_{2}\right) \\
& +\cdots+x_{c, N_{c}}\left(\varphi_{N_{c}}^{\prime}-\varphi_{N_{c}}\right) \tag{4}
\end{align*}
$$

So we can obtain the new measurement vector $y^{\prime}$ based on the original measurement vector $y$ and the error between the highly coherent columns and random columns:

$$
\begin{align*}
y^{\prime}= & y+x_{c, 1}\left(\varphi_{1}^{\prime}-\varphi_{1}\right)+x_{c, 2}\left(\varphi_{2}^{\prime}-\varphi_{2}\right)  \tag{5}\\
& +\cdots+x_{c, N_{c}}\left(\varphi_{N_{c}}^{\prime}-\varphi_{N_{c}}\right)
\end{align*}
$$

This ends the proof.
However, in reality the original signal $x$ is unknown, and it is difficult to obtain the exact value of $x_{c}\left(x_{c}=\right.$ $\left.\left[x_{c, 1}, x_{c, 2}, \ldots, x_{c, N_{c}}\right]^{T}\right)$ in advance. If the number of highly coherent columns is small (e.g. $N_{c} \leq 10$ ), we could list all the configurations of $x_{c}$ with each element's value chosen as one or zero. Based on each configuration, we could obtain a candidate signal using a reconstruction algorithm. The error between the true measurement and the estimate measurement based on each candidate signal is calculated and then normalized. The candidate signal with the smallest error is the one closest to the original sparse signal and is what we pursuit. The detailed procedure is listed in Algorithm 2.

Algorithm 2. The coherent column replacement method for a partially deterministic sensing matrix is as follows.
(1) The sensing matrix $\Phi$ is changed into a random Gaussian matrix $\Phi^{\prime}$ through replacing the highly coherent columns $\left[\varphi_{1}, \varphi_{2}, \ldots, \varphi_{N_{c}}\right]$ with random Gaussian columns $\left[\varphi_{1}^{\prime}, \varphi_{2}^{\prime}, \ldots, \varphi_{N_{c}}^{\prime}\right]$.
(2) List all the configurations of $x_{c}$ with each element's value chosen as one or zero. The total number of configurations $C N$ equals $2^{N_{c}}$. The $i$ th configuration could be represented as $\lambda^{i}\left(x_{c, 1}^{i}, x_{c, 2}^{i}, \ldots, x_{c, N_{c}}^{i}\right), i=$ $1, \ldots, C N$ and is abbreviated as $\lambda^{i}$ for briefness.
(3) For the $i$ th configuration $\lambda^{i}, i=1, \ldots, C N$, calculate the new measurement vector $y^{\prime, i}$ via (5) and obtain a candidate signal $x^{i}$ using a reconstruction algorithm as

$$
\begin{equation*}
x^{i}=\operatorname{Reconstruct}\left(y^{\prime, i}, \Phi^{\prime}\right) \tag{6}
\end{equation*}
$$

(4) For the $i$ th candidate signal $x^{i}, i=1, \ldots, C N$, calculate $\mathrm{ERR}^{i}$, which is defined as the normalized error between the true measurement and the estimate measurement based on $x^{i}$,

$$
\begin{equation*}
\operatorname{ERR}^{i}=\frac{\left\|y-\left(\Phi x^{i}\right)\right\|_{2}}{\|y\|_{2}} \tag{7}
\end{equation*}
$$

where $\|\cdot\|$ denotes the $l_{2}$-norm.
(5) Find the smallest one in $\left\{\operatorname{ERR}^{i}, i=1, \ldots, C N\right\}$ and define it as ERR ${ }^{\text {min }}$. The candidate signal corresponding to ERR ${ }^{\text {min }}$ is what we pursuit and is defined as $x^{\text {estimate }}$.

In the above algorithm, "Reconstruct" in (6) refers to any available reconstruction algorithm and the basis pursuit denoising (BPDN) method [14] is chosen as the reconstruction algorithm here.

## 3. Compressed Sensing Based DOA Estimation in Phased Array Radar System

In this section, the proposed column replacement method is extended to solve the DOA estimation problem in phased array radar system. Firstly, the signal model for DOA estimation in phased array radar system is represented in a standard compressed sensing form in Section 3.1, where the sparse radar scene is abstracted as a sparse signal. In Section 3.2, the proposed method is then used to generate a new sensing matrix with low coherence, based on which the original sparse signal could be reconstructed well with large probability.
3.1. Signal Model for DOA Estimation and Sparse Representation. Assume a phased array radar system consisting of half-wavelength spaced uniform linear arrays (ULA). Targets may appear at directions represented by DOA angles. The task of signal processing is to estimate the directions to the targets and the corresponding complex amplitudes (DOA estimation, see [17]). We assume that the other parameters
like range and Doppler frequency have been isolated before by appropriate processing.

The ULA of the phased array radar system consists of $M$ antennas, which are used to emit the transmitted signal $s(t)$. The $M \times 1$ received complex vector of array observations is defined as $F(t)=\left[f_{1}(t), \ldots, f_{M}(t)\right]^{T}$. Assuming a hypothetical target located at a DOA angle of $\theta$ in the far field, the received complex vector of array observations can be written as

$$
\begin{equation*}
F(t)=\beta(\theta) s(t) a(\theta)+E(t), \tag{8}
\end{equation*}
$$

where $\beta(\theta)$ is the reflection coefficient of the hypothetical target, and $E(t)$ is a $M \times 1$ complex Gaussian noise vector. $a(\theta)$ is the $M \times 1$ steering vector, which is defined as

$$
\begin{equation*}
a(\theta)=\left[1 e^{j(2 \pi d \sin \theta / \lambda)}, \ldots, e^{j(M-1)(2 \pi d \sin \theta / \lambda)}\right]^{T} \tag{9}
\end{equation*}
$$

where $d$ is the distance between the elements of the arrays, and $\lambda$ denotes wavelength.

Assuming $D$ targets are observed with reflection coefficients $\left\{\beta_{i}\right\}_{i=1}^{D}$ and DOA angles $\left\{\theta_{i}\right\}_{i=1}^{D}$, the $M \times 1$ received complex vector of array observations can be written as

$$
\begin{equation*}
F_{M}(t)=\sum_{i=1}^{D} \beta\left(\theta_{i}\right) s(t) a\left(\theta_{i}\right)+E_{M}(t) \tag{10}
\end{equation*}
$$

where $E_{M}(t)$ is a $M \times 1$ complex Gaussian noise vector. Equation (10) could be rewritten as

$$
\begin{equation*}
F_{M}(t)=A(\theta) S(t, \theta)+E_{M}(t) \tag{11}
\end{equation*}
$$

where $A(\theta)=\left[a\left(\theta_{1}\right), a\left(\theta_{2}\right), \ldots, a\left(\theta_{D}\right)\right]$ is a $M \times D$ steering matrix, and $S(t, \theta)=s(t)\left[\beta\left(\theta_{1}\right), \beta\left(\theta_{2}\right), \ldots, \beta\left(\theta_{D}\right)\right]^{T}$ denotes a $D \times 1$ reflection vector.

Since the radar scene is generally in practice sparse, compressed sensing is a valid candidate for estimating the DOA angles for multiple targets. To do so, the DOA angle plane is divided into $N$ fine grids, each cell generally with the same size $\Delta \theta$. The $i$ th grid represents the DOA angle of $\theta_{0}+(i-1) \Delta \theta$, where $\theta_{0}$ is the initial angle of the DOA plane. Each cell has a unique mathematical representation as well as physical explanation: for example, if a target's DOA angle occupies the $i$ th grid, its contribution could be uniquely written as $s(t) \beta\left(\theta_{0}+(i-1) \Delta \theta\right) \vec{a}\left(\theta_{0}+(i-1) \Delta \theta\right)$. Now, the DOA estimation problem is recast as the search for the grid cells in which the targets lie.

As the system has no knowledge of the numbers and locations of the targets, the information of all the grids in the DOA plane should be considered. Therefore, the steering matrix and reflection vector in (11) are extended to obtain the $M \times N$ extended steering matrix $\Phi$ and the $N \times 1$ extended reflection vector $x$, which are defined as $\Phi=\left[a\left(\theta_{0}\right), a\left(\theta_{0}+\right.\right.$ $\left.\Delta \theta), \ldots, a\left(\theta_{0}+(i-1) \Delta \theta\right) a\left(\theta_{0}+(N-1) \Delta \theta\right)\right]$ and $x=$ $s(t)\left[\beta\left(\theta_{0}\right) \beta\left(\theta_{0}+\Delta \theta\right), \ldots, \beta\left(\theta_{0}+(i-1) \Delta \theta\right) \beta\left(\theta_{0}+(N-1) \Delta \theta\right)\right]^{T}$. Since small numbers of grids are occupied by the targets, $x$ is a sparse vector with the $i$ th element defined as $x(i)=$ $s(t) \beta\left(\theta_{0}+(i-1) \Delta \theta\right)$ if the $i$ th grid is occupied by the target;
otherwise, $x(i)=0$. As a result, the $M \times 1$ received complex vector of array observations $y$ could be written as follows:

$$
\begin{equation*}
y=\Phi x+e \tag{12}
\end{equation*}
$$

where $e$ is a $M \times 1$ complex Gaussian noise vector. Though in (12) the radar vectors and matrices are complex valued in contrary to the original compressed sensing environment, it is easy to transfer it to real variables according to $[9,18]$.

Discussion. In $[10,19]$, it is assumed that the discretized step is small enough so that each target falls on some specific grid point. However, no matter how finely the parameter space is gridded, the sources may not lie in the center of the grid cells, and consequently there is a mismatch between the assumed and the actual bases for sparsity. The sensitivity of compressed sensing to mismatch between the assumed and the actual sparsity bases is studied in [20]. The effect of basis mismatch is analyzed on the best $k$-term approximation error, and some achievable bounds for the $l_{1}$ error of the best $k$-term approximation are provided. The readers can refer to [20] for a detailed analysis on the influence of the griding operations on the estimation performance.
3.2. DOA Estimation Based on the Column Replacement Method. The proposed column replacement method is then extended to solve the DOA estimation problem in phased array radar system where the sensing matrix is comprised of highly coherent columns and incoherent columns. In order to distinguish the highly coherent columns from the incoherent columns, the coherence of the sensing matrix is adopted, which is defined in $[21,22]$ as follows.

Definition 3. For the sensing matrix $\Phi$, its coherence is defined as the largest absolute and normalized inner product between different columns in $\Phi$. Formally, this reads

$$
\begin{equation*}
\mu\{\Phi\}=\max _{1 \leq i, j \leq N, i \neq j} \frac{\left|\Phi_{i}^{T} \Phi_{j}\right|}{\left\|\Phi_{i}\right\| \cdot\left\|\Phi_{j}\right\|} \tag{13}
\end{equation*}
$$

The coherence provides a measure of the worst similarity between the sensing matrix columns.

A different way to understand the coherence is by considering the Gram matrix $G$ which is defined as

$$
\begin{equation*}
G=\widetilde{\Phi^{T}} \widetilde{\Phi} \tag{14}
\end{equation*}
$$

where $\widetilde{\Phi^{T}}$ is the normalized sensing matrix obtained from the original sensing matrix with each column normalized. The off-diagonal entries in $G$ are the inner products that appear in (13). The coherence is the off-diagonal entry $g_{i, j}$ with the largest magnitude.

In the proposed method, the Gram matrix $G$ is firstly built via (14), and a threshold $T$ is then set properly to distinguish the highly coherent columns from the incoherent columns as follows. For each off-diagonal entry $\left\{g_{i, j}, i=1, \ldots, N, j=\right.$ $1, \ldots, N\}$, if $g_{i, j}$ is larger than $T$, the columns $i$ and $j$ are added to the set of highly coherent columns. The remaining columns
that do not belong to the set of highly coherent columns form the set of incoherent columns.

The set of highly coherent columns and the set of incoherent columns are denoted by $\alpha_{j}, j=1, \ldots, N_{h c}$ and $\beta_{j}, j=1, \ldots, N_{i c}$, respectively. $N_{h c}$ is the number of highly coherent columns, and $N_{i c}$ is the number of incoherent columns with $N_{h c}+N_{i c}=N$. Without loss of generality, the highly coherent columns are put at the leftmost of the sensing matrix, while the incoherent columns are put next to them as $\Phi=\left[\alpha_{1}, \alpha_{2}, \ldots, \alpha_{N_{h c}}, \beta_{1}, \beta_{2}, \ldots, \beta_{N_{i c}}\right]$. Accordingly, the signal $x$ could be divided into two groups, $x=\left[x_{h c} ; x_{i c}\right]^{T}=$ $\left[x_{h c, 1}, x_{h c, 2}, \ldots, x_{h c, N_{h c}}, x_{i c, 1}, x_{i c, 2}, \ldots, x_{i c, N_{i c}}\right]^{T}$, where $x_{h c}=$ $\left[x_{h c, 1}, x_{h c, 2}, \ldots, x_{h c, N_{h c}}\right]^{T}$ and $x_{i c}=\left[x_{i c, 1}, x_{i c, 2}, \ldots, x_{i c, N_{i c}}\right]^{T}$ correspond to the highly coherent columns and incoherent columns respectively. It is assumed that the value of each element of $x_{h c},\left\{x_{h c, 1}, x_{h c, 2}, \ldots, x_{h c, N_{h c}}\right\}$ is chosen as either one or zero.

The sensing matrix $\Phi$ is changed into a new matrix $\Phi^{\prime}$ through replacing the highly coherent columns $\left[\alpha_{1}, \alpha_{2}, \ldots\right.$, $\alpha_{N_{h c}}$ ] with random Gaussian columns [ $\left.\alpha_{1}^{\prime}, \alpha_{2}^{\prime}, \ldots, \alpha_{N_{h c}}^{\prime}\right]$. The resulted new sensing matrix $\Phi^{\prime}$ could be represented as $\Phi^{\prime}=$ $\left[\alpha_{1}^{\prime}, \alpha_{2}^{\prime}, \ldots, \alpha_{N_{h c}}^{\prime}, \beta_{1}, \beta_{2}, \ldots, \beta_{N_{i c}}\right]$.

Theorem 4 (see [23]). Let $x$ be a $K$-sparse signal, and write $y=\Phi x+e$, where $e \sim N\left(0 ; \sigma^{2} I\right)$. Suppose that

$$
\begin{equation*}
K<\frac{1}{3 \mu(\Phi)} \tag{15}
\end{equation*}
$$

and consider the BPDN optimization problem (16) with $\lambda=$ $\sqrt{16 \sigma^{2} \log M}$ :

$$
\begin{equation*}
\widehat{x}=\underset{x \in R^{N}}{\arg \min }\|x\|_{1}+\lambda\|y-\Phi x\|_{2} . \tag{16}
\end{equation*}
$$

Then, with probability on the order of $1-1 / M^{2}$, the solution $\widehat{x}$ of (16) is unique, and its error is bounded by

$$
\begin{equation*}
\|x-\widehat{x}\|_{2} \leq C \sigma \sqrt{K \log M} \tag{17}
\end{equation*}
$$

and its support is a subset of the true $K$-element support of $x$.
Lemma 5. Given the standard model in compressed sensing $y=\Phi x+e$ and given the sensing matrix $\Phi^{\prime}$ defined above, the $K$-sparse signal $x$ could be reconstructed from the new measurement vector $y^{\prime}=\Phi^{\prime} x+e$ perfectly using the BPDN optimization method with probability on the order of $1-1 / M^{2}$. The new measurement vector $y^{\prime}$ could be calculated via (18) provided that part of the signal $x, x_{h c}=$ $\left[x_{h c, 1}, x_{h c, 2}, \ldots, x_{h c, N_{h c}}\right]^{T}$ is known:

$$
\begin{align*}
y^{\prime}= & y+x_{h c, 1}\left(\alpha_{1}^{\prime}-\alpha_{1}\right)+x_{h c, 2}\left(\alpha_{2}^{\prime}-\alpha_{2}\right)  \tag{18}\\
& +\cdots+x_{h c, N_{h c}}\left(\alpha_{N_{h c}}^{\prime}-\alpha_{N_{h c}}\right) .
\end{align*}
$$

Proof. The threshold $T$ is set to distinguish the highly coherent columns from the incoherent columns. In theory, the threshold $T$ could be designed as small as possible to obtain
a very small coherence $\mu(\Phi)$. As a consequence, the new sensing matrix $\Phi^{\prime}$ satisfies (15) properly with set threshold $T$. We could reconstruct $x$ perfectly from the new measurement vector $y^{\prime}=\Phi^{\prime} x+e$ using the BPDN optimization method with probability on the order of $1-1 / M^{2}$ according to Theorem 4. The deviation of (18) is similar to that of (1). This ends the proof.

However, in reality the original signal $x$ is unknown, and it is difficult to obtain the exact value of $x_{h c}\left(x_{h c}=\right.$ $\left[x_{h c, 1}, x_{h c, 2}, \ldots, x_{h c, N_{h c}}\right]^{T}$ ) in advance. If the number of highly coherent columns is small (e.g., $N_{h c} \leq 10$ ), we could list all the configurations of $x_{h c}$ with each element's value chosen as one or zero. Based on each configuration, we could obtain a candidate signal using a reconstruction algorithm. The error between the true measurement and the estimate measurement based on each candidate signal is calculated and then normalized. The candidate signal with the smallest error is the one closest to the original sparse signal and is what we pursuit. The detailed procedure is listed in Algorithm 6.

Algorithm 6. The coherent column replacement method for a deterministic sensing matrix is as follows.
(1) The Gram matrix $G$ is firstly built via (14), and a threshold $T$ is then set properly to distinguish the highly coherent columns from the incoherent columns as follows. For each off-diagonal entry $\left\{g_{i, j}, i=1, \ldots, N, j=1, \ldots, N\right\}$, if $g_{i, j}$ is larger than $T$, the columns $i$ and $j$ are added to the set of highly coherent columns. The remaining columns that do not belong to the set of highly coherent columns form the set of incoherent columns.
(2) The sensing matrix $\Phi$ is changed into a new sensing matrix $\Phi^{\prime}$ through replacing the highly coherent columns $\left[\alpha_{1}, \alpha_{2}, \ldots, \alpha_{N_{h c}}\right]$ with random Gaussian columns $\left[\alpha_{1}^{\prime}, \alpha_{2}^{\prime}, \ldots, \alpha_{N_{h c}}^{\prime}\right]$.

Steps (3)-(6) are the same as steps (2)-(5) in Algorithm 2.

## 4. Simulation Results and Analysis

In this section, a simple example is firstly carried out to verify the performance of the proposed column replacement algorithm for a partially deterministic sensing matrix. The proposed method is then extended to cope with the DOA estimation problem in phased array radar system.
4.1. A Simple Example. In this section, a simple example is used to evaluate the performance of the proposed algorithm. The original sensing matrix $\Phi$ is a $20 \times 30$ matrix with $5\left(N_{c}\right)$ highly coherent columns, which are put at the leftmost of it, and $25\left(N_{r}\right)$ random Gaussian columns. The original signal $x$ is shown in Figure 1, which shows that its nonzero entries are in indexes $\{1,3,5,11,17,20,27\}$. Since the highly coherent columns are put at the leftmost of $\Phi$, the true value of $x_{c}$ is $x_{c}=\left\{x_{c, 1}=1, x_{c, 2}=0, x_{c, 3}=\right.$ $\left.1, x_{c, 4}=0, x_{c, 5}=1\right\}$ (abbreviated as 10101). In the proposed method, each element of $x_{c},\left\{x_{c, 1}, x_{c, 2}, \ldots, x_{c, 5}\right\}$ is chosen as


Figure 1: The original signal $x$.


Figure 2: Reconstruction error based on configurations of $x_{c}$ in one trial.
one or zero, resulting in totally $2^{5}=32$ configurations as $\{00000,00001,00010, \ldots, 11111\}$. The reconstruction error obtained based on each configuration in one trial is shown in Figure 2, which shows that the configuration number with the smallest reconstruction error is 21 (10101 in binary format). This matches the true $x_{c}$ exactly. The whole reconstructed signal corresponding to configuration 21 is shown in Figure 3. Moreover, five hundred Monte Carlo simulations are carried out, and the average reconstruction error is shown in Figure 4, which shows that the configuration 21 is with the smallest average reconstruction error 0.06868 .
4.2. DOA Estimation Based on the Proposed Column Replacement Method. In this section, a synthetic example about DOA estimation based on the phased array radar system is provided. A hybrid system is built which consists of a bottom subsystem and a top subsystem. The bottom subsystem is built based on the specific hardware structure of the phased


Figure 3: Reconstructed signal corresponding to the configuration with the smallest reconstruction error.


Figure 4: Average reconstruction error based on configurations of $x_{c}$.
array system, which consists of half-wavelength spaced uniform linear arrays (ULA). The number of transmit/receive antennas is 20. The antennas transmit independent orthogonal quadrature phase shift keyed (QPSK) waveforms, and the carrier frequency is 8.62 GHz . The SNR of the measurement noise is set to a fixed value $(20 \mathrm{~dB})$. The range of the DOA plane is $\left[0^{\circ}, 90^{\circ}\right]$, which is divided into 30 cells with the initial angle $\left(\theta_{0}\right)$ and angle interval $(\Delta \theta)$ equaling $0^{\circ}$ and $3^{\circ}$, respectively. A maximum of $L=512$ snapshots are considered at the receive node. Targets may appear at directions represented by DOA angles. The task of signal processing is to estimate the directions to the targets and the corresponding complex amplitudes (DOA estimation; see [17]).

In the proposed method, the Gram matrix $G$ of the original sensing matrix $\Phi$ is built via (14). In the next, we will set the threshold $T$ to distinguish the highly coherent


Figure 5: Histogram of the absolute off-diagonal entries of $G$ based on the original sensing matrix.
columns from the incoherent columns. In theory, the threshold $T$ is designed as small as possible to obtain a very small coherence $\mu(\Phi)$, which guarantees the perfect reconstruction of the $K$-sparse signal $x$. However, small $T$ will result in a large number of highly coherent columns, leading to a huge number of configurations of $x_{h c}$. This will increase the computing time dramatically. While the restrict selection of $T$ is true from a worst-case standpoint, it turns out that the coherence as defined previously does not do justice to the actual behavior of sparse representations and pursuit algorithms' performance. Thus, if we relax our expectations and allow a small fraction of failed reconstructions, then values substantially beyond the above bound are still leading to successful compressed sensing [24]. In this simulation example, the threshold $T$ is set as 0.6 , resulting in 10 highly coherent columns and 20 incoherent columns.

A new sensing matrix $\Phi^{\prime}$ is then generated, based on which a new Gram matrix $G^{\prime}$ is built. Figures 5 and 6 present the histograms of the absolute off-diagonal entries of $G$ and $G^{\prime}$ respectively. As can be seen, there is a shift towards the origin of the histogram after using the proposed method. The tail representing the higher values in Figure 5 disappears in Figure 6. Therefore the coherence of the new sensing matrix $\Phi^{\prime}$ is far less than that of the original sensing matrix $\Phi$.

Firstly, the performance of the proposed column replacement method is compared to the compressed sensing method using the original sensing matrix (abbreviated as standard compressed sensing) and other three commonly used estimation methods, the Capon, APES, and GLRT method, in one trial. Figure 7 shows the original scene, the modulus of the reflection coefficients $\beta_{k}$, as functions of the DOA. Figures $8,9,10,11$, and 12 correspond to the DOA estimates obtained via the standard compressed sensing method, the proposed method, Capon, APES, and GLRT, respectively. The proposed method and the standard compressed sensing method use one snapshot only, and the other three methods use 512 snapshots each. One can see that the presence of the four targets is clearly evident via the proposed method


Figure 6: Histogram of the absolute off-diagonal entries of $G^{\prime}$ based on the new sensing matrix, using a fixed threshold $t=0.6$.

TABLE 1: Performance comparison.

|  | Average <br> reconstruction error | Average RMSE <br> of DOA angles <br> (degree) |
| :--- | :---: | :---: |
| Standard compressed <br> sensing <br> The proposed <br> algorithm | 0.64 | 2.5 |

(Figure 9), while the standard compressed sensing method fails in identifying the targets (Figure 8). Secondly, five hundred Monte Carlo simulations are carried out, and in each trial four targets are located randomly within the DOA range of $\left[0^{\circ}, 90^{\circ}\right]$, and the corresponding reflection coefficients are set as $\left\{\beta_{k}=1, k=1, \ldots, 4\right\}$. The performance of the proposed method is compared to the standard compressed sensing method via the average reconstruction error and the average root mean square error (RMSE) [25] of the estimated DOA angles of all four targets. The results in Table 1 show that the proposed method is with less reconstruction error and RMSE. This shows that the proposed method outperforms the standard compressed sensing with more accurate estimated DOA angles.

## 5. Conclusion

In this paper, the coherent column replacement method is proposed to reduce the coherence of a partially deterministic sensing matrix, which is comprised of highly coherent columns and random Gaussian columns. The proposed method is then extended to a more practical condition when highly coherent columns and incoherent columns are considered, for example, the direction of arrival (DOA) estimation problem in phased array radar system using compressed sensing. Numerical simulations show that the proposed method obtains more precise estimation of DOA using one snapshot compared with the traditional estimation methods such as Capon, APES, and GLRT, based on hundreds of snapshots.


Figure 7: The original scene.


Figure 8: DOA estimation using the standard compressed sensing method.


Figure 9: DOA estimation using the proposed coherent column replacement method.


Figure 10: DOA estimation using Capon.


Figure 11: DOA estimation using APES.


Figure 12: DOA estimation using GLRT.

## Acknowledgments

This work was supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (61221063), the Natural Science Foundations of China (no. 61104051, no. 61074176, and no. 61004087), the Fundamental Research Funds for the Central Universities, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

[1] J. Romberg, "Imaging via compressive sampling: introduction to compressive sampling and recovery via convex programming," IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 14-20, 2008.
[2] Z. Liu, A. Elezzabi, and H. Zhao, "Maximum frame rate video acquisition using adaptive compressed sensing," IEEE Transactions on Circuits and Systems For Video Technology, vol. 21, no. 11, pp. 1704-1718, 2011.
[3] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: from subspace methods to compressed sensing," IEEE Transactions on Signal Processing, vol. 58, no. 3, pp. 17081721, 2010.
[4] R. Baraniuk and P. Steeghs, "Compressive radar imaging", in Proceedins of the IEEE Radar Conference, pp. 128-133, April 2007.
[5] A. Gurbuz, J. McClellan, and W. Scott, "Compressive sensing for GPR imaging," in Proceedings of the 41th Asilomar Conference on Signals, Systems and Computers, pp. 2223-2227, November 2007.
[6] M. Herman and T. Strohmer, "Compressed sensing radar," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2223-2227, Las Vegas, Nev, USA, 2008.
[7] G. Shi, J. Lin, X. Chen, F. Qi, D. Liu, and L. Zhang, "UWB echo signal detection with ultra-low rate sampling based on compressed sensing," IEEE Transactions on Circuits and Systems II, vol. 55, no. 4, pp. 379-383, 2008.
[8] C. R. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, "Signal processing for passive radar using OFDM waveforms," IEEE Journal on Selected Topics in Signal Processing, vol. 4, no. 1, pp. 226-238, 2010.
[9] J. H. Ender, "On compressive sensing applied to radar," Signal Processing, vol. 90, pp. 1402-1414, 2010.
[10] Y. Yu, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing-based MIMO radar," IEEE Transactions on Signal Processing, vol. 59, no. 11, pp. 5338-5352, 2011.
[11] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comтипications on Pure and Applied Mathematics, vol. 59, no. 8, pp. 12071223, 2006.
[12] D. L. Donoho, "Compressed sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306, 2006.
[13] E. Candès, "Compressive sampling", in Proceedings of the International Congress of Mathematicians, vol. 3, pp. 1433-1452, Madrid, Spain, 2006.
[14] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: from theory to applications," IEEE Transactions on Signal Processing, vol. 59, no. 9, pp. 4053-4085, 2011.
[15] J. J. Fuchs, "Linear programming in spectral estimation. Application to array processing," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 6, pp. 3161-3164, May 1996.
[16] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Transactions on Signal Processing, vol. 53, no. 8, pp. 3010-3022, 2005.
[17] A. Gurbuz, J. McClellan, and V. Cevher, "A compressive beamforming method," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2617-2620, Las Vegas, Nev, USA, 2008.
[18] I. Stojanovic, W. Karl, and M. Cetin, "Compressed sensing of mono-static and multi-static SAR," in Defense and Security Symposium, vol. 7337 of Proceedings of SPIE. Algorithms for Synthetic Aperture Radar Imagery XVI, Orlando, Fla, USA.
[19] J. Zhang, D. Zhu, and G. Zhang, "Adaptive compressed sensing radar oriented toward cognitive detection in dynamic sparse target scene," IEEE Transactions on Signal Processing, vol. 60, no. 4, pp. 1718-1729, 2012.
[20] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," IEEE Transactions on Signal Processing, vol. 59, no. 5, pp. 2182-2195, 2011.
[21] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $l^{1}$ minimization," Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 5, pp. 2197-2202, 2003.
[22] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," IEEE Transactions on Information Theory, vol. 49, no. 12, pp. 3320-3325, 2003.
[23] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5030-5043, 2010.
[24] M. Elad, "Optimized projections for compressed sensing," IEEE Transactions on Signal Processing, vol. 55, no. 12, pp. 5695-5702, 2007.
[25] Y. Bar-Shalom and T. E. Fortmann, Tracking and Data Association, vol. 179, Academic Press, San Diego, Calif, USA, 1988.

## Research Article

# Splitting Matching Pursuit Method for Reconstructing Sparse Signal in Compressed Sensing 

Liu Jing, ${ }^{1}$ Han ChongZhao, ${ }^{1,2}$ Yao XiangHua, ${ }^{1}$ and Lian Feng ${ }^{1,2}$<br>${ }^{1}$ The School of Electronic and Information Engineering, Xian Jiaotong University, Xian, Shaanxi 710049, China<br>${ }^{2}$ The MOE Key Lab for Intelligent and Networked Systems, The School of Electronic and Information Engineering, Xian Jiaotong University, Xian, Shaanxi 710049, China<br>Correspondence should be addressed to Liu Jing; elelj20080730@gmail.com

Received 14 March 2013; Accepted 16 April 2013
Academic Editor: Xianxia Zhang
Copyright © 2013 Liu Jing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


#### Abstract

In this paper, a novel method named as splitting matching pursuit (SMP) is proposed to reconstruct $K$-sparse signal in compressed sensing. The proposed method selects $F l(F l>2 K)$ largest components of the correlation vector $c$, which are divided into $F$ split sets with equal length $l$. The searching area is thus expanded to incorporate more candidate components, which increases the probability of finding the true components at one iteration. The proposed method does not require the sparsity level $K$ to be known in prior. The Merging, Estimation and Pruning steps are carried out for each split set independently, which makes it especially suitable for parallel computation. The proposed SMP method is then extended to more practical condition, e.g. the direction of arrival (DOA) estimation problem in phased array radar system using compressed sensing. Numerical simulations show that the proposed method succeeds in identifying multiple targets in a sparse radar scene, outperforming other OMP-type methods. The proposed method also obtains more precise estimation of DOA angle using one snapshot compared with the traditional estimation methods such as Capon, APES (amplitude and phase estimation) and GLRT (generalized likelihood ratio test) based on hundreds of snapshots.


## 1. Introduction

The standard noiseless model in compressed sensing is

$$
\begin{equation*}
y=\Phi x \tag{1}
\end{equation*}
$$

where $x \in \Re^{N}$ is a $K$-sparse signal $(K \ll N), y \in \mathfrak{R}^{M}$ is a measurement of $x$, and $\Phi$ is an $M \times N$ sensing matrix. The compressed sensing recovery problem is defined as follows: given $y$ and $\Phi$, find a signal $x$ within the class of interest satisfies (1) exactly. The compressed sensing recovery process consists of a search for the sparsest signal $x$ that yields the measurement $y$. By defining the $l_{0}$ "norm" of a vector $\|w\|_{0}$ as the number of nonzero entries in $w$, the simplest way to pose a recovery algorithm is using the optimization

$$
\begin{equation*}
x=\underset{y=\Phi w}{\arg \min }\|w\|_{0} \tag{2}
\end{equation*}
$$

Solutions to (2) therefore lead to algorithms for recovering $K$-sparse signals from $M$ linear measurements. In general,
the minimization of (2) is NP-hard. An alternative to the $l_{0}$ "norm" used in (2) is to use the $l_{1}$ "norm", defined as $\|w\|_{1}=$ $\sum_{n=1}^{N}|w(n)|$. The resulting adaptation of (2), known as basis pursuit (BP) [1], is formally defined as

$$
\begin{equation*}
x=\underset{y=\Phi w}{\arg \min }\|w\|_{1} . \tag{3}
\end{equation*}
$$

Since the $l_{1}$ "norm" is convex, (3) can be seen as a convex relaxation of (2). The optimization (3) can be modified to allow for noise in the measurements $y=\Phi x+e$, where $e$ denotes an $M \times 1$ measurement noise vector. We simply change the constraint on the solution to

$$
\begin{equation*}
x=\underset{\|y-\Phi w\|_{2} \leq \epsilon}{\arg \min }\|w\|_{1} \tag{4}
\end{equation*}
$$

where $\epsilon>\|e\|_{2}$ is an appropriately chosen bound on the noise magnitude. This modified optimization is known as basis pursuit with inequality constraints (BPIC) and is a quadratic program with polynomial complexity solvers
[2]. The Lagrangian relaxation of this quadratic program is written as

$$
\begin{equation*}
x=\arg \min \|w\|_{1}+\lambda\|y-\Phi w\|_{2} \tag{5}
\end{equation*}
$$

and is known as basis pursuit denoising (BPDN). There exist many efficient solvers to find BP, BPIC, and BPDN solutions; for an overview, see [3]. Unfortunately, the complexity of the linear programming algorithms for solving (3)~(5) is highly impractical for large-scale applications.

An alternative approach to sparse signal recovery is based on the idea of iterative greedy pursuit. The basic greedy algorithm is the matching pursuit (MP) [4]. OMP [5] is a variation of MP method, which adds a least-squares minimization step to MP method to obtain the best approximation over the chosen atoms. Unlike MP and OMP choosing just one atom at each time, ROMP, StOMP, SP, CoSaMP, and BAOMP methods choose several atoms at each iteration. Furthermore, regularized OMP (ROMP) [6], subspace pursuit (SP) [7], compressive sampling matching pursuit (CoSaMP) [8], and backtracking-based matching pursuit (BAOMP) [9] methods also use a two-step selection technique to carefully choose the atoms. While SP and CoSaMP have offered comparable theoretical reconstruction quality to the linear programming methods along with low reconstruction complexity, they require the sparsity level to be known for exact recovery. As an improvement, the BAOMP algorithm achieves the blind sparse signal reconstruction without requiring the sparsity level $K$. However, the BAOMP algorithm adopts two parameters (atom-adding constant $\mu_{1}$ and atom-deleting constant $\mu_{2}$ ) which are related with sparsity level and required to be tuned online.

In this paper, a novel method named as splitting matching pursuit (SMP) is proposed to reconstruct sparse signal in compressed sensing. The SP/CoSaMP algorithm assumes that the true indices in the support set of the original signal correspond to the large components of the correlation vector and choose $K / 2 K$ largest components at each iteration. However, in practice some true indices may correspond to a set of small components. The proposed method selects Fl ( $\mathrm{Fl}>2 \mathrm{~K}$ ) largest components of the correlation vector, which expands the searching area and increases the probability of finding the true components at one iteration. The Fl candidate components are divided into $F$ split sets with equal length $l$. The proposed method does not require the sparsity level $K$ to be known in prior. Different from BAOMP algorithm which adopts two tuning parameters related with sparsity level, the proposed algorithm uses two parameters $F$ and $l$ which are preset and determined by $N$ and $M$. The candidate components are divided into $F$ split sets, which could be processed simultaneously and suitable for parallel computation.

The proposed SMP method is then extended to more practical condition, for example, the direction of arrival (DOA) estimation problem in phased-array radar system using compressed sensing. Numerical simulations show that the proposed method succeeds in identifying multiple targets in a sparse radar scene, outperforming other OMP-type methods. The proposed method also obtains more precise
estimation of DOA angle using one snapshot compared with the traditional estimation methods such as Capon [10], APES [11], and GLRT [12] based on hundreds of snapshots.

The rest of the sections are organized as follows. The proposed SMP method is introduced in Section 2. The simulation results are listed in Section 3, and the paper is summarized in Section 4.

## 2. The Splitting Matching Pursuit Method

The most difficult part of signal reconstruction is to identify the locations of $K$ largest components in the target signal. The SP/CoSaMP algorithm assumes that the true indices in the support set of the original signal correspond to the large components of the correlation vector and choose $K / 2 K$ largest components at each iteration. However, in practice some true indices may correspond to a set of small components. The proposed SMP selects $F l(F l>2 K)$ largest components of the correlation vector, which expands the searching area and increases the probability of finding the true components at one iteration. In the following, a short introduction of the proposed method is given in Section 2.1, and the detailed algorithm is introduced in Section 2.2. The convergency analysis, parameters setting, complexity analysis, and convergency speed analysis of the proposed method are provided in Sections 2.3 to 2.6, respectively.
2.1. Brief Introduction to Splitting Matching Pursuit Method. A schematic diagram of the proposed algorithm is depicted in Figure 1. At the beginning of each iteration, the proposed method selects $\mathrm{Fl}(\mathrm{Fl}>2 \mathrm{~K})$ largest components of the correlation vector, which are divided into $F$ split sets with equal length $l$. Each split set is merged with the estimated support set from the previous iteration, resulting in a sequence of merged split sets. The proposed algorithm then approximates the target signal on each merged split set to obtain a sequence of split estimates using least-square algorithm. A set of pruned split sets are then built by retaining only the $l$ largest magnitude entries in the split estimates. The obtained pruned split sets are then combined to a final merged set, which contains as much as possible true components. An interim estimate is then calculated based on the final merged set using least-square algorithm. An estimated support set is obtained by retaining $l$ indices corresponding to the largest magnitude entries in the interim estimate, based on which a final estimate is generated. The iterations repeat if the $l_{2}$ norm (magnitude) of the calculated residual is less than a threshold $T$.

[^1]

FIgURE 1: Description of reconstruction procedures of the SMP method.

Algorithm 1 (the SMP method).
Input. Sensing matrix $\Phi$, measurement vector $y$, parameters $F$ and $l$, and threshold $T$ to halt the iterations.

Output. The estimated signal $x_{\text {out }}$.
Initialization
(1) $a^{0}=\emptyset$, where $a^{0}$ indicates the initial estimated support set and $\emptyset$ denotes empty set. $a$ is the estimated support set with length $l$.
(2) $r^{0}=y$, where $r^{0}$ denotes the initial residual.
(3) $c^{0}=\Phi^{*} r^{0}$, where $c^{0}$ denotes the initial correlation vector and $\Phi^{*}$ denotes the transpose of matrix $\Phi$.

Iteration. At the $k$ th iteration, go through the following steps.
(1) Splitting. Locate the Fl largest components of the correlation vector at the $(k-1)$ th iteration, and divide them into $F$ split sets as

$$
\begin{equation*}
I_{j}^{k}=c_{(j-1) l+1: j l}^{k-1}, \quad j=1, \ldots, F \tag{6}
\end{equation*}
$$

where $I_{j}^{k}$ denotes the $j$ th split set at the $k$ th iteration, and $c^{k-1}$ denotes the correlation vector at the $(k-1)$ th iteration. Furthermore, $c_{(j-1) l+1: j l}^{k-1}$ denotes the $((j-$ $1) l+1)$ th largest magnitude entry to the $(j l)$ th largest magnitude entry of $c^{k-1}, j=1, \ldots, F$.
(2) Support Merging. Each newly identified split set is united with the estimated support set from the previous iteration, resulting in a sequence of merged split sets as

$$
\begin{equation*}
J_{j}^{k}=I_{j}^{k} \cup a^{k-1}, \quad j=1, \ldots, F, \tag{7}
\end{equation*}
$$

where $J_{j}^{k}$ denotes the $j$ th merged split set at the $k$ th iteration, and $a^{k-1}$ denotes the estimated support set at the $(k-1)$ th iteration.
(3) Estimation. The proposed algorithm then solves a least square problem to approximate the nonzero
entries of the target signal on each merged split set $\left(J_{j}^{k}, j=1, \ldots, F\right)$ and sets other entries as zero, resulting in a sequence of split estimates as

$$
\begin{gather*}
\left.b_{j}^{k}\right|_{J_{j}^{k}}=\left(\Phi_{I_{j}^{k}}\right)^{\dagger} y,  \tag{8}\\
\left.b_{j}^{k}\right|_{\left(J_{j}^{k}\right)^{c}}=0, \quad j=1, \ldots, F,
\end{gather*}
$$

where $b_{j}^{k}$ denotes the $j$ th split estimate of the target signal at the $k$ th iteration. The vector $\left.b_{j}^{k}\right|_{J_{j}^{k}}$ is composed of the entries of $b_{j}^{k}$ indexed by $i \in J_{j}^{k}$, and $\left.b_{j}^{k}\right|_{\left(J_{j}^{k}\right)^{c}}$ is composed of the entries of $b_{j}^{k}$ indexed by $i \in\left(J_{j}^{k}\right)^{c}$. $\dagger$ indicates pseudoinverse operation. The matrix $\Phi_{J_{j}^{k}}$ consists of the columns of $\Phi$ with indices $i \in J_{j}^{k}$.
(4) Pruning. Obtain a sequence of pruned split sets via retaining only the largest $l$ indices corresponding to the largest magnitude entries in $\left.b_{j}^{k}\right|_{J_{j}^{k}}, j=1, \ldots, F$, for example,
$H_{j}^{k}=\left\{\right.$ indices of $l$ largest magnitude entries in $\left.\left.b_{j}^{k}\right|_{J_{j}^{k}}\right\}$,

$$
\begin{equation*}
j=1, \ldots, F \tag{9}
\end{equation*}
$$

where $H_{j}^{k}$ denotes the $j$ th pruned split set at the $k$ th iteration.
(5) Split Sets Merging. The pruned split sets are merged to form a final merged set, $G^{k}$, as

$$
\begin{equation*}
G^{k}=\operatorname{union}\left\{H_{1}^{k}, H_{2}^{k}, \ldots, H_{F}^{k}\right\} \tag{10}
\end{equation*}
$$

(6) Estimation. An interim estimate of the original signal is calculated based on the final merged set $G^{k}$ using least-square algorithm as

$$
\begin{gather*}
\left.q^{k}\right|_{G^{k}}=\left(G^{k}\right)^{\dagger} y,  \tag{11}\\
\left.q^{k}\right|_{\left(G^{k}\right)^{c}}=0,
\end{gather*}
$$

where $q^{k}$ denotes the interim estimate of the original signal at the $k$ th iteration. The vector $\left.q^{k}\right|_{G^{k}}$ is composed of the entries of $q^{k}$ indexed by $i \in G^{k}$, and $\left.q^{k}\right|_{\left(G^{k}\right)^{c}}$ is composed of the entries of $q^{k}$ indexed by $i \in\left(G^{k}\right)^{c}$.
(7) Pruning. Obtain the estimated support set by retaining $l$ indices corresponding to the largest magnitude entries in the vector $\left.q^{k}\right|_{G^{k}}$, as
$a^{k}=\left\{\right.$ indices of $l$ largest magnitude entries in $\left.\left.q^{k}\right|_{G^{k}}\right\}$,
where $a^{k}$ denotes the estimated support set at the $k$ th iteration.
(8) Estimation. Obtain the final estimate at each iteration, based on $a^{k}$ using least-square algorithm as

$$
\begin{gather*}
\left.x_{F}^{k}\right|_{a^{k}}=\left(\Phi_{a^{k}}\right)^{\dagger} y,  \tag{13}\\
\left.x_{F}^{k}\right|_{\left(a^{k}\right)^{c}}=0,
\end{gather*}
$$

where $x_{F}^{k}$ denotes the final estimate at the $k$ th iteration. The vector $\left.x_{F}^{k}\right|_{a^{k}}$ is composed of the entries of $x_{F}^{k}$ indexed by $i \in a^{k}$, and $\left.x_{F}^{k}\right|_{\left(a^{k}\right)^{c}}$ is composed of the entries of $x_{F}^{k}$ indexed by $i \in\left(a^{k}\right)^{c}$. The matrix $\Phi_{a^{k}}$ consists of the columns of $\Phi$ with indices $i \in a^{k}$.
(9) Residual calculation:

$$
\begin{equation*}
r^{k}=y-\left.\left(\Phi_{a^{k}}\right) x_{F}^{k}\right|_{a^{k}} \tag{14}
\end{equation*}
$$

where $r^{k}$ denotes the residual at the $k$ th iteration.
(10) If $\left\|r^{k}\right\|_{2}>T$, perform the correlation calculation $c^{k}=$ $\Phi^{*} r^{k}$, and then go to step (1) of the $(k+1)$ th iteration; otherwise, set $x_{\text {out }}=x_{F}^{k}$ and quit the iteration.
2.3. Convergency Analysis. Here, we will discuss the convergency of the proposed SMP method.

Theorem 2 (Theorem 2.1 in [8]). Let $x \in \mathfrak{R}^{N}$ be a $K$-sparse signal, and let its corresponding measurement be $y=\Phi x+$ $e \in \mathfrak{R}^{M}$. If the sampling matrix satisfies the restricted isometry property (RIP) with constant

$$
\begin{equation*}
\delta_{4 K}<0.1, \tag{15}
\end{equation*}
$$

then the signal approximation $\hat{x}^{k}$ is $K$-sparse and

$$
\begin{equation*}
\left\|x-\hat{x}^{k+1}\right\|_{2} \leq 0.5\left\|x-\hat{x}^{k}\right\|_{2}+10 v . \tag{16}
\end{equation*}
$$

In particular,

$$
\begin{equation*}
\left\|x-\hat{x}^{k}\right\|_{2} \leq 2^{-k}\|x\|_{2}+20 v, \tag{17}
\end{equation*}
$$

where $v$ denotes the unrecoverable energy in the signal.
Proposition 3. Let $x \in \mathfrak{R}^{N}$ be a $K$-sparse signal, and let its corresponding measurement be $y=\Phi x+e \in \mathfrak{R}^{M}$. If the sampling matrix satisfies the RIP with constant

$$
\begin{equation*}
\delta_{F l}<0.1, \tag{18}
\end{equation*}
$$

then the proposed SMP algorithm is guaranteed to recover $x$ from $y$ via a finite number of iterations.

Proof. The proving process is very similar to that of Theorem 2 (Theorem 2.1 in [8]). The CoSaMP algorithm selects the 2 K largest components of the correlation vector at each iteration, while the proposed SMP method selects the Fl largest components of the correlation vector. We can obtain the similar results through replacing $2 K$ with $F l$ in the derivation process in [8].
2.4. Parameters Setting. Here, we will discuss how to set values of the number of split sets $F$ and the length of the split set $l$.

Proposition 4. Note that $F l<4 K$ guarantees $\delta_{F l}<0.1$ if $\delta_{4 K}<0.1$.

Proof. Considering the monotonicity of $\delta_{K}$ : for any two integers $K \leq K^{\prime}, \delta_{K} \leq \delta_{K}^{\prime}$ according to Lemma 1 in [7]. So we have $\delta_{F l}<\delta_{4 K}$ provided that $F l<4 K$.

Proposition 5. Step (7) in Algorithm 1 guarantees $K \leq l$.
Proof. As the estimated support set, $a^{k}$ contains at least $K$ elements, resulting in $K \leq l$.

According to Propositions $3 \sim 5$, in order to guarantee a perfect recovery of the $K$-sparse vector $x$ from $y$ via a finite number of iterations, we have $F l<4 K \leq 4 l$, resulting in $F<4$. Since $F$ is set large enough to expand the searching area, $F$ is set as 3 in the proposed method. We then have $l<(4 / 3) K$ according to $F l<4 K$. Propositions $3 \sim 5$ are based on Theorem 2, which has a rigid setting for $\delta_{4 K}$. We can relax the range of $l$ to $[K, 2 K)$. As a result, we need not know the exact value of $K$ and can select an integer randomly from [ $K, 2 K$ ) based on an estimated value of $K$.
2.5. Complexity Analysis. The process of complexity analysis is similar to that of [7]. In each iteration, the correlation maximization procedure requires $M N$ computations in general, while the cost of computing the projections is of the order of $O\left(F(2 s)^{2} M+(2 F s)^{2} M\right)$, which could be approximated as $O\left(s^{2} M\right)$ when $F$ is chosen as a small value (e.g. 3 in the simulation setup). As a result, the total cost of computing is of the order of $O\left(M N+s^{2} M\right)$, which is comparable to the SP/CoSaMP algorithm.
2.6. Convergence Speed Analysis. Here, we will discuss the convergence speed of the proposed SMP method.

Theorem 6 (Theorem 8 in [7]). The number of iterations $\left(n_{i t}^{\mathrm{SP}}\right)$ of the SP algorithm is upper bounded by

$$
\begin{equation*}
n_{i t}^{S P} \leq \frac{1.5 K}{-\log \left(c_{K}\right)} \tag{19}
\end{equation*}
$$

where $c_{K}=2 \delta_{3 K}\left(1+\delta_{3 K}\right) /\left(1-\delta_{3 K}\right)^{3}$.
Proposition 7. The number of iterations $\left(n_{i t}^{S M P}\right)$ of the SMP algorithm is upper bounded by

$$
\begin{equation*}
n_{i t}^{S M P}<\frac{n_{i t}^{S P}}{F} \tag{20}
\end{equation*}
$$

Proof. The SP algorithm selects the $K$ largest components of the correlation vector at each iteration, while the proposed SMP method selects the $F l(F l \geq F K$ according to Proposition 5) largest components of the correlation vector. The searching area is thus expanded to at least $F$ times of that of SP algorithm. The probability of finding the true components at one iteration is increased to at least $F$ times of that of SP algorithm. As a result, the number of iterations is decreased to $1 / F$ of that of the SP algorithm by average according to Theorem 6.

## 3. Simulation Results and Analysis

In this section, a simple example is firstly carried out to verify the performance of the proposed SMP method in reconstructing zero-one binary and Gaussian signals. The proposed method is then extended to cope with the DOA estimation problem in phased-array radar system.
3.1. A Simple Example. The simulations are carried out to compare the accuracy of different reconstruction algorithms empirically. The proposed SMP algorithm is compared with some popular greedy algorithms (including OMP, ROMP, StOMP, SP, CoSaMP, and BAOMP algorithms) and the convex optimization algorithm (BP method).

In the simulation setup, a signal sparsity level $K$ is chosen such that $K \leq M / 2$ given the length of the original signal $N$ ( $N=256$ ) and the length of the measurement vector $M(M=$ 128). An $M \times N$ sampling matrix $\Phi$ is randomly generated from the standard i.i.d. Gaussian ensemble. A support set $S$ of size $K$ is selected uniformly at random, and the original sparse signal vector $x$ is chosen as either Gaussian signal or zero-one signal [7]. The estimate of the original signal, $\widehat{x}$, is computed based on the measurement vector $y$ generated through $y=\Phi x$. In the experiments, OMP uses $K$ iterations, StOMP and BP methods use the default settings (OMP, StOMP, and BP tools use SparseLab [13]), and ROMP and SP methods use the parameters given in $[6,7]$, respectively. The proposed SMP method use $F=3, l=84, n_{\max }=M, \varepsilon=10^{-5}$, as the input parameters.

The signal sparsity level $K$ is varied from 0 to $M / 2$. For each fixed $K$, five hundred Monte Carlo simulations


Figure 2: Zero-one signal.


Figure 3: Gaussian signal.
are carried out for each algorithm. The reconstruction is considered to be exact when the $l_{2}$ norm of the difference between the original signal $x$ and the reconstructed one $\widehat{x}$ is smaller than $10^{-5}$, that is, $\|x-\widehat{x}\|_{2}<10^{-5}$. The frequency of exact reconstruction $(\xi)$ is used to evaluate the reconstruction performance of the different methods, which is defined as

$$
\begin{equation*}
\xi=\frac{\alpha}{N_{\mathrm{MC}}} \tag{21}
\end{equation*}
$$

where $\alpha$ denotes the number of exact reconstructions for each algorithm given a fixed $K$, and $N_{\mathrm{MC}}$ denotes the number of Monte Carlo simulations. The frequency of exact reconstruction is also adopted by the SP method to evaluate the reconstruction performance [7].

Figures 2 and 3 show the reconstruction results for binary zero-one and Gaussian sparse signals, respectively. We only present the results of the SP algorithm since the SP and CoSaMP algorithms are almost the same with different
deviation process, and both obtain the same simulation results. As can be seen in Figure 2, for binary zero-one sparse signal which is a difficult case for OMP-type methods, the performance of the proposed SMP method is much better than all other OMP-type methods and comparable to the BP minimization method. Of particular interest is the sparsity level at which the recovery rate drops below $100 \%$, that is, the critical sparsity defined in [7]. It could be seen from Figure 2 that the proposed method is with the largest critical sparsity (39), which exceeds that of the BP method (36). For the Gaussian sparse signal, as shown in Figure 3, the proposed method also gives the comparable performance to the BP method.
3.2. DOA Estimation Based on Splitting Matching Pursuit Method. In this section, the proposed SMP method is extended to solve the DOA estimation problem in phasedarray radar system. The signal model for DOA estimation in phased-array radar system is represented in a standard compressed sensing form in Section 3.2.1, where the sparse radar scene is abstracted as a sparse signal. And the simulation results are shown in Section 3.2.2, which shows that the proposed method succeeds in identifying multiple targets in a sparse radar scene.
3.2.1. Signal Model for DOA Estimation and Sparse Representation. Assume that a phased-array radar system consists of half wavelength spaced uniform linear arrays (ULAs). Targets may appear at directions represented by DOA angles. The task of signal processing is to estimate the directions to the targets and the corresponding complex amplitudes (DOA estimation, see [14]). We assume that the other parameters like range and Doppler frequency have been isolated before by appropriate processing.

The ULA of the phased-array radar system consists of $M$ antennas, which are used to emit the transmitted signal $s(t)$. The $M \times 1$ received complex vector of array observations is defined as $f(t)=\left[f_{1}(t), \ldots, f_{M}(t)\right]^{T}$. Assuming a hypothetical target located at a DOA angle of $\theta$ in the far field, the received complex vector of array observations can be written as

$$
\begin{equation*}
f(t)=\beta(\theta) s(t) a(\theta)+n(t) \tag{22}
\end{equation*}
$$

where $\beta(\theta)$ is the reflection coefficient of the hypothetical target, and $n(t)$ is an $M \times 1$ complex Gaussian noise vector. $a(\theta)$ is the $M \times 1$ steering vector, which is defined as

$$
\begin{equation*}
a(\theta)=\left[1 e^{j(2 \pi d \sin \theta / \lambda)}, \ldots, e^{j(M-1)(2 \pi d \sin \theta / \lambda)}\right]^{T} \tag{23}
\end{equation*}
$$

where $d$ is the distance between the elements of the arrays and $\lambda$ denotes wavelength.

Assuming $D$ targets are observed with reflection coefficients $\left\{\beta_{i}\right\}_{i=1}^{D}$ and DOA angles $\left\{\theta_{i}\right\}_{i=1}^{D}$, the $M \times 1$ received complex vector of array observations can be written as

$$
\begin{equation*}
\alpha(t)=\sum_{i=1}^{D} \beta\left(\theta_{i}\right) s(t) a\left(\theta_{i}\right)+\gamma(t) \tag{24}
\end{equation*}
$$

where $\gamma(t)$ is an $M \times 1$ complex Gaussian noise vector. Equation (24) could be rewritten as

$$
\begin{equation*}
\alpha(t)=A(\theta) S(t, \theta) \gamma(t) \tag{25}
\end{equation*}
$$

where $A(\theta)=\left[a\left(\theta_{1}\right) a\left(\theta_{2}\right) \cdots a\left(\theta_{D}\right)\right]$ is an $M \times D$ steering matrix, and $S(t, \theta)=s(t)\left[\beta\left(\theta_{1}\right) \beta\left(\theta_{2}\right) \cdots \beta\left(\theta_{D}\right)\right]^{T}$ denotes a $D \times 1$ reflection vector.

Since the radar scene is generally in practice sparse, compressed sensing is a valid candidate for estimating the DOA angles for multiple targets. To do so, the DOA angle plane is divided into $N$ fine grids, each cell generally with the same size $\Delta \theta$. The $i$ th grid represents the DOA angle of $\theta_{0}+(i-1) \Delta \theta$, where $\theta_{0}$ is the initial angle of the DOA plane. The steering matrix and reflection vector in (25) are extended to obtain the $M \times N$ extended steering matrix $\Phi$ and the $N \times 1$ extended reflection vector $x$, which are defined as $\Phi=\left[a\left(\theta_{0}\right) a\left(\theta_{0}+\Delta \theta\right) \cdots a\left(\theta_{0}+(i-1) \Delta \theta\right) a\left(\theta_{0}+(N-1) \Delta \theta\right)\right]$ and $x=s(t)\left[\beta\left(\theta_{0}\right) \beta\left(\theta_{0}+\Delta \theta\right) \cdots \beta\left(\theta_{0}+(i-1) \Delta \theta\right) \beta\left(\theta_{0}+\right.\right.$ $(N-1) \Delta \theta)]^{T}$. Since small number of grids are occupied by the targets, $x$ is a sparse vector with the $i$ th element defined as $x(i)=s(t) \beta\left(\theta_{0}+(i-1) \Delta \theta\right)$ if the $i$ th grid is occupied by the target; otherwise, $x(i)=0$. As a result, the $M \times 1$ received complex vector of array observations $y$ could be written as

$$
\begin{equation*}
y=\Phi x+e \tag{26}
\end{equation*}
$$

where $e$ is an $M \times 1$ complex Gaussian noise vector. Though in (26) the radar vectors and matrices are complex valued in contrary to the original compressed sensing environment, it is easy to transfer it to real variables according to $[15,16]$.

Discussion. In $[17,18]$, it is assumed that the discretized step is small enough so that each target falls on some specific grid point. However, no matter how finely the parameter space is gridded, the sources may not lie in the center of the grid cells, and consequently there is mismatch between the assumed and the actual bases for sparsity. The sensitivity of compressed sensing to mismatch between the assumed and the actual sparsity bases is studied in [19]. The effect of basis mismatch is analyzed on the best $k$-term approximation error, and some achievable bounds for the $l_{1}$ error of the best $k$-term approximation are provided. The readers can refer to [19] for a detailed analysis on the influence of the griding operations on the estimation performance.
3.2.2. Simulation Results. In this section, an example about DOA estimation is provided based on the phased-array radar system, which consists of half wavelength spaced uniform linear arrays (ULAs). The number of transmit/receive antennas is 20. The antennas transmit independent orthogonal quadrature phase shift keyed (QPSK) waveforms and the carrier frequency is 8.62 GHz . The SNR of the measurement noise is set to a fixed value $(20 \mathrm{~dB})$. The range of the DOA plane is [ $0^{\circ}, 90^{\circ}$ ], which is divided into 30 cells with the initial angle $\left(\theta_{0}\right)$ and angle interval $(\Delta \theta)$ equaling $0^{\circ}$ and $3^{\circ}$, respectively. A maximum of $L=512$ snapshots are considered at the receive node. Targets may appear at directions represented by DOA angles.

Table 1: Performance comparison.

|  | Average reconstruction <br> error | Average RMSE of DOA <br> angles (degree) |
| :--- | :---: | :---: |
| SMP | 0.04 | 0.05 |
| OMP | 0.2 | 0.19 |
| ROMP | 0.24 | 0.23 |
| StOMP | 0.23 | 0.25 |
| SP/CoSaMP | 0.15 | 0.13 |
| BAOMP | 0.17 | 0.16 |
| BPDN | 0.09 | 0.07 |
| Capon | 0.38 | 1.1 |
| APES | 0.35 | 0.54 |
| GLRT | 0.27 | 0.17 |

The proposed SMP method is compared to several popular greedy algorithms, for example, the OMP, ROMP, StOMP, SP, CoSaMP and BAOMP algorithms, and the convex optimization algorithm, BPDN method. The proposed method is further compared to three commonly used methods in DOA estimation, for example, the Capon, APES, and GLRT methods. The compressed sensing based methods (the SMP method, the greedy algorithms, and the BPDN method) use one snapshot only, and the Capon, APES, and GLRT methods use 512 snapshots each. Five hundred Monte Carlo simulations are carried out, and in each trial four targets locate randomly within the DOA range of $\left[0^{\circ}, 90^{\circ}\right]$, and the corresponding reflection coefficients are set as $\left\{\beta_{k}=1, k=\right.$ $1, \ldots, 4\}$.

The average reconstruction error is adopted to evaluate the reconstruction performance of the methods, which is defined as

$$
\begin{equation*}
\chi_{\text {average }}=\sum_{i=1}^{N_{\mathrm{MC}}} \frac{\chi^{i}}{N_{\mathrm{MC}}} \tag{27}
\end{equation*}
$$

where $\chi^{i}$ denotes the reconstruction error at the $i$ th Monte Carlo simulation, which is defined as

$$
\begin{equation*}
x^{i}=\frac{\left\|x_{\text {estimate }}^{i}-x^{i}\right\|_{2}}{\left\|x^{i}\right\|_{2}} \tag{28}
\end{equation*}
$$

where $x^{i}$ and $x_{\text {estimate }}^{i}$ represent the true and estimated signal representing the sparse radar scene at the $i$ th Monte Carlo simulation, respectively. The average root mean square error (RMSE) is also adopted to evaluate the DOA estimation performance of the methods, which is defined in [20].

The results in Table 1 show that the proposed SMP method is with the smallest average reconstruction error and average RMSE. The proposed method succeeds in identifying multiple targets in a sparse radar scene, outperforming other OMP-type methods. It also obtains more precise estimation of DOA angle using one snapshot compared with the traditional estimation methods such as Capon, APES, and GLRT based on 512 snapshots.

## 4. Conclusion

We have presented a novel SMP method for sparse signal reconstruction in compressed sensing. The proposed method expands the searching area and increases the probability of finding the true components at one iteration. It also does not require the sparsity level $K$ to be known in prior. The proposed method is then extended to more practical condition, for example, the direction of arrival (DOA) estimation problem in phased-array radar system using compressed sensing. Numerical simulations show that the proposed method succeeds in identifying multiple targets in a sparse radar scene, outperforming other OMP-type methods. The proposed method also obtains more precise estimation of DOA angle using one snapshot compared with the traditional estimation methods such as Capon, APES, and GLRT based on hundreds of snapshots.

## Acknowledgments

This work was supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (61221063), the Natural Science Foundations of China (No. 61104051, No. 61074176, and No. 61004087), the Fundamental Research Funds for the Central Universities, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33-61, 1998.
[2] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.
[3] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," Proceedings of the IEEE, vol. 98, no. 6, pp. 948-958, 2010.
[4] S. G. Mallat and Z. Zhang, "Matching pursuits with timefrequency dictionaries," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3397-3415, 1993.
[5] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," IEEE Transactions on Information Theory, vol. 50, no. 10, pp. 2231-2242, 2004.
[6] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," Foundations of Computational Mathematics, vol. 9, no. 3, pp. 317-334, 2009.
[7] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," IEEE Transactions on Information Theory, vol. 55, no. 5, pp. 2230-2249, 2009.
[8] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 301-321, 2009.
[9] H. Huang and A. Makur, "Backtracking-based matching pursuit method for sparse signal reconstruction," IEEE Signal Processing Letters, vol. 18, no. 7, pp. 391-394, 2011.
[10] J. Capon, "High resolution frequency-wave number spectrum analysis," Proceedings of the IEEE, vol. 57, no. 8, pp. 1408-1418, 1969.
[11] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," IEEE Transactions on Signal Processing, vol. 44, no. 6, pp. 1469-1484, 1996.
[12] E. J. Kelly, "An adaptive detection algorithm," IEEE Transactions on Aerospace and Electronic Systems, vol. 22, no. 2, pp. 115-127, 1986.
[13] SparseLab, http://dsp.rice.edu/cs.
[14] A. C. Gürbüz, J. H. McClellan, and V. Cevher, "A compressive beamforming method," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08), pp. 2617-2620, Las Vegas, Nev, USA, April 2008.
[15] I. Stojanovic, W. C. Karl, and M. Cetin, "Compressed sensing of mono-static and multi-static SAR," in Algorithms for Synthetic Aperture Radar Imagery XVI, Proceedings of SPIE, Orlando, Fla, USA, April 2009.
[16] J. H. G. Ender, "On compressive sensing applied to radar," Signal Processing, vol. 90, no. 5, pp. 1402-1414, 2010.
[17] Y. Yu, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing-based MIMO radar," IEEE Transactions on Signal Processing, vol. 59, no. 11, pp. 5338-5352, 2011.
[18] J. Zhang, D. Zhu, and G. Zhang, "Adaptive compressed sensing radar oriented toward cognitive detection in dynamic sparse target scene," IEEE Transactions on Signal Processing, vol. 60, no. 4, pp. 1718-1729, 2012.
[19] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," IEEE Transactions on Signal Processing, vol. 59, no. 5, pp. 2182-2195, 2011.
[20] Y. Bar-Shalom and T. E. Fortmann, Tracking and Data Association, Academic Press, Orlando, Fla, USA, 1988.


[^0]:    - Error $=y_{s 2}(k T)-y_{\text {okid } 2}(k T)$

[^1]:    2.2. Detailed Procedures of Splitting Matching Pursuit Method. The detailed procedure of the SMP method is listed as follows.

