Mathematical Problems in Engineering

Theory, Methods, and Applications

Editor-in-Chief: José Manoel Balthazar

Special Issue
Dynamics and Control in Sciences and Engineering

Guest Editor
José Manoel Balthazar

Hindawi Publishing Corporation
http://www.hindawi.com

Volume 2007

Mathematical Problems in Engineering

F. E. Udwadia
University of Southern California, USA
fudwadia@usc.edu

Ferdinand Verhulst
University of Utrecht, The Netherlands
f.verhulst@math.uu.nl

Jerzy Warminski
Lublin University of Technology, Poland
j.warminski@pollub.pl

Mohammad I. Younis
Binghamton University, USA
myounis@binghamton.edu

T. Zolezzi
University of Genova, Italy
zolezzi@dima.unige.it

Mathematical Problems in Engineering
Volume 2007

Special Issue

Dynamics and Control in Sciences and Engineering
Guest Editor: José Manoel Balthazar

Contents

Dynamics and Control in Sciences and Engineering, *J. M. Balthazar*
Volume 2007, Article ID 10432, 3 pages

Successive Bifurcation Conditions of a Lorenz-Type Equation for the Fluid Convection Due to the Transient Thermal Field, *Xiaoling He*
Volume 2007, Article ID 81514, 24 pages

Patrol Mobile Robots and Chaotic Trajectories, *Luiz S. Martins-Filho and Elbert E. N. Macau*
Volume 2007, Article ID 61543, 13 pages

Inductorless Chua's Circuit: Experimental Time Series Analysis, *R. M. Rubinger, A. W. M. Nascimento, L. F. Mello, C. P. L. Rubinger, N. Manzanares Filho, and H. A. Albuquerque*
Volume 2007, Article ID 83893, 16 pages

Chaos Synchronization Criteria and Costs of Sinusoidally Coupled Horizontal Platform Systems, *Jianping Cai, Xiaofeng Wu, and Shuhui Chen*
Volume 2007, Article ID 86852, 10 pages

Stabilizability and Motion Tracking Conditions for Mechanical Nonholonomic Control Systems, *Elżbieta Jarzębowska*
Volume 2007, Article ID 31267, 20 pages

Stabilization and Observability of a Rotating Timoshenko Beam Model, *Alexander Zuyev and Oliver Sawodny*
Volume 2007, Article ID 57238, 19 pages

Numerical and Analytical Study of Optimal Low-Thrust Limited-Power Transfers between Close Circular Coplanar Orbits, *Sandro da Silva Fernandes and Wander Almodovar Golfetto*
Volume 2007, Article ID 59372, 23 pages

Fault Detection and Control of Process Systems, *Vu Trieu Minh, Nitin Afzulpurkar, and W. M. Wan Muhamad*
Volume 2007, Article ID 80321, 20 pages

Joint Dynamics Modeling and Parameter Identification for Space Robot Applications, *Adenilson R. da Silva, Luiz C. Gadelha de Souza, and Bernd Schäfer*
Volume 2007, Article ID 12361, 19 pages

Quadratic Stabilization of LPV System by an LTI Controller Based on ILMI Algorithm, *Wei Xie*
Volume 2007, Article ID 28262, 9 pages

Modal Formulation of Segmented Euler-Bernoulli Beams, *Rosemaira Dalcin Copetti, Julio C. R. Claeyssen, and Teresa Tsukazan*
Volume 2007, Article ID 36261, 18 pages

Asymptotic Solution of the Theory of Shells Boundary Value Problem, *I. V. Andrianov and J. Awrejcewicz*
Volume 2007, Article ID 82348, 25 pages

Dynamic Stationary Response of Reinforced Plates by the Boundary Element Method, *Luiz Carlos Facundo Sanches, Euclides Mesquita, Renato Pavanello, and Leandro Palermo Jr.*
Volume 2007, Article ID 62157, 17 pages

Simple Orbit Determination Using GPS Based on a Least-Squares Algorithm Employing Sequential Givens Rotations, *Rodolpho Vilhena de Moraes, Aurea Aparecida da Silva, and Helio Koiti Kuga*
Volume 2007, Article ID 49781, 8 pages

Evaluation of Tropospheric and Ionospheric Effects on the Geographic Localization of Data Collection Platforms, *C. C. Celestino, C. T. Sousa, W. Yamaguti, and H. K. Kuga*
Volume 2007, Article ID 32514, 11 pages

Incompressible Turbulent Flow Simulation Using the $k - \varepsilon$ Model and Upwind Schemes, *V. G. Ferreira, A. C. Brandi, F. A. Kurokawa, P. Seleghim Jr., A. Castelo, and J. A. Cuminato*
Volume 2007, Article ID 12741, 26 pages

Dynamical Simulation and Statistical Analysis of Velocity Fluctuations of a Turbulent Flow behind a Cube, *T. F. Oliveira, R. B. Miserda, and F. R. Cunha*
Volume 2007, Article ID 24627, 28 pages

A Stochastic Model for the HIV/AIDS Dynamic Evolution, *Giuseppe Di Biase, Guglielmo D'Amico, Arturo Di Girolamo, Jacques Janssen, Stefano Iacobelli, Nicola Tinari, and Raimondo Manca*
Volume 2007, Article ID 65636, 14 pages

Models for Master-Slave Clock Distribution Networks with Third-Order Phase-Locked Loops, *José Roberto Castilho Piqueira and Marcela de Carvalho Freschi*
Volume 2007, Article ID 18609, 17 pages

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 10432, 3 pages
doi:10.1155/2007/10432

Editorial
Dynamics and Control in Sciences and Engineering

J. M. Balthazar

Received 19 August 2007; Accepted 19 August 2007

The aim of this special issue is to present discussions of some recent developments in the area of dynamics and control, in all branches of science and engineering.

The subject of dynamics and control systems is wonderfully broad and it has important applications in fields ranging from several engineering branches, physics, and computer science to the life sciences, sociology, and finance.

So, this special issue of MPE is designed to present the state-of-the-art research and the latest theoretical, numerical, and practical achievements to contribute to the advancement of this field, in a significant way.

This special issue involves 19 original papers, selected by the editor, related to the various researches themes on dynamics and control, in order to present recent results on the mentioned fields.

These papers are related to various subjects: mechanical (modeling, dynamics, robotics, structures, chaos); electrical (telecommunications), aerospace science, biological (modeling), fluids and control processes (optimization).

This special issue on Dynamics and Control in Sciences and Engineering papers is organized as follows:

Concerning chaos, the 4 papers are:

Successive Bifurcation Conditions of a Lorenz-Type Equation for the Fluid Convection Due to the Transient Thermal Field, Xiaoling He.

Patrol Mobile Robots and Chaotic Trajectories, Luiz S. Martins-Filho and Elbert E. N. Macau.

Inductorless Chua's Circuit: Experimental Time Series Analysis, R. M. Rubinger, A. W. M. Nascimento, L. F. Mello, C. P. L. Rubinger, N. Manzanares Filho, and H. A. Albuquerque.

Chaos Synchronization Criteria and Costs of Sinusoidally Coupled Horizontal Platform Systems, Jianping Cai, Xiaofeng Wu, and Shuhui Chen.

Concerning control and identification, the 6 papers are:

Stabilizability and Motion Tracking Conditions for Mechanical Nonholonomic Control Systems, Elżbieta Jarzębowska.

Stabilization and Observability of a Rotating Timoshenko Beam Model, Alexander Zuyev and Oliver Sawodny.

Numerical and Analytical Study of Optimal Low-Thrust Limited-Power Transfers between Close Circular Coplanar Orbits, Sandro da Silva Fernandes and Wander Almodovar Golfetto.

Fault Detection and Control of Process Systems, Vu Trieu Minh, Nitin Afzulpurkar, and W. M. Wan Muhamad.

Joint Dynamics Modeling and Parameter Identification for Space Robot Applications, Adenilson R. da Silva, Luiz C. Gadelha de Souza, and Bernd Schäfer.

Quadratic Stabilization of LPV System by an LTI Controller Based on ILMI Algorithm, Wei Xie.

Concerning dynamics, the 3 papers are:

Modal Formulation of Segmented Euler-Bernoulli Beams, Rosemaira Dalcin Copetti, Julio C. R. Claeyssen, and Teresa Tsukazan.

Asymptotic Solution of the Theory of Shells Boundary Value Problem, I. V. Andrianov and J. Awrejcewicz.

Dynamic Stationary Response of Reinforced Plates by the Boundary Element Method, Luiz Carlos Facundo Sanches, Euclides Mesquita, Renato Pavanello, and Leandro Palermo Jr.

Concerning dynamics and control applications, the 2 papers are:

Simple Orbit Determination Using GPS Based on a Least-Squares Algorithm Employing Sequential Givens Rotations, Rodolpho Vilhena de Moraes, Aurea Aparecida da Silva, and Helio Koiti Kuga.

Evaluation of Tropospheric and Ionospheric Effects on the Geographic Localization of Data Collection Platforms, C. C. Celestino, C. T. Sousa, W. Yamaguti, and H. K. Kuga.

Concerning turbulence, the 2 papers are:

Incompressible Turbulent Flow Simulation Using the κ-ε Model and Upwind Schemes, V. G. Ferreira, A. C. Brandi, F. A. Kurokawa, P. Seleghim Jr., A. Castelo, and J. A. Cuminato.

Dynamical Simulation and Statistical Analysis of Velocity Fluctuations of a Turbulent Flow behind a Cube, T. F. Oliveira, R. B. Miserda, and F. R. Cunha.

Concerning Biological applications, the paper is:

A Stochastic Model for the HIV/AIDS Dynamic Evolution, Giuseppe Di Biase, Guglielmo D'Amico, Arturo Di Girolamo, Jacques Janssen, Stefano Iacobelli, Nicola Tinari, and Raimondo Manca.

Concerning telecommunications, the paper is:

Models for Master-Slave Clock Distribution Networks with Third- Order Phase-Locked Loops, José Roberto Castilho Piqueira and Marcela de Carvalho Freschi.

J. M. Balthazar: Department of Statistics, Applied Mathematics and Computation,
State University of São Paulo at Rio Claro, P.O. Box 178, 13500-230 Rio Claro, SP, Brazil
Email address: jmbaltha@rc.unesp.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 81514, 24 pages
doi:10.1155/2007/81514

Research Article

Successive Bifurcation Conditions of a Lorenz-Type Equation for the Fluid Convection Due to the Transient Thermal Field

Xiaoling He

Received 13 October 2006; Revised 13 February 2007; Accepted 7 June 2007

Recommended by José Manoel Balthazar

This paper investigates the convection flow between the two parallel plates in a fluid cell subject to the transient thermal field. We use the modal approximations similar to that of the original Lorenz model to obtain a generalized Lorenz-type model for the flow induced by the transient thermal field at the bottom plate. This study examines the convection flow bifurcation conditions in relation to the transient temperature variations and the flow properties. We formulated successive bifurcation conditions and illustrated the various flow behaviors and their steady-state attractors affected by the thermal field functions and fluid properties.

1. Introduction

The study of the thermally induced convection flow, or the Rayleigh-Benard convection problem, has centered on the Lorenz equation since 1963 when Lorenz used the 3-mode truncation of the Fourier series to obtain a nonlinear model [1]. Lorenz used the Boussinesq approximation adopted by Saltzman [2] who solved the convection flow problem in a seven-mode Fourier series approximation. The Lorenz equation represents the Rayleigh-Benard convection for both the parallel and circular plates [3, 4]. Essentially, the Boussinesq approximation originates from the Navier-Stokes equation and the heat conduction equation when the variation of the fluid density is negligible. The Lorenz model concerns the thermally induced convection flow by a steady-state temperature difference between the two parallel plates. Lorenz's simplification to the nonlinear equation allows for identification of the convection flow characteristics, such as the strange attractors and flow stabilities. Major investigations of the Lorenz system have been on the bifurcation, stability, and chaos at different Rayleigh numbers and at both the small and large Prandtl numbers [5–8]. These earlier studies are largely based on numerical computations or experimental observations, which demonstrate various behaviors, including the sequential bifurcations with respect to the Rayleigh numbers and chaos with sensitive dependence on the initial conditions. In addition, both the homoclinic and heteroclinic bifurcations occur leading to periodic orbits [9–11]. The study by McLaughlin found that the sequential bifurcation of the Lorenz system itself can give rise to chaos [12, 13]. Curry observed that chaos also persists when the system is subject to a harmonic forcing [14]. However, a formulation to explain the sequential bifurcations has not been well established yet.

A sustained interest in the nonlinear convection flow extends the nonlinear model further to higher order systems than the Lorenz three-dimensional model. Curry subsequently expanded the Lorenz model to a generalized Lorenz model of 14 modes. Curry found different bifurcation and stability conditions with this high-dimensional analogue of the convection flow problem [15]. Specifically, Curry's computation results indicate that a torus of a periodic orbit appeared at a higher r with period doubling bifurcations, where r is the ratio between the Rayleigh number and the critical Rayleigh number. Curry showed that the stability of the origin gives its way to the Hopf bifurcation at a critical Rayleigh number. This critical number r differs from that established from the original Lorenz model. In a separate study, Boldrighini and Franceschini [16] and Franceschini and Tebaldi [17] investigated a five-term truncation of the convection equations. They found that the system has a four-fold symmetry with four stable points and undergoes both Hopf bifurcation and the period doubling bifurcation at large Rayleigh numbers to produce four stable periodic orbits. Further, saddle node bifurcations exist at a larger r. Gibbon and McGuinness studied another variation of the five-mode truncation of the fluid convection model [18]. Their stability condition renders the Hopf bifurcation at $r = 1$ and bifurcations into nonstable torus at a high r, which is consistent with Curry's results. In general, the numerical computation results of the high-dimensional convection flow reveal a different stability and bifurcation condition from that of the original Lorenz model. It is apparent that such a deviation comes from the different modal truncations. For the Fourier series, although a higher order truncation gives a closer approximation of the system, the fundamental mode plays a dominant role compared with other modes. This makes the low-dimensional system, such as the Lorenz model, remain a valid approximation.

In spite of all the attention paid to the Lorenz system, major efforts have focused on a thermal field defined the same as that in the original Rayleigh's description, that is, a constant temperature difference between the two parallel plates is maintained externally [3]. This restriction excluded the transient thermal process in the plate. Therefore, the conclusions drawn from the Lorenz equation or a generalized higher-dimensional Lorenz-type model become invalid when a transient thermal field is present. The nonuniform transient temperature difference arises from a thermal and fluid energy transfer without external thermal modulation. Therefore, a formulation taking into account the nonuniform transient thermal field will better explain the relevant flow behaviors.

In this paper, we investigate the Rayleigh-Benard convection problem with a transient thermal field at the bottom layer. We derive the equation of motion with a transient thermal field using the same truncation modes as that of Lorenz. Our purpose for this study is to see how the transient thermal field influences the flow behavior, such as the bifurcation and chaos with respect to the Rayleigh number and fluid properties. We will answer questions on the sequential bifurcations to convection flow attractors and flow stability in quantified terms to justify the numerical computation results from prior models and from the current model. The study could reveal the difference and analogy between the original Lorenz system and the system with a nonuniform transient thermal field.

This paper is organized as follows. Subsequent to this introduction on the previous study of the original and the generalized Lorenz system, we derive the convection flow model with a nonuniform transient thermal field. Next, we examine the steady-state attractors of the flow subject to different thermal fields. In this part, we formulate various bifurcation conditions, such as the Hopf bifurcation, period doubling, and saddle node bifurcations that affect the attractor behavior and stability. In the fourth section, we illustrate the numerical computation results for the sequential bifurcations and the transient response behavior. We pay special attention to the homoclinic bifurcations at large r and offer our explanation of the phenomena. This paper concludes with discussions and a summary of the influence of a transient thermal field on flow behaviors.

2. The model

The flow within a parallel plates with a transient heat source at the bottom layer is shown in Figure 2.1. The flow is parallel in the y-direction. The flow velocity u, w in the horizontal x- and the vertical z-direction, respectively, are related to the stream function $\psi(x,z,t)$ by the continuity equation as

$$\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}=0, \qquad u=-\frac{\partial \psi}{\partial z}, \qquad w=\frac{\partial \psi}{\partial x}. \tag{2.1}$$

Using the Boussinesq approximation, that is, the variation of the fluid density is negligible, the equilibrium equation for the flow field is [2]

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+w\frac{\partial u}{\partial z}+\frac{\partial P}{\partial x}-\nu\nabla^2 u=0, \tag{2.2a}$$

$$\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial x}+w\frac{\partial w}{\partial z}+\frac{\partial P}{\partial z}-g\varepsilon T_1-\nu\nabla^2 w=0, \tag{2.2b}$$

$$\frac{\partial T}{\partial t}=-u\frac{\partial T}{\partial x}-w\frac{\partial T}{\partial z}+\kappa\nabla^2 T, \tag{2.3}$$

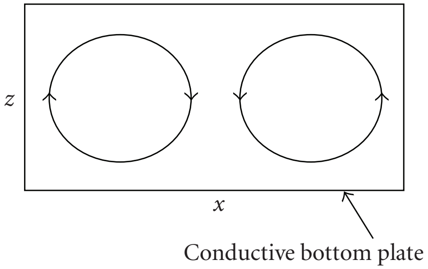


FIGURE 2.1. A flow field in two parallel plates.

where the thermal field is defined as

$$\begin{aligned} T(x,z,t) &= T_1(x,z,t)+\theta(x,z,t) = \left(1-\frac{z}{H}\right)\Delta T+\theta(x,z,t), \\ T_1(x,z,t) &= \Delta T(x,0,t)-\frac{\Delta T(x,0,t)z}{H}, \\ \theta(x,z,t) &= \theta_{11}(x,z,t)+\theta_{02}(z,t). \end{aligned} \tag{2.4}$$

The laminate temperature variation is independent of the cell height, that is,

$$\frac{\partial \Delta T}{\partial z}=0. \tag{2.5}$$

$\theta(x,z,t)$ is the transient temperature variation of the of the flow field, which is composed of the 2D variation $\theta_{11}(x,z,t)$ and the vertical variation $\theta_{02}(z,t)$, respectively. $\Delta T(x,0,t)$ is the temperature difference between the two parallel plates or equivalently the temperature variation of the bottom plate when the upper plate is as the reference. $\Delta T(x,0,t)$ causes a linear temperature variation along the vertical direction, $T_1(x,z,t)$. $\Delta T(x,0,t)$ has both transient and nonuniform spatial variations in the x-direction, that is, $\partial\Delta T(x,0,t)/\partial t \neq 0$, $\partial\Delta T(x,0,t)/\partial x \neq 0$.

Introducing (2.1) into (2.2a) and (2.2b), the governing equation of motion for the thermally induced convection flow is transformed to be

$$\frac{\partial}{\partial t}(\nabla^2\psi)-\frac{\partial\psi}{\partial z}\frac{\partial}{\partial x}(\nabla^2\psi)+\frac{\partial\psi}{\partial x}\frac{\partial}{\partial z}(\nabla^2\psi)-g\varepsilon\frac{\partial T_1}{\partial x}-\nu\nabla^4\psi=0. \tag{2.6}$$

Here, g and ε are the gravitational acceleration and the coefficient of volume expansion of the fluid. With both spatial and temporal variations of the temperature, (2.3) becomes

$$\frac{\partial\theta}{\partial t}+\left[\left(\frac{\kappa}{\kappa_L}-1\right)\frac{z}{H}\right]\frac{\partial\Delta T}{\partial t}=-\frac{\partial\theta}{\partial z}\frac{\partial\psi}{\partial x}+\frac{\partial\theta}{\partial x}\frac{\partial\psi}{\partial z}-\frac{z}{H}\frac{\partial\Delta T}{\partial x}\frac{\partial\psi}{\partial z}+\kappa\nabla^2\theta+\frac{\Delta T}{H}\frac{\partial\psi}{\partial x}. \tag{2.7}$$

The Rayleigh number is a function of the temperature difference between the two parallel plates, that is,

$$R_a = \frac{g\varepsilon H^3 \Delta T(x,z,t)}{\kappa\nu}. \tag{2.8a}$$

The critical Rayleigh number is

$$R_c = \frac{\pi^4}{a^2}\left(1+a^2\right)^3. \tag{2.8b}$$

$R_c = 27\pi^4/4$ when the convection occurs at a wave length of $a^2 = 1/2$ [2]. Considering the transient conductive thermal field in the form $\Delta T(x,t) = \Delta T(x) * g_T(t)$, the ratio between the two Rayleigh numbers r is

$$r(t) = \frac{R_a}{R_c} * g_T(t) = r * g_T(t). \tag{2.8c}$$

The function $g_T(t)$ represents the transient temperature variation with respect to time of a conductive plate. For example, $g_T(t)$ is an exponential function arising from solution of the diffusion equation

$$\frac{\partial \Delta T}{\partial t} = \kappa_L \nabla^2(\Delta T). \tag{2.9}$$

$g_T(t)$ can also assume other forms for different thermal processes in the solids.

By incorporating (2.9) for the heat conduction of the plate along with the heat conductivities at the boundaries, such as the Newmann or Dirichlet boundary conditions as explained below, we obtain a Lorenz-type model with a nonuniform transient thermal field as

$$\begin{aligned} \frac{dX}{d\tau} &= -\sigma X + \sigma Y, \\ \frac{dY}{d\tau} &= cXZ - Y + rX, \\ \frac{dZ}{d\tau} &= drX + XY - bZ + e_\kappa r\beta, \end{aligned} \tag{2.10a}$$

where

$$\begin{gathered} \sigma = \frac{\nu}{\kappa}, \qquad \beta = \frac{g_{T,\tau}}{g_T}, \qquad b = \frac{4}{1+a^2}, \qquad c = 2\cos\left(\frac{2\pi z}{H}\right), \\ e_\kappa = \frac{1}{2}\left(\frac{\kappa}{\kappa_L} - 1\right)\left[1 + \frac{1}{2}\left(\frac{\pi z}{H}\right)^2\right]. \end{gathered} \tag{2.10b}$$

Note that here $g_{T,\tau}$ means $\partial g_T(\tau)/\partial\tau$. We adopted the same truncation modes X, Y, Z as that in the original Lorenz equation, which are dimensionless functions of the normalized time $\tau = (\pi/H)^2(1+a^2)\kappa t$ alone. The parameters b, σ, κ, ν, τ are defined the same as in the original Lorenz equation, that is, the geometry factor b, Prandtl number σ, kinematic thermal viscosity ν, and thermal diffusivity of the fluid κ and that of the solid κ_L. It is worth mentioning that our derivation verified that the original Lorenz model represents the points of $z = H/3$ or $z = 2H/3$ only of the 2-dimensional flow field by assuming that $c = -1$ based on the expression $c = 2\cos(2\pi z/H)$ of this model.

In the above derivation, a series approximation is used for the temperature variation $\Delta T(x)$ with respect to x, in order to be consistent with the form of the functions X, Y, Z for the purpose of reduction. This variation of $\Delta T(x)$ introduces the thermal parameter d, which is related to heat conduction at the boundaries as follows:

(a) the von Neumann condition $\Delta T(x) = T_0 \sin(2\pi x/L)$ satisfies $\Delta T(x = 0,\ x = L) = 0$,

$$\Delta T(x,t) = \Delta T(x) g_T(\tau),$$
$$d = d_N(x) = -2\sqrt{2}\left(\frac{\pi x}{L}\right)^2, \quad d_N = [-2\sqrt{2}\pi^2, 0] = [-27.92, 0]. \tag{2.11a}$$

(b) the Dirichlet condition $\Delta T(x) = T_0 \cos(2\pi x/L)$ satisfies $\partial \Delta T/\partial x(x = 0,\ x = L) = 0$,

$$\Delta T(x,t) = \Delta T(x) g_T(\tau),$$
$$d = d_D(x) = -\frac{\sqrt{2}\pi x}{L}, \quad d_D = [-\sqrt{2}\pi, 0] = [-4.44, 0]. \tag{2.11b}$$

At the center of the plate $x = L/2$, $d_N|_{x=L/2} = -6.98$, $d_D|_{x=L/2} = -2.22$, $d_N = d_D|_{x=L/2\pi} = -\sqrt{2}/2$.

By incorporating the nonuniform transient thermal field, we obtain this Lorenz-type model that differs from the original Lorenz equation. The model correlates the convective flow with the transient temperature fluctuation function $\beta(\tau)$ in the conductive plate, the spatial temperature variation and the thermal boundary condition of the plate $d(x)$, and the plate-fluid thermal diffusion rate parameter $e_\kappa(z)$, respectively. The transient thermal field acts as a forcing source in the form $\beta(\tau) = g_{T,\tau}(\tau)/g_T(\tau)$, which measures the rate of change of the temperature or the thermal fluctuation of the plate. $\beta(\tau)$ drives the vertical temperature of the fluid Z directly to influence the flow field stream function X and the temperature variation Y. When $\beta(\tau) = 0$ and $d(x) = 0$, this model reduces to that by Lorenz.

As the definition entails, $e_\kappa(z)$ concerns the thermal diffusion between the fluid and the plate; $e_\kappa(z)$ influences the vertical temperature variation Z due to the heat exchange between the plate and the fluid. Since the thermal diffusivity of the fluid is generally greater than that of the solid, that is, $\kappa > \kappa_L$, therefore, $e_\kappa(z) > 0$. In addition, $e_\kappa(z)$ increases as the fluid-solid heat exchange rate subsides at the high end of the cell. As an example, the glycerin has conductivity in the range of $\kappa = 0.14$ [W/cm K], engine oil has $\kappa = 0.28$ [W/cm K], and a conductive metal plate has $\kappa = 0.2$ [W/cm K]. If $\kappa/\kappa_L = 5$ for the solid layer, $e_\kappa|_{\min} = 2$ at $z = 0$, and $e_\kappa|_{\max} = 11.89$ at $z = H$. At the same points for $c = -1$, that is, $z = H/3$ and $z = 2H/3$, $e_\kappa(z = H/3) = 3.08$ and $e_\kappa(z = 2H/3) = 6.38$, respectively.

This expanded model has the same negative divergence as the original Lorenz system, when the transient function $\beta(\tau)$ is considered as an external forcing, that is,

$$\frac{\partial \dot{X}}{\partial X}+\frac{\partial \dot{Y}}{\partial Y}+\frac{\partial \dot{Z}}{\partial Z}=-(1+b+\sigma)<0, \tag{2.12}$$

which suggests that the flow is dissipative. Geometrically, a dissipative system has all trajectories confined when the transient temperature rise is restricted. On the other hand, a rapid temperature rise certainly will cause oscillation without bound if $\beta(\tau)$ is unbounded.

3. The steady-state attractors and bifurcations

We examine the linearized system for the stability of the flow at the steady state, that is,

$$J=\begin{bmatrix} -\sigma & \sigma & 0 \\ cZ+r & -1 & cX \\ dr+Y & X & -b \end{bmatrix}. \tag{3.1}$$

3.1. The steady state at the origin for $X = Y = Z = 0$**.** The eigenvalues determine the stability and bifurcation behaviors of the system. For a steady-state attractor appearing at $X = Y = Z = 0$, the eigenvalues are given by

$$(\lambda+b)\big[(\lambda+1)(\lambda+\sigma)-\sigma r\big]=0. \tag{3.2}$$

It is evident that the eigenvalues are independent of $d(x)$ and the thermal fluctuation function $\beta(\tau)$. This defines the same eigenvalues as the Lorenz model, that is,

$$\begin{aligned} \lambda_1 &= -b<0, \\ \lambda_{2,,3} &= \frac{1}{2}\Big[-(1+\sigma)\pm\sqrt{(1-\sigma)^2+4\sigma r}\Big] \\ &= \frac{1}{2}\Big[-(1+\sigma)\pm(1-\sigma)\sqrt{1+\delta}\Big], \quad \delta=\frac{4\sigma r}{(1-\sigma)^2}>0. \end{aligned} \tag{3.3}$$

Using the series approximation,

$$\lambda_2=\sigma\left[-1+\frac{r}{(1-\sigma)}\right], \qquad \lambda_3=-\left[1+\frac{\sigma r}{(1-\sigma)}\right]. \tag{3.4}$$

$\lambda_2 < 0$ for $\sigma > 1$or $r < (1-\sigma)$. Note that $\sigma > 1$ is a typical condition for the convection flow problem. $\lambda_3 < 0$ when $r > (1 - 1/\sigma)$. The negative eigenvalues produce a stable flow to the nodal attractor at the origin. The condition for the onset of the convection flow is $r < (1 - 1/\sigma)$. For $\sigma\to\infty$, this means that $r \to 1$.

An unstable saddle node bifurcation occurs at $\lambda_3 = 1$, corresponding to

$$r=2\left(1-\frac{1}{\sigma}\right). \tag{3.5}$$

This condition can be satisfied by numerous combinations of parameters, such as $\sigma = 2$, $r = 1$ and $\sigma = 10$, $r = 1.8$, which suggests that the system experiences a sequential saddle node bifurcation as r varies. In a similar fashion, we can find that the condition for period doubling bifurcation at $\lambda_1 = -b$ is $b = 1$. However, the period doubling bifurcation will not occur at $\lambda_{2,3} = -1$. This is because the physical parameter $r > 0$; $\lambda_2 = -1$ requires $r = -(1-\sigma)^2/\sigma < 0$ and $\lambda_3 = -1$ requires $r = 0$. Since all the eigenvalues are real, the steady state attractor at the origin does not undergo the Hopf bifurcation. However, a successive saddle node and period doubling bifurcations can occur at different r, b and σ.

3.2. The nonzero steady-state attractors. The steady-state attractor at the nonorigin, that is, at $X, Y, Z \neq 0$ is determined by $\dot{X} = \dot{Y} = \dot{Z} = 0$ from (2.10a), which yields

$$\begin{gathered} X = Y, \\ cZ = (1-r), \\ drX + X^2 - bZ + e_\kappa r\beta(\tau) = 0. \end{gathered} \tag{3.6}$$

This defines the attractors at

$$\begin{gathered} X^2 + drX + \eta = 0, \quad \eta = \frac{b}{c}(r-1) + e_\kappa r\beta(\tau) = \frac{1}{c}\left[r\left(b + ce_\kappa\beta(\tau)\right) - b\right], \\ X = Y = \frac{1}{2}\left(-dr \pm \sqrt{(dr)^2 + \frac{4}{c}\left[b - r\left(b + ce_\kappa\beta(\tau)\right)\right]}\right), \\ Z = \frac{1}{c}(1-r). \end{gathered} \tag{3.7}$$

The original Lorenz attractor at $c = -1$, $d = 0$, $e_\kappa = 0$ or $\beta(\tau) = 0$ is at

$$X = Y = \pm\sqrt{b(1-r)}, \qquad Z = r - 1. \tag{3.8}$$

X, Y can be either real or complex, depending on the value of r. To ensure the physical parameter X is a real parameter, the following condition should be satisfied:

$$(r^*)^2 - \frac{4b + 4ce_\kappa\beta(\tau)}{cd^2}r + \frac{4b}{cd^2} > 0. \tag{3.9}$$

For a real attractor X, Y, $r > r_1^*$ or $r < r_2^*$, $r_2^* < r_1^*$. The condition in (3.9) is alternatively expressed as

$$f_0(r^*) = (dr^*)^2 + \frac{4}{c}\left[b - r^*\left(b + ce_\kappa\beta(\tau)\right)\right] > 0. \tag{3.10}$$

The characteristic equation for the stability of the attractor is in the form

$$\lambda^3 + I\lambda^2 + II\lambda + III = 0, \tag{3.11}$$

where

$$\begin{aligned} I &= -(\lambda_1+\lambda_2+\lambda_3) = (1+b+\sigma), \\ II &= (\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3) = b^2+b\sigma-cX^2, \\ III &= -\lambda_1\lambda_2\lambda_3 = \sigma cdrX, \end{aligned} \tag{3.12}$$

where X is determined by (3.7). The characteristics equation can lead to various bifurcation conditions determined by the eigenvalues, as we analyze below.

(a) *Periodic orbits with purely imaginary eigenvalues.* The steady-state attractors will not occur with a pair of purely imaginary eigenvalues. This is because the condition requires that

$$b+1=0. \tag{3.13}$$

As $b>0$, this is impossible. Similarly, it can be demonstrated that neither will a flow initiate due to the real eigenvalues of

$$\lambda_1=-b, \qquad \lambda_2=-\lambda_3. \tag{3.14}$$

(b) *All negative real eigenvalues and complex conjugacy.* For the Hopf bifurcation in a complex conjugacy $\lambda_{2,3}=\alpha\pm i\gamma$ and $\alpha=-1/2(I+\lambda_1)$, the coefficients of the characteristic equation become

$$\begin{aligned} I &= -(\lambda_1+2\alpha) = (1+b+\sigma), \\ II &= (2\alpha\lambda_1+\alpha^2+\gamma^2) = b^2+b\sigma-cX^2, \\ III &= -\lambda_1(\alpha^2+\gamma^2) = \sigma cdrX. \end{aligned} \tag{3.15}$$

For any $\alpha\lambda_1>0$ in either of the Hopf bifurcations, $II=2\alpha\lambda_1+\alpha^2+\gamma^2>0$. This defines the necessary condition associated with three possibilities: (a) all real eigenvalues to make $II>0$; (b) the subcritical Hopf bifurcation when $\alpha<0$ with $\lambda_1<0$; (c) the supercritical Hopf bifurcation when $\alpha>0$ with $\lambda_1>0$. However, $\alpha>0$ leads to $\lambda_1<-I<0$, since $I>0$ and $\alpha=-1/2(I+\lambda_1)$. Therefore, only cases (a) and (b) are possible. In the case of $\lambda_1=0$, $III=0$ and $X=0$. This means that there is no periodic orbit due to either of the Hopf bifurcations alone with $\lambda_1=0$. The condition $II>0$ generates the steady-state attractor at

$$X>-\frac{r(b+ce_\kappa\beta(\tau))-b+b(b+\sigma)}{cdr}=-\frac{(b+ce_\kappa\beta(\tau))}{cd}-\frac{b(b+\sigma-1)}{cdr}=\Delta_{nh}. \tag{3.16a}$$

For the steady-state response, this is equivalent to

$$f_{nh}(r)=\frac{1}{2}\left(-dr\pm\sqrt{(dr)^2+\frac{4}{c}[b-r(b+ce_\kappa\beta(\tau))]}\right)-\Delta_{nh}>0. \tag{3.16b}$$

Note that the condition (3.16) does not differentiate between cases of all real eigenvalues in $II>0$ and the case of complex conjugate eigenvalues with $\alpha<0$ and $\lambda_1<0$. This means that the necessary condition is not exclusive for either of the cases.

(c) *The Hopf bifurcation with complex eigenvalues.* In the case of the Hopf bifurcation with a real eigenvalues, we can further identify the necessary conditions for different cases. The condition can be expressed in the form identical to that in (3.16) except for the operator Δ_{nh}. That is,

$$f_j(r) = \frac{1}{2}\left(-dr \pm \sqrt{(dr)^2 + \frac{4}{c}\left[b - r\left(b + ce_\kappa\beta(\tau)\right)\right]}\right) - \Delta_j. \tag{3.17}$$

Here the function $f_j(r)$ is associated with the specific operator Δ_j, of which three conditions can be drawn based on the relation between λ_1 and α:

(1) $I * II - III < 0$, a necessary condition for the supercritical Hopf bifurcation with $\lambda_1 < -I < 0$, $\alpha > 0$, which requires the steady-state attractor to satisfies

$$X < -\frac{(1+b+\sigma)}{cdr(1+b)}\left[b(b+\sigma-1) + r\left(b + ce_\kappa\beta(\tau)\right)\right] = \Delta_H^-. \tag{3.18a}$$

Therefore, the necessary condition is

$$f_{\text{sup}}^-(r) = \frac{1}{2}\left(-dr \pm \sqrt{(dr)^2 + \frac{4}{c}\left[b - r\left(b + ce_\kappa\beta(\tau)\right)\right]}\right) - \Delta_H^- < 0. \tag{3.18b}$$

(2) $I * II - III > 0$ for the subcritical Hopf bifurcation with $-I < \lambda_1 < 0$ and $-I/2 < \alpha < 0$, which is the opposite condition to that for the supercritical Hopf bifurcation with $\lambda_1 < -I < 0$. Therefore, the necessary condition is

$$f_{\text{sub}}^-(r) = \frac{1}{2}\left(-dr \pm \sqrt{(dr)^2 + \frac{4}{c}\left[b - r\left(b + ce_\kappa\beta(\tau)\right)\right]}\right) - \Delta_H^- > 0. \tag{3.19}$$

(3) $II^2 - 4I * III > 0$ for the subcritical Hopf bifurcation with $\lambda_1 > 0$, $\alpha < -I/2 < 0$, this defines

$$X < \frac{\left\{\left(b^2 + b\sigma - b + \left(ce_\kappa\beta(\tau) + b\right)r\right)^2 - c(dr)^2\left[\left(ce_\kappa\beta(\tau) + b\right)r - b\right]\right\}}{\left\{4(1+b+\sigma)\sigma - 2\left(b^2 + b\sigma - b + \left(ce_\kappa\beta(\tau) + b\right)r\right) + c(dr)^2\right\}cdr} = \Delta_{\text{sub}}^+. \tag{3.20a}$$

This leads to the necessary condition of

$$f_{\text{sub}}^+(r) = \frac{1}{2}\left(-dr \pm \sqrt{(dr)^2 + \frac{4}{c}\left[b - r\left(b + ce_\kappa\beta(\tau)\right)\right]}\right) - \Delta_{\text{sub}}^+ < 0. \tag{3.20b}$$

The above three conditions are exclusive necessary condition for each Hopf bifurcation.

(d) *The Hopf bifurcation concurrent with the saddle node/period doubling bifurcations.* When the real eigenvalue is specified, the necessary and sufficient condition can be uniquely defined for the Hopf bifurcation with a real eigenvalue. For example, the period doubling or saddle node bifurcation can occur at $\lambda_1 = -1$ and $\lambda_1 = 1$, respectively, concurrent with the subcritical Hopf bifurcation. This is because $\alpha < 0$ as long as $\lambda_1 > -(1+b+\sigma)$. Therefore,

$$\alpha_{\lambda_1=1} = \frac{1}{2}[-I-1] = -\frac{1}{2}(2+b+\sigma), \qquad \alpha_{\lambda_1=-1} = \frac{1}{2}[-I+1] = -\frac{1}{2}(b+\sigma). \tag{3.21}$$

Notice that a combination of $\lambda_1 = \pm 1$ and $\alpha = 0$ for the purely imaginary eigenvalues is impossible since $\lambda_1 < -1$ at $\alpha = 0$.The necessary and sufficient condition for the Hopf bifurcation with any real λ_1 is

$$III + \lambda_1 II = -(\lambda_1)^2 (I + \lambda_1), \tag{3.22}$$

which defines the attractor at

$$X = -\frac{\{(\lambda_1)^2(1+\sigma+b+\lambda_1) + \lambda_1[b^2 + b\sigma - b + r(b + ce_\kappa\beta(\tau))]\}}{(\lambda_1+\sigma)cdr} = \Delta_H^*. \tag{3.23a}$$

Equivalently, this gives the bifurcation condition of:

$$f_H^*(r) = \frac{1}{2}\left(-dr \pm \sqrt{(dr)^2 + \frac{4}{c}[b - r(b + ce_\kappa\beta(\tau))]}\right) - \Delta_H^* = 0. \tag{3.23b}$$

The condition in (3.22) ensures that all the necessary conditions for the Hopf bifurcation in (3.18), (3.19), and (3.20) are satisfied. For example, in the case of the supercritical Hopf bifurcation with $\lambda_1 < -I < 0$, the condition $I * II - III < 0$ becomes

$$\begin{aligned} I * II - III &= I * II + \lambda_1^- II + (\lambda_1^-)^2 I + (\lambda_1^-)^3 < I * II - I * II + (\lambda_1^-)^2 I + (\lambda_1^-)^3 \\ &= (\lambda_1^-)^2 [I + (\lambda_1^-)] < 0. \end{aligned} \tag{3.24a}$$

For the subcritical Hopf bifurcation with $-I < \lambda_1 < 0$, $\alpha < 0$, $I * II - III > 0$ is satisfied by

$$\begin{aligned} I * II - III &= I * II + \lambda_1^- II + (\lambda_1^-)^2 I + (\lambda_1^-)^3 > I * II - I * II + (\lambda_1^-)^2 I + (\lambda_1^-)^3 \\ &= (\lambda_1^-)^2 [I + (\lambda_1^-)] > 0. \end{aligned} \tag{3.24b}$$

For the subcritical Hopf bifurcation with $\lambda_1 > 0$, $\alpha < 0$, the condition $II^2 - 4I * III > 0$ is

$$II^2 + 4I * (\lambda_1^- II + (\lambda_1^-)^2 I + (\lambda_1^-)^3) = [II + 2(\lambda_1^- I)]^2 + 4(\lambda_1^-)^3 * I > 0. \tag{3.24c}$$

The condition (3.23) can also determine the concurrent saddle node bifurcation at $\lambda_1 = 1$ and the period doubling bifurcation at $\lambda_1 = -1$, respectively, since, (3.23) is the necessary and sufficient condition for the Hopf bifurcations with a specified eigenvalue. This condition defines the Hopf bifurcation curve, whereas those necessary conditions in (3.18), (3.19), and (3.20) define the domain boundaries for the bifurcation they are associated with. These conditions describe a bifurcation map with respect to the physical parameters. The fact that all of the conditions are in the third order polynomials of r

suggests that several ranges of parameters could coexist to satisfy the condition. As a result, the system exhibits sequential bifurcations discussed above. One exception is the Hopf bifurcation with a pair of purely imaginary eigenvalues, which will not occur due to the restriction of the physical parameters. These bifurcation conditions provide qualified and quantified terms to define the steady-state attractors and describe their stability due to various bifurcations.

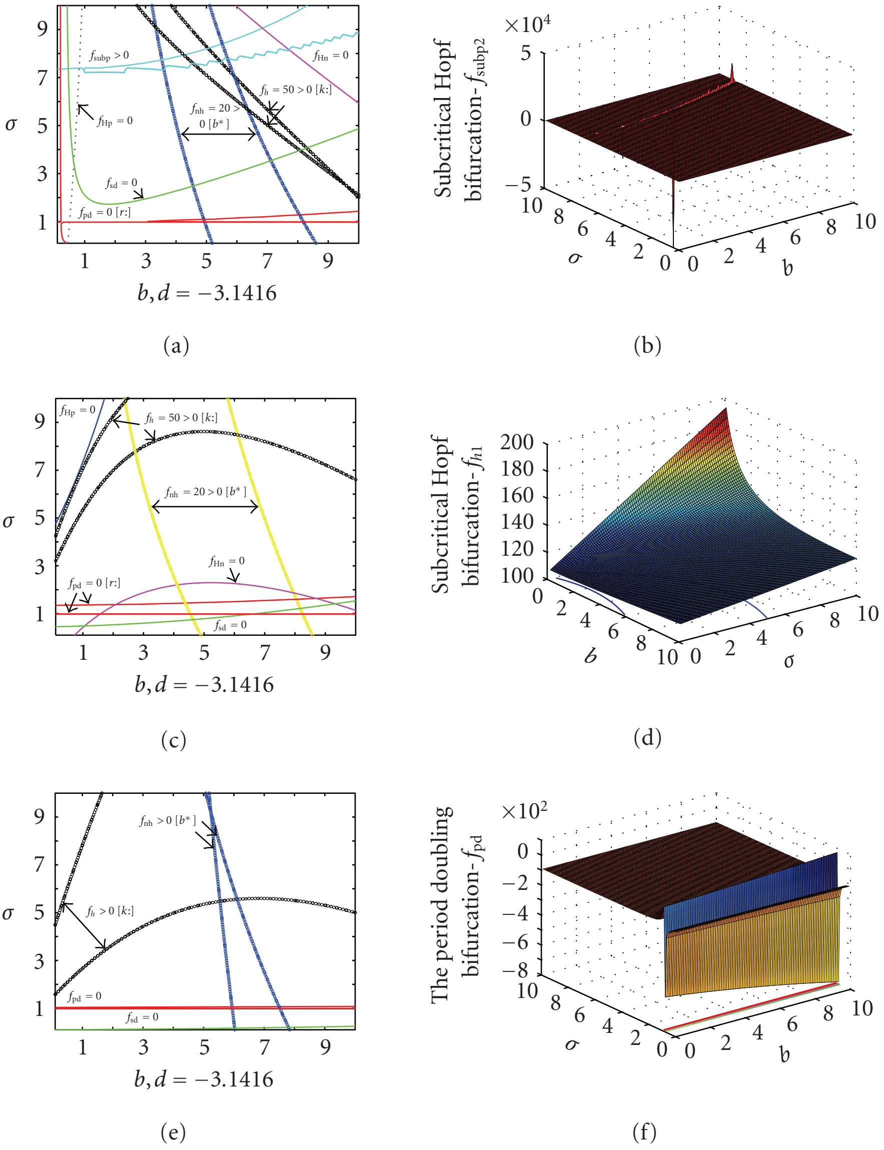
4. Computation results

We use a numerical computation in the 4th-order Runge-Kutta method to study the bifurcation behavior and the response of the system. We fixed parameters at $c = -1$, $d = -\pi$ for the response with respect to σ, b, and r.

4.1. Bifurcation conditions. The computation for the bifurcation map assumes the fluid property parameters b and σ in the range of [0 10]. We study various bifurcation conditions that can be satisfied by the parameters in this range with respect to σ and b when r and $\beta(\tau)$ are specified. The purpose of the computation is to demonstrate the influence of the transient thermal field effect on the convective flow of the specified geometry loci. For the Hopf bifurcations with a specified eigenvalue λ_1, we select the period doubling bifurcation at $\lambda_1 = -1$ and the saddle node bifurcation at $\lambda_1 = 1$, respectively. In addition, the Hopf bifurcation conditions at both $\lambda_1 = 15$ and $\lambda_1 = -15$ are examined. At $\lambda_1 = -15$, either a supercritical or a subcritical Hopf bifurcation can occur since the real part of the complex conjugate α varies between $\alpha = -3$ and $\alpha = 7$ for b, σ in the range of [0 10]. For $\lambda_1 = 15$, only a subcritical Hopf bifurcation is possible as α varies between $\alpha = -8$ and $\alpha = -18$. These four curves are marked by fsd for the Hopf bifurcation with the saddle node, fpd for that with the period doubling, fHp for the subcritical Hopf bifurcation at $\lambda_1 = 15$ and fHn for the Hopf bifurcation at $\lambda_1 = -15$, respectively. Curves $f\,\mathrm{sub}\,p > 0$, $fh > 0$ and $fnh > 0$ represent the contour projections of the function $f_{\mathrm{sub}}^{+}(r) > 0$ in (3.20b), $f_{\mathrm{sub}}^{-}(r) > 0$ in (3.19), and $f_{nh}(r) > 0$ in (3.16b), respectively, for the three distinct necessary conditions associated with the Hopf bifurcations and other possible cases. Therefore, curves fsd, fpd, and fHn and fHp define the necessary and sufficient conditions while each other one represents the necessary condition only.

Figures 4.1(a), 4.1(c), and 4.1(e) show the bifurcation curves at a specified r and $\beta(\tau)$. Figure 4.1(a) is for $r = 5$ and $\beta(\tau) = 0$ whereas Figure 4.1(c) shows the bifurcation at $r = 5$ and $\beta(\tau) = -10$. In both figures, we observe the curve with the saddle node bifurcation, fsd, the curve with the period doubling fpd and the curve fHn for the Hopf bifurcation at $\lambda_1 = -15$. The curve$f\,\mathrm{sub}\,p > 0$ in Figure 4.1(a) defines parameter range that satisfies the condition for a successive subcritical Hopf bifurcation with $\lambda_1 > 0$. The 3D plot for the condition $f\,\mathrm{sub}\,p$ is shown in Figure 4.1(b), indicating that the intercepted area by the two blue curves satisfy $f\,\mathrm{sub}\,p > 0$. Curve fHp also appears in both Figures 4.1(a) and 4.1(c) for the subcritical Hopf bifurcation with $\lambda_1 = 15$, with different parameter ranges.

Note that there are two curves of $fh > 0$ for the subcritical Hopf bifurcation with $\lambda_1 < 0$, that is, $f_{\mathrm{sub}}^{-}(r) > 0$, in Figures 4.1(a) to 4.1(c), as a result of the projection at a specified value at $fh = 50$, similar to the curve $f\,\mathrm{sub}\,p > 0$ in Figure 4.1(a). For the case in Figure 4.1(c) with $r = 5$ and $\beta = -10$, the necessary condition $f_{\mathrm{sub}}^{-}(r) > 0$ in (3.17),



(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 4.1. Bifurcation conditions and map. (a) Bifurcation map at $r = 5$, $\beta = 0$. (b) The subcritical Hopf bifurcation condition $f^{+}_{\text{sub}}(r) > 0$ at $r = 5$, $\beta = 0$. (c) Bifurcation map at $r = 5$. (d) The subcritical Hopf bifurcation condition with $\lambda_1 < 0$ $f^{-}_{\text{sub}}(r) > 0$ at $r = 28$. (e) Bifurcation map at $r = 28$. (f) Condition for the period doubling bifurcation at $f_{PD}(r)$ at $r = 28$.

shown by $fh = 50 > 0$, is uniformly satisfied by the parameters in the 2D domain as shown in Figure 4.1(d) for $f^{-}_{\text{sub}}(r) > 0$. The curves in Figures 4.1(a) and 4.1(c) illustrate only a selective projection of the contour at $fh = 50$. In both Figures 4.1(a) and 4.1(c), the necessary condition for $f_{nh}(r) > 0$ is also satisfied, marked by the curve $fnh > 0$ where the projection is at $f_{nh}(r) = 20$ in Figure 4.1(a) and $f_{nh}(r) = 50 > 0$ in Figure 4.1(c). Both conditions $fh > 0$ for $f_{\text{sub}}^{-}(r) > 0$ and $fnh > 0$ for $f_{nh}(r) > 0$ are valid in the entire domain of b and σ in Figures 4.1(a) and 4.1(c).

The variation of $\beta(\tau)$ changes the bifurcation conditions as it is evident from comparison of Figures 4.1(a) and 4.1(c) where the same condition is satisfied by different parameters. In the same token, a different r also alters the bifurcation map. Figure 4.1(e) is the map for $r = 28$ and $\beta = -10$, where the Hopf bifurcation curve with the period doubling and the saddle node bifurcations are illustrated. The curves $fh > 0$ and $fnh > 0$ are again selective contour projections of functions that are satisfied by the parameters b, σ in the range of [0 10]. Figure 4.1(f) shows that the Hopf bifurcation concurrent with the period doubling bifurcation occurs around $\sigma = 1$, which is characteristic for all different range of parameters as seen from Figures 4.1(a), 4.1(c), and 4.1(e) also.

4.2. Transient thermal field functions.

We study three different transient thermal field functions, which are as follows.

(a) *The harmonic function $g_T(\tau)$.* A cyclic function of $\beta(\tau)$ results. Specifically,

$$g_T(\tau) = \cos(\omega\tau), \qquad \beta(\tau) = -\omega\tan(\omega\tau)\begin{cases} < 0, & 0 < (\omega\tau) < \pi/2, \\ > 0, & \pi/2 < (\omega\tau) < \pi. \end{cases} \tag{4.1a}$$

Note that the sign of ω does not affect the sign of $\beta(\tau) = -\omega\tan(\omega\tau)$ in each bounded interval $k\pi < (\omega\tau) < (k+1)\pi + \pi/2$.The function $\beta(\tau)$ goes to infinity at the boundaries $(\omega\tau) = (k+1)\pi \pm \pi/2$. The function $\beta(\tau)$ causes instantaneous change of the attractors due to transitions of the bifurcation conditions, which makes the condition $\dot{X} = \dot{Y} = \dot{Z} = 0$ invalid. Equivalently, this suggests that a steady-state attractor at $\dot{X} = \dot{Y} = \dot{Z} = 0$ does not exist in this case.

(b) *The exponential function $g_T(\tau)$ for the thermal conduction.* The exponential function $g_T(\tau)$ produces a constant driving force as

$$g_T(\tau) = \exp(-\omega\tau), \qquad \beta(\tau) = -\omega. \tag{4.1b}$$

The function $\beta(\tau) > 0$, if $\omega < 0$ for a temperature rise or vice versa. Therefore, a temperature rise or decline for $\beta(\tau) > 0$ or $\beta(\tau) < 0$ will influence the steady state response in a different fashion. However, in each case the steady-state attractor, as defined in (3.7), remains stationary since $\beta(\tau)$ is a constant.

(c) *The linear function $g_T(\tau)$.* The linear function $g_T(\tau)$ makes $\beta(\tau) > 0$ and $\beta(\tau) \to 0$ as $\tau \to \infty$, that is,

$$g_T(\tau) = \omega\tau, \qquad \beta(\tau) = \frac{1}{\tau} > 0. \tag{4.1c}$$

The three transient thermal field functions $\beta(\tau)$ discussed above suggest that the steady state is not stationary with a harmonic function $g_T(\tau)$, while the exponential and linear functions lead to stationary attractors.

4.3. Transient responses. We use a time step $h = 0.001$ second to generate the transient response for steps of $N = 10{,}000$ for this system with respect to different $\beta(\tau)$ and other parameters. The initial condition is assumed to be $X = 0$, $Y = 20$, $Z = 10$, unless otherwise specified.

(a) $r = 1$, $\beta(\tau) = 0$, *nodal attractors at the origin.* For $r = 1$ and $\beta(\tau) = 0$ without the transient thermal effect, the response could converge to the steady-state $X = Y = 0$, $Z = r - 1$ or $X = Y = -d$, $Z = r - 1$, depending on the initial conditions. Figures 4.2(a) and 4.2(b) show that all X, Y, and Z converge to the zero steady-state attractor at $r = 1$, $b = 8/3$, $\sigma = 6.33$ with the given initial conditions. The steady state reaches the nodal point $X = 0$, $Y = 0$, $Z = 0$ quickly after a transient oscillation. The corresponding eigenvalues are all real and negative, that is, $\lambda = 0, -10.4273, -3.2394$, which makes the nodal attractor stable.

(b) $r = 5$, $\beta(\tau) = 0$, *the Hopf bifurcation.* The parameters $b = 8/3$, $\sigma = 10$, $r = 5$ induce the subcritical Hopf bifurcation with eigenvalues of $\lambda_1 = 11.5572$ and $\lambda_{2,3} = -12.6119 \pm 20.8298i$. Figures 4.2(c) and 4.2(d) show a spiral oscillation leading to the nonzero steady-state attractor. A similar behavior exists at $r = 28$, shown in Figures 4.2(e) and 4.2(f). A higher oscillation frequency during the transition to the steady-state is evident at $r = 28$ in comparison with that at $r = 5$, as shown in the phase diagrams of X-Y and Y-Z for $r = 5$ and $r = 28$, respectively. This is caused by the eigenvalue with an increased imaginary part at $r = 28$, which is $\lambda_1 = 23.6679$ and $\lambda_{2,3} = -18.7065 \pm 91.5255i$.

(c) $r = 28$ *with different transient thermal field functions.* The response behaves differently with a different transient thermal field function $\beta(\tau)$. Figures 4.3(a) and 4.3(b) show the transient response with $b = 8/3$, $\sigma = 10$, $r = 28$ and $\beta(\tau) = 1/\tau$. The transient response vanishes after certain iterations with the ensuing oscillation approaching the steady state by way of the Hopf bifurcation as $\beta(\tau) = \lim_{\tau\to\infty}(1/\tau) = 0$. The steady state attractor is identical to that shown in Figures 4.2(e) and 4.2(f), respectively.

The response with an exponential function $g(\tau) = \exp^{-(\pi/2)\tau}$ and $\beta(\tau) = (\pi/2)$ is shown in Figures 4.3(c) and 4.3(d) with $b = 8/3$, $\sigma = 10$, and $r = 28$. The transient exponential function, in fact, produces a constant driving force to the system. This function $\beta(\tau)$ directly influences the vertical temperature $Z(\tau)$ and modifies the attractor X, Y, and Z. The steady-state attractor is at $X = Y = 88.64$, $Z = 27$ for $\beta(\tau) = (\pi/2)$ shown in Figures 4.3(c) and 4.3(d) for the phase diagrams of X-Y and Y-Z, respectively. Note that the subcritical Hopf bifurcation produces a spiral for either $\beta(\tau) = (\pi/2)$ or $\beta(\tau) = 1/\tau$ approaching a stationary steady-state attractor. Notably, the attractor position Z is independent of $\beta(\tau)$, although $\beta(\tau)$ influences the transient behavior of Z prior to the steady state. Each initial response goes through a period of increasing amplitude for different r and different function $\beta(\tau)$. This is caused by the positive real eignevalue λ_1. Such increase is eventually balanced by the negative real part of the complex eigenvalues, which eventually makes the periodic oscillation dominate in a subcritical Hopf bifurcation.

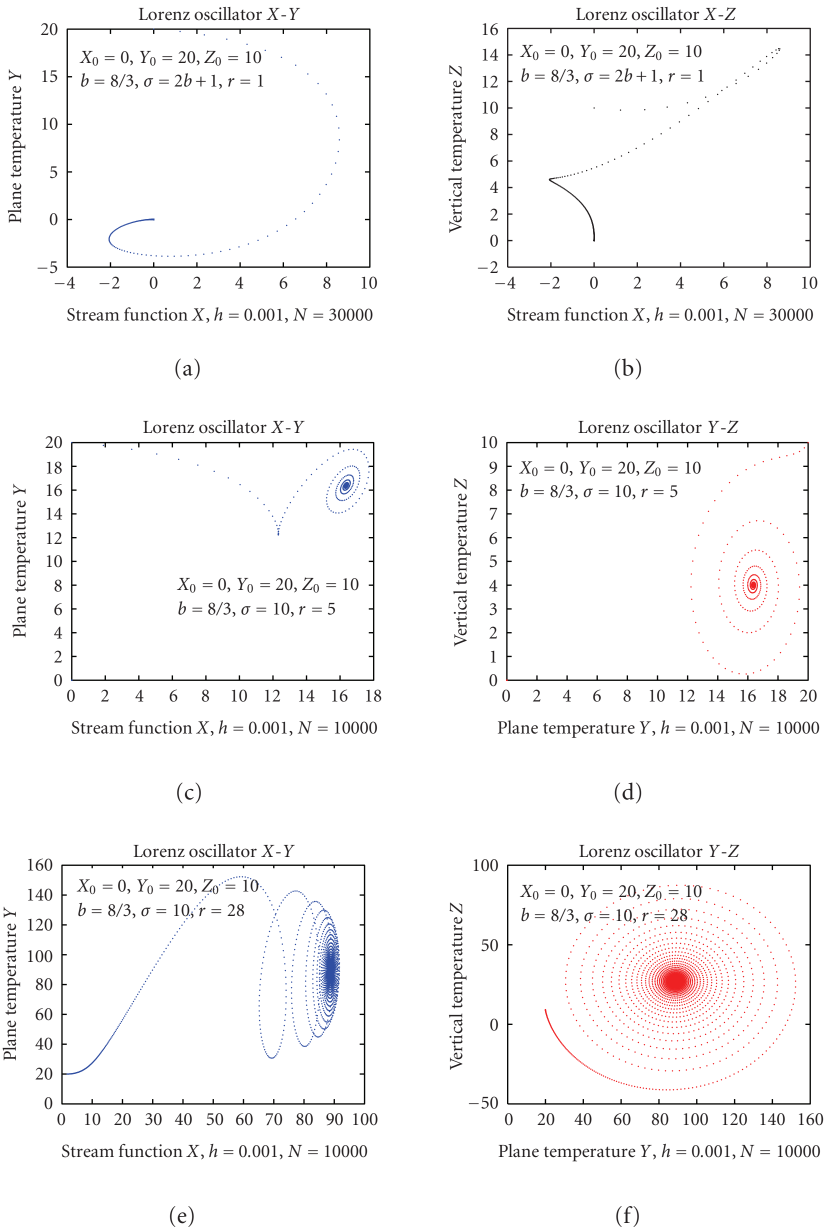


FIGURE 4.2. Phase diagrams without the transient effect. (a) X-Y at $b = 8/3$, $\sigma = 2b + 1$, $r = 1$. (b) X-Z at $b = 8/3$, $\sigma = 2b + 1$, $r = 1$. (c) X-Y at $b = 8/3$, $\sigma = 10$, $r = 5$. (d) Y-Z at $b = 8/3$, $\sigma = 10$, $r = 5$. (e) X-Y at $b = 8/3$, $\sigma = 10$, $r = 28$. (f) Y-Z at $b = 8/3$, $\sigma = 10$, $r = 28$.

The attractor's behavior with an assumed constant function $\beta(\tau)$ is an instant representation of the oscillatory attractors with a transient function $\beta(\tau)$, whereas the attractor

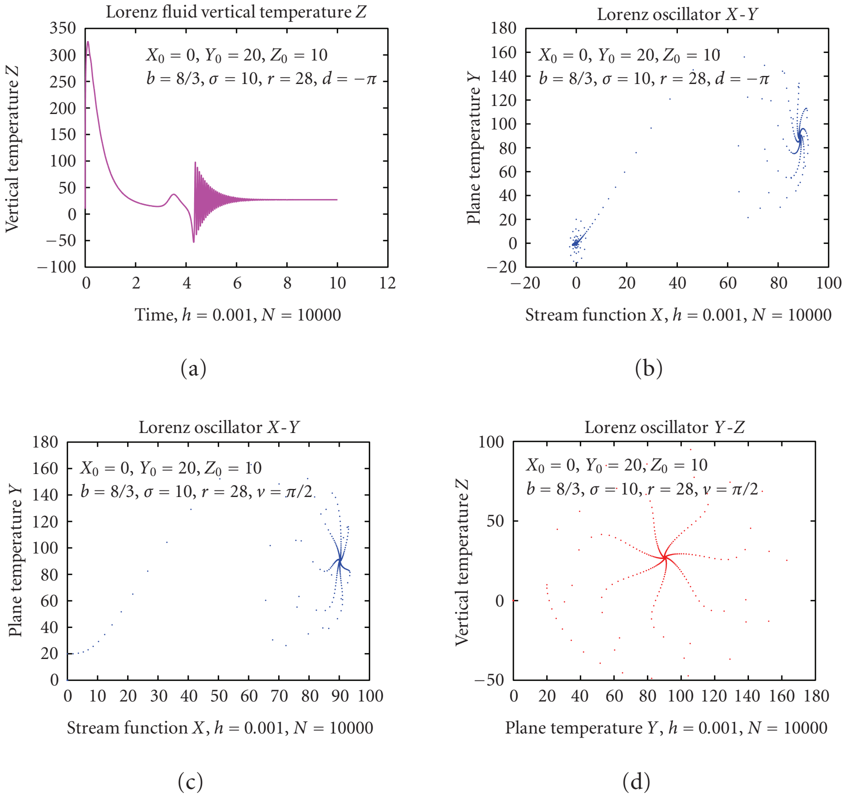


FIGURE 4.3. Response with a transient function $\beta(\tau)$, $b = 8/3$, $\sigma = 10$, $r = 28$. (a) Z versus time, $\beta(t) = 1/\tau$. (b) X-Y phase diagram, $\beta(t) = 1/\tau$. (c) X-Y phase diagran $\beta(t) = \pi/2$, $g(t) = \exp^{-(\pi/2)\tau}$. (d) Y-Z phase diagram, $\beta(t) = \pi/2$, $g(t) = \exp^{-(\pi/2)\tau}$.

experiences instantaneous oscillations with respect to the transient eigenvalues and function $\beta(\tau)$. The steady-state attractor also experiences instability when the function $\beta(\tau)$ approaches infinity such as in the case of $\beta(\tau) = -\omega\tan(\omega\tau)$. The oscillatory behavior of attractors associated with the transient function $\beta(\tau)$ means that the steady-state attractors can not be predicted based on the assumption of $\dot{X} = \dot{Y} = \dot{Z} = 0$, as that in (3.7). In fact, a steady-state attractor does not exist for a case of $\dot{X} \neq 0$, $\dot{Y} \neq 0$, $\dot{Z} \neq 0$.

(d) *The homoclinic bifurcation at* $r = 1000$*, a periodic oscillation.* In contrast with the bifurcation conditions associated with the steady-state attractors, another type of bifurcation occurs independent of these conditions, that is, the homoclinic explosion, a phenomenon that transform the steady-state oscillation to a newly born set of orbits. Figures 4.4(a) to 4.4(f) show a periodic oscillation as a result of the explosion at $b = 8/3$, $\sigma = 10$, $r = 1000$ and $\beta(\tau) = \pi$ with a transient function $g_T(\tau) = \exp^{-\pi\tau}$. The time history of X, Y and Z in Figures 4.4(a), 4.4(c), and 4.4(e), respectively, indicates a burst of the homoclinic explosion after the steady state is sustained for a certain period of time. The

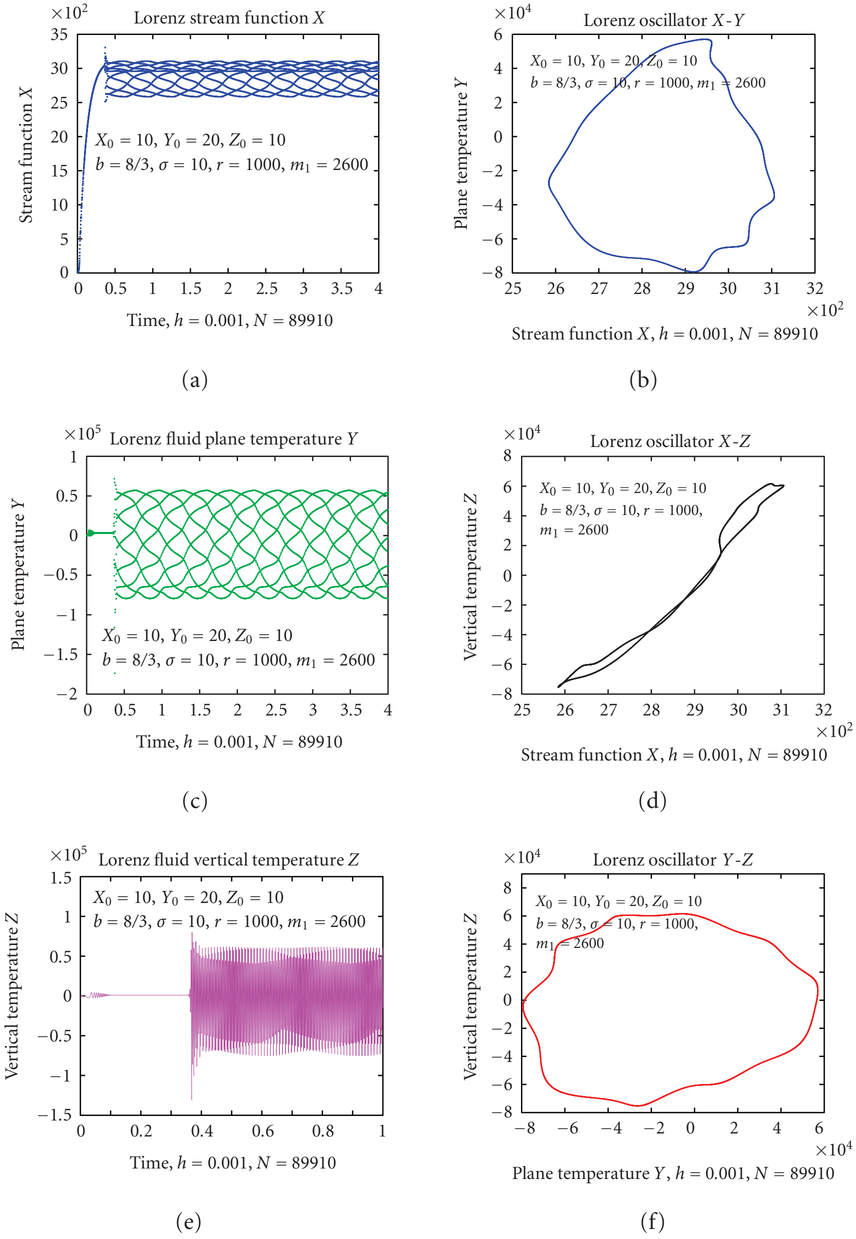


FIGURE 4.4. Response with a transient function $\beta(t) = \pi$, $b = 8/3$, $\sigma = 10$, $r = 1000$. (a) X versus time. (b) steady-state Poincare map X-Y. (c) Y versus time. (d) steady-state Poincare map X-Z. (e) Z versuss time. (f) steady-state Poincare map Y-Z.

Poincare maps shown in Figures 4.4(b), 4.4(d), and 4.4(f) are sampled at a frequency interval $\Delta h = 0.01\,\text{s}$, equivalent to 10 iterations for each point. Totally 87,310 periodic traces are taken in each map after eliminating the transient 2,600 iterations. These orbits remain the same as those with fewer sampling points, such as $N = 10{,}000$, or those at a different sampling frequency. These identical Poincare maps confirm that the orbit is periodic. The numerical computation results also reveal that the response with $\beta(\tau)$ from a linear transient function arrives at the same periodic orbit due to the explosion after reaching the steady state predicted by (3.7). $\beta(\tau)$ only influences the duration of the steady state prior to such an explosion, but not the orbit after the explosion.

(e) *The steady state at* $r = 903$*, a quasi-periodic oscillation.* Another bifurcation behavior exists, as can be observed from the phase diagrams at $b = 8/3$, $\sigma = 10$, and $r = 903$, where the homoclinic bifurcation leads to multiple periodic orbits. Figures 4.5(a), 4.5(c), and 4.5(e) show the bifurcation explosion after an initial steady-state sustained for about 2000 iterations. The phase diagrams in Poincare maps show a finite number of orbits in Figures 4.5(b), 4.5(d), and 4.5(f), after eliminating the initial 2600 transient iterations. Our extensive computation results verified that such a homoclinic bifurcation initiates at a higher Rayleigh number, that is, about $r = 900$. This transition number r is also affected by the exponential function frequency ω.

Figures 4.4 and 4.5 together suggest that the system experiences homoclinic bifurcations that lead to another steady state. This phenomenon agrees with the earlier observation from that of the original Lorenz model in that homoclinic explosions at a large r lead to periodic orbits [11]. Our results ascertain that such homoclinic explosions persist with different transient functions $\beta(\tau)$. Namely, a different $\beta(\tau)$ function leads to identical orbits as a result of the explosion, that is, either a monotonic periodic orbit or multiple period orbits.

There exists no valid explanation for such explosion phenomena that occur at a large r except consistent computation observations. However, examining the eigenvalues for each case suggests that cases with a large number r are associated with an insignificant real eigenvalues, that is, $\lambda_1 = 26.6$ at both $r = 1000$ and $r = 903$. At the same time, the complex conjugates have a trivial real part where the real and the imaginary part has a ratio in the order of 10^2, that is, $\lambda_{2,3} = -20.18 \pm 3139.10i$ for $r = 1000$; $\lambda_{2,3} = -20.16 \pm 2837.40i$ for $r = 903$, respectively. The steady-state attractors are at $X = 3138.9$, $Z = 999$ and $X = 2837.3$, $Z = 902$, respectively. These eigenvalues suggest that the amplitude of the steady-state oscillation is insignificant, due to the canceling effect between the real and the complex eigenvalues. Such a behavior is evident in the time history of the oscillation prior to the explosion. However, such trivial oscillation is subject to computing errors which can alter the eigenvalues and give birth to new periodic orbits. Therefore, the homoclinic orbit is a manifestation of the transition of eigenvalues as a result of the computation error perturbation. A single periodic oscillation is the consequence of a pair of stable eigenvalues due to such a perturbation, whereas a multiperiod oscillation occurs when the eigenvalues are unstable, experiencing multiple transitions among different values. Therefore, the phase diagram embodies either a finite number of periodic orbits when the eigenvalues are finite or an infinite number of orbits when the eigenvalues vary continuously. Although the function $\beta(\tau)$ influences the transient behavior, we observe that the initiation of the explosion is dependent on $\beta(\tau)$. The skew-shaped periodic orbit for the Poincare map in Figures 4.4(d) and 4.5(d) with $X \neq Y$ is the consequence of such perturbation

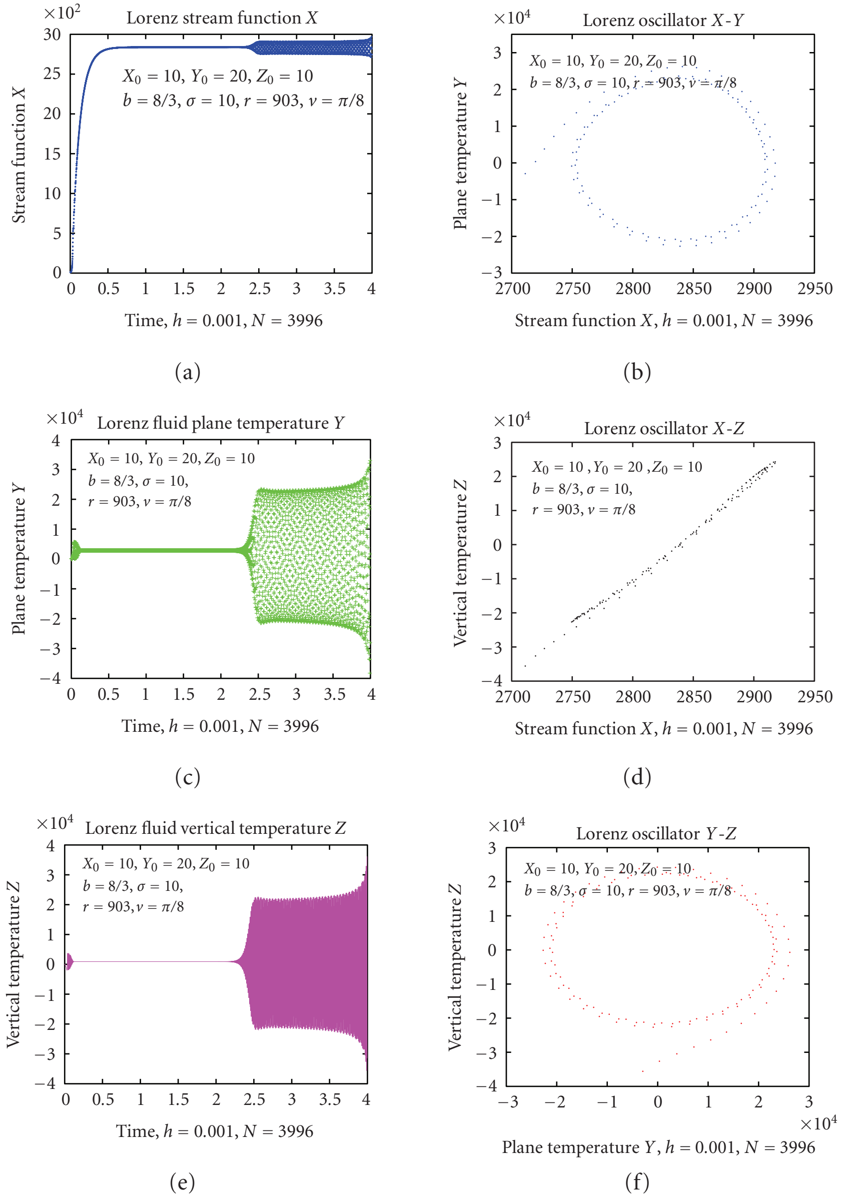


Figure 4.5. Response with a harmonic function $\beta(t) = -\nu\tan(\nu t)$, $b = 8/3$, $\sigma = 10$, $r = 903$, $\nu = \pi/8$. (a) X versus time. (b) Transient response Y versus time. (c) Z versus time. (d) steady-state Poincare map X-Y. (e) Steady-state Poincare map X-Z. (f) Steady-state Poincare map Y-Z.

that produces the orbits different from those predicted by the steady-state analysis. Another fact that can be verified from the expression of the coefficients in (3.12) is that a large r reduces the absolute values of λ_1 and α is in an reduced magnitude. In addition, the two values are related by the constant coefficient $I = 1 + b + \sigma = -(\lambda_1 + 2\alpha) < 0$. This makes the two in a comparable scale to reinforce the canceling effect that leads to the trivial oscillation magnitude. It is apparent that the characteristics of the eigenvalues offers a compelling argument for the homoclinic explosions at a large r.

5. Discussions and conclusions

Our bifurcation analysis and the computation results indicate that the thermally induced convection flow presents drastically different behaviors when the transient thermal field drives the flow. The transient form Lorenz model incorporates the influence of the conductive layer and the heat transfer boundary condition for the flow behavior of the entire 2D field. The transient thermal field function influences the steady-state and transient oscillations. We identified stationary steady-state attractors that exist subject to certain transient thermal field functions. The fluctuation of the thermal field modifies the attractors, bifurcation conditions for the initiation of the unstable flow. It also affects the bursts of the homoclinic bifurcation, though not the homoclinic orbit itself. The transient thermal field variation is likely to cause transitions among different bifurcation behaviors, which could generate turbulence or chaos due to instantaneous transitions of the attractors.

The bifurcation analysis from this study provides a quantified justification for the sequential bifurcations at different thermal and fluid parameters. This explains the successive bifurcations exhibited by the original Lorenz model as well as the current model at a different range of parameters. Further, we identified the mechanisms of the bursts of the homoclinic explosions at a large r. We attribute the explosions to the trivial effect of the oscillation amplitude determined by these eigenvalues at large r that is sensitive to numerical computation errors to alter the oscillation orbits.

This study revealed the typical behaviors of the thermally induced convection flow with a transient thermal source and predicted the system response in both qualitative and quantitative terms for the bifurcations of steady-state attractors. These bifurcation conditions shed light on the turbulence of the thermally induced convection flow.

Nomenclature

a: critical wave number
b: geometry factor
c: geometry factor
d: coefficient for the thermal boundary condition effect
d_N, d_D: coefficients d for the Newmann and Dirichlet condition, respectively
e_κ: coefficient for the thermal diffusivity between the fluid and the solid
$f^-_{\text{sup}}(r)$: necessary condition for the supercritical Hopf bifurcation with $\lambda_1 < 0$
$f^-_{\text{sub}}(r)$: necessary condition for the subcritical Hopf bifurcation with $\lambda_1 < 0$
$f^+_{\text{sub}}(r)$: necessary condition for the subcritical Hopf bifurcation with $\lambda_1 > 0$
$f^*_H(r)$: necessary and sufficient condition for the Hopf bifurcation with
a specified eigenvalue $\lambda_1 > 0$ or $\lambda_1 < 0$

fsd: bifurcation curve for the Hopf bifurcation with the concurrent saddle node bifurcation

fpd: bifurcation curve for the Hopf bifurcation with the concurrent period doubling bifurcation

fHp: bifurcation curve for the subcritical Hopf bifurcation at $\lambda_1 = 15$

fHn: bifurcation curve for the Hopf bifurcation at $\lambda_1 = -15$

$fh > 0$: the contour projection of the function $f_{\text{sub}}^{-}(r) > 0$

$fnh > 0$: the contour projection of the function $f_{nh}(r) > 0$

$f\,\text{sub}\,p > 0$: curve for the contour projection of the function $f_{\text{sub}}^{+}(r) > 0$

g: gravitational acceleartion

$g_T(t)$, $g_T(\tau)$: transient thermal field function

$g_{T,\tau}(\tau)$: time derivative of the transient thermal field function $g_T(\tau)$

H: height of the fluid cell

I, II, III: coefficients

J: Jocobi of the system

L: length of the fluid cell

r: ratio between R_a and R_c

r^*: threshold value of r for $r > 0$

R_a: Rayleigh number

R_c: critical Rayleigh number

$T(x,z,t)$: temperature of the flow field

T_0: magnitude of the temperature variation at the bottom layer

$T_1(x,z,t)$: linear temperature variation along z

$\Delta T(x,0,t)$: temperature difference between the two parallel plates

X: variable for the function $\theta_{11}(x,z,t)$

Y: variable for the function $\theta_{02}(x,z,t)$

Z: variable for the function $\psi(x,z,t)$

u, w flow in x and z, respectively

ε: the coefficient of volume expansion of the fluid

α: the real part of the eigenvalue

γ: the imaginary part of the eigenvalue

$\beta(\tau)$: ratio between $g_{T,\tau}(\tau)$ and $g_T(t)$

δ: intermediate variable

κ: thermal diffusivity of fluid

κ_L: thermal diffusivity of solid at the bottom plate

σ: Prandtl number

η: intermediate variable

λ: eigenvalue

τ: normalized time

ω: frequency of the transient thermal field function

θ, θ_{11}, θ_{02}: temperature variation of the flow field

ν: kinematic thermal viscosity

$\psi(x,z,t)$: flow field stream function

Δ_{nh}: operator for the attractors when the eigenvalues are real or complex

Δ_H^-: operator for the attractor at the Hopf bifurcation with $\lambda_1 < 0$

Δ_{sub}^+: operator for the attractor at the Hopf bifurcation with $\lambda_1 > 0$

Δ_H^*: operator for the attractor at the Hopf bifurcation with a specified λ_1

References

[1] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[2] B. Saltzman, "Finite amplitude free convection as an initial value problem—I," *Journal of the Atmospheric Sciences*, vol. 19, no. 4, pp. 329–341, 1962.

[3] L. Rayleigh, "On convection currents in a horizontal layer of fluid, when the higher temperature is on the other side," *Philosophical Magazine and Journal of Science*, vol. 32, pp. 529–546, 1916.

[4] J. A. Yorke, E. D. Yorke, and J. Mallet-Paret, "Lorenz-like chaos in a partial differential equation for a heated fluid loop," *Physica D*, vol. 24, no. 1–3, pp. 279–291, 1987.

[5] A. C. Fowler and M. J. McGuinness, "A description of the Lorenz attractor at high prandtl number," *Physica D*, vol. 5, no. 2-3, pp. 149–182, 1982.

[6] E. J. Kostelich and J. A. Yorke, "Lorenz cross sections of the chaotic attractor of the double rotor," *Physica D*, vol. 24, no. 1–3, pp. 263–278, 1987.

[7] E. N. Lorenz, "The local structure of a chaotic attractor in four dimensions," *Physica D*, vol. 13, no. 1-2, pp. 90–104, 1984.

[8] D. D. Joseph, "On the stability of the Boussinesq equations," *Archive for Rational Mechanics and Analysis*, vol. 20, no. 1, pp. 59–71, 1965.

[9] J. Guckenheimer, "Sensitive dependence to initial conditions for one dimensional maps," *Communications in Mathematical Physics*, vol. 70, no. 2, pp. 133–160, 1979.

[10] J. Guckenheimer, "A strange, strange attractor," in *The Hopf Bifurcation and Its Applications*, J. E. Marsden and M. McCracke, Eds., vol. 19 of *Applied Mathematical Science*, pp. 368–381, Springer, New York, NY, USA, 1976.

[11] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, vol. 41 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1982.

[12] J. B. McLaughlin and P. C. Martin, "Transitions to turbulence in a statistically stressed fluid system," *Physical Review A*, vol. 12, no. 1, pp. 186–203, 1975.

[13] J. B. McLaughlin, "Successive bifurcations leading to stochastic behavior," *Journal of Statistical Physics*, vol. 15, no. 4, pp. 307–326, 1976.

[14] J. H. Curry, "Chaotic response to periodic modulation of model of a convecting fluid," *Physical Review Letter*, vol. 43, no. 14, pp. 1013–1016, 1979.

[15] J. H. Curry, "A generalized Lorenz system," *Communications in Mathematical Physics*, vol. 60, no. 3, pp. 193–204, 1978.

[16] C. Boldrighini and V. Franceschini, "A five-dimensional truncation of the plane incompressible Navier-Stokes equations," *Communications in Mathematical Physics*, vol. 64, no. 2, pp. 159–170, 1979.

[17] V. Franceschini and C. Tebaldi, “Sequences of infinite bifurcations and turbulence in a five-mode truncation of the Navier-Stokes equations,” *Journal of Statistical Physics*, vol. 21, no. 6, pp. 707–726, 1979.

[18] J. D. Gibbon and M. J. McGuinness, “The real and complex Lorenz equations in rotating fluids and lasers,” *Physica D*, vol. 5, no. 1, pp. 108–122, 1982.

Xiaoling He: Department of Materials Engineering, College of Engineering and Applied Science,
University of Wisconsin-Milwaukee, 3200 North Cramer Street, P.O. Box 784, Milwaukee,
WI 53201-784, USA
Email address: xiaoling@uwm.edu

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 61543, 13 pages
doi:10.1155/2007/61543

Research Article

Patrol Mobile Robots and Chaotic Trajectories

Luiz S. Martins-Filho and Elbert E. N. Macau

Received 16 October 2006; Revised 19 January 2007; Accepted 16 April 2007

Recommended by José Manoel Balthazar

This paper presents a study of special trajectories attainment for mobile robots based on the dynamical features of chaotic systems. This method of trajectories construction is envisaged for missions for terrain exploration, with the specific purpose of search or patrol, where fast scanning of the robot workspace is required. We propose the imparting of chaotic motion behavior to the mobile robot by means of a planner of goal positions sequence based on an area-preserving chaotic map. As a consequence, the robot trajectories seem highly opportunistic and unpredictable for external observers, and the trajectories's characteristics ensure the quick scanning of the patrolling space. The kinematic modeling and the closed-loop control of the robot are described. The results and discussion of numerical simulations close the paper.

1. Introduction

Mobile robotics, after decades of continuous development, keeps up as an intensive research issue because of its ever-increasing application to different domains and its economical and technological relevances. Interesting applications can be seen in robots performing floor-cleaning tasks [1], executing industrial transportation [2], exploring volcanos, scanning areas to find explosive devices, and so on [3].

This work deals with the problem in which a mobile robot is used for searching or patrolling a defined region. To avoid risks to human life, it is very wise to use autonomous or remotely operated robots to deal with the hazardous tasks of detecting dangerous materials or an intruder. For these applications, the physical robot and its systems architecture and software have been studied extensively for military and civil operations. Some mobile patrol robots are already commercially available (e.g., [4]). Many examples of studies involving vigilance robots can be found in [5–9]. These works mainly focus on target perception and identification, robot localization, terrain map updating, optimization of communication with the operation center or other robots. Here, we are specifically interested in investigating a very critical and still open issue that is paramount for the success of these applications: the path planning. The main goal is to generate a very convenient trajectory to be followed by the rover so that it increases as much as possible the probability of finding the intruder inside the surveillance region. This trajectory must be as opportunistic as possible so that its developed thread on the region cannot be easily understood or predicted by the intruder. Otherwise, he can come up with a way that allow him to avoid the rover. As so, this problem of using a robot to patrol a defined region is actually a problem of conceiving a proper strategy that generates an opportunistic and properly crafted trajectory to be followed by the robot. We can affirm that high unpredictability for the robot trajectories as well as fast scanning of the workspace area are strongly required. In this work, we introduce a very convenient strategy to accomplish these basic requirements. Our strategy exploits the dynamical behavior of a conservative chaotic system to generate trajectories to be followed by the rover. As a spin-off of this approach, previous terrain mapping is no longer necessary.

The interaction between mobile robotics and chaos theory has been studied only recently, as can be seen in [10–12]. For instance, the integration between the robot motion system and a chaotic system, the Arnold dynamical system, is used to impart chaotic behavior to a robot in [13]. An extension of this motion control strategy, applying diverse chaotic systems on integration with the robot kinematics model, can be found in [14]. In [15], the same principle of systems integration is used, however a Van der Pol equation is associated with the target. In [16], an open-loop control approach is proposed to produce unpredictable trajectories so that the state variables of the Lorenz chaotic system are used to command the velocities of the robot's wheels.

Here, we introduce a new strategy that generates an opportunistic and proper crafted trajectory that works as follows. We associate a path-planning generator module with a closed-loop locomotion control module. At each step, the first one generates a position goal defined by its coordinates in the phase space. This position goal is provided to the second module, which drives the mobile rover from its actual position to the desired one. When the rover arrives at the desired position, the path-planning generator module is used again to give another position goal, which is subsequently provided to the second module. This sequence of action is repeated over and over again. The path-planning generator module is implemented by exploiting the dynamics of an area-preserving map in a chaotic regime. Different to a dissipative chaotic map, in which the chaotic evolution takes place on attractors, the chaotic region of an area-preserving map for specific parameter values extends practically over all of its phase spaces. For these parameter values, the entire phase space is covered by a single chaotic orbit. It is therefore possible to make an association between the physical region that we wish to scan with the phase space defined for the area-preserving map. Thus, the position goals are generated as transformed iterations of an area-preserving map and so present a chaotic dynamics with its remarkable characteristic that makes long-time prediction based on measurements very difficult. Between the position goals, the rover trajectory is driven by a closed-loop locomotion control module which allows a short trajectory between the points and introduces an element of regularity to the strategy. By properly combining the parameters of both modules, we can exploit a multitude of possibilities between a quick scanning of a region by using a trajectory with "small" level of unpredictability to a "slow" scanning with a highly unpredictable trajectory. Actually, by properly combining the parameters, we are able to come up with a very efficient, opportunistic, and properly crafted trajectory that fits the desired requirements for patrol missions.

This paper presents the proposed path planning (Section 2). The robot model and the adopted control are described in Section 3. Section 4 presents the simulation results and discussions, and Section 5 concludes the paper.

2. Chaotic trajectory planning

In the context of deterministic systems, sensitive dependence on initial conditions is the main well-known characteristic of the chaotic behavior [17]. This means that arbitrarily close initial conditions imply trajectories that move far away from each other after some time. This property of a chaotic system makes long-term prediction of a chaotic trajectory based on finite-time measurements practically impossible because of the limited accuracy associated to measurement sensors. Another intrinsic characteristic of a chaotic evolution is the transitivity [17]. A deterministic system is transitive on an invariant set if for any two open subsets U and V of this invariant set, there exist trajectories originating from U that pass through V after some time. This property means that we always can use a chaotic trajectory as a transportation path between regions belonging to the chaotic invariant set. It implies the "mixing property" founded in chaotic systems that ensures that the system cannot be broken down into subsystems that do not interact with each other. From the perspective of an external observer, a chaotic trajectory presents a complicated behavior that does not exhibit any recurrent pattern and seems to be random. In other words, a chaotic trajectory is reported by an external observer as an erratic trajectory that quickly moves among different regions of a certain invariant set. It is precisely this behavior that we exploit in this work to orient the movement of a robot to make it very suitable to be used as an opportunistic patrolling engine.

The space to be scanned by the robot can be viewed as a kind of a continuous subset with an integer dimension 2. In order to accomplish our goal, the chaotic trajectory must fill densely this integer dimension subset. This requirement provides an extra ingredient that delimits the class of chaotic system to be used. In a dissipative chaotic system, the chaotic invariant set is an attractor. In general, this attractor has a noninteger dimension, that is, its geometric picture on the phase space is a fractal. However, in our problem, the robot must cover densely its patrol region. Consequently, we consider it more appropriate to choose another class of chaotic systems: an area-preserving chaotic system. One of the basic properties of an area-preserving system is that it preserves volumes in the phase space [18]. As a consequence of this property, these systems do not have attractors, and the chaotic regions spread densely over regions of the phase space. Note that this fact individualizes our approach in the scenario of previous works that uses a chaotic system to run the robot dynamics. Because previous approaches use dissipative chaotic systems (e.g., [13]), they require subterfuges to make the robot wander opportunistically through the patrolling area. This is not necessary in this approach. For the knowledge of the authors, this is the first time that this concept is applied in the area of mobile robots.

2.1. The standard map-based path planning. The planning procedure is based on determining a sequence of intermediary goal points (coordinates x and y) that will compose the robot trajectory. The path-planning generator module uses an area-preserving map that is considered as a paradigm for area-preserving chaotic systems: the standard map, also called Taylor-Chirikov map [19]. It is a two-dimensional map which results from a periodic impulsive kicking of a rotor. This map was firstly proposed by Brian Taylor and then independently obtained by Chirikov [20] to describe the dynamics of magnetic field lines on the kicked rotor [18]. The dynamic effect of this system is expressed mathematically through the map equations, given by

$$\begin{aligned} x_{n+1} &= x_n + K\sin y_n, \\ y_{n+1} &= y_n + x_{n+1}, \end{aligned} \tag{2.1}$$

where x is a periodic configuration variable (angular position) and y is the momentum variable (angular speed). These map variables are both computed mod(2π). The map parameter K represents the strength of the nonlinear kick applied in the rotor mechanism. In its phase space and according to the value associated with the parameter K, it has stable and unstable periodic orbits, Kolmogorov-Arnold-Moser (KAM) surfaces, and chaotic regions. Depending on the nonlinear parameter K, the regions of regular motion and the regions of chaotic motion are complexly interwoven, but the chaotic regions are confined between KAM tori. As this parameter is increased, the KAM surfaces start to be destroyed, chaotic regions occupy increasingly large areas until, for a specific value of K, the last KAM torus is destroyed and the entire region of the phase space appears to be densely covered by a single chaotic orbit. Our path-planning generator module is implemented based on the standard map that presents this dynamics.

Let us now show an example of how the standard map is used in the context of our path-planning generator module. By numerically simulating the map equations, we can analyze the properties of terrain covering considering the basic mission requirements for fast terrain scanning. We define a square terrain with dimensions 100×100 in a normalized measurement unit. The map simulation begins with an arbitrary initial position, and considers the gain value $K = 7$. We can see that the third case can cover completely the considered terrain (in fact, the necessary condition for the complete scan is $K > 6$ [20]). The results of passage locations planned for 100 and 3000 iterations can be seen in Figure 2.1.

The terrain covering can be judged through a performance index. This index is defined using a terrain division on square unit cells (e.g., 1×1, i.e., 10 000 cells), and computing the visited cells percentage after the robot locations planning. This index of terrain covering is presented in the form of a plotting index *versus* planning evolution, as can be seen in Figure 2.2 (where index = 1 represents 100% of cells visited). However, this analysis does not consider the robot trajectories between two subsequent locations, that will be

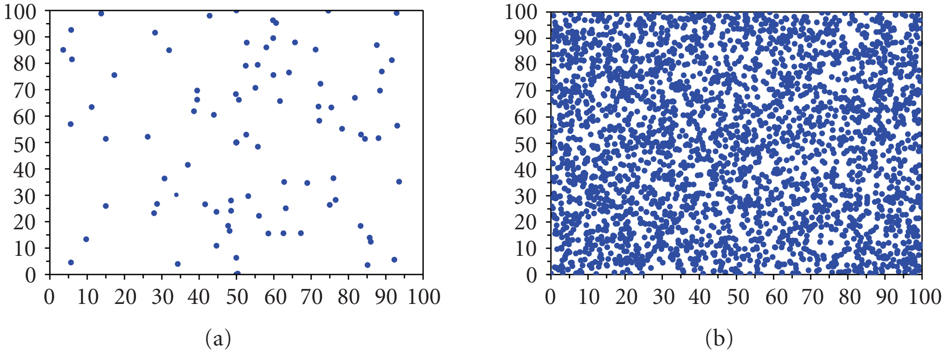


Figure 2.1. Terrain covering by subgoals planned points using standard maps after 100 and 3000 planning iterations (considered map gain value $K = 7$).

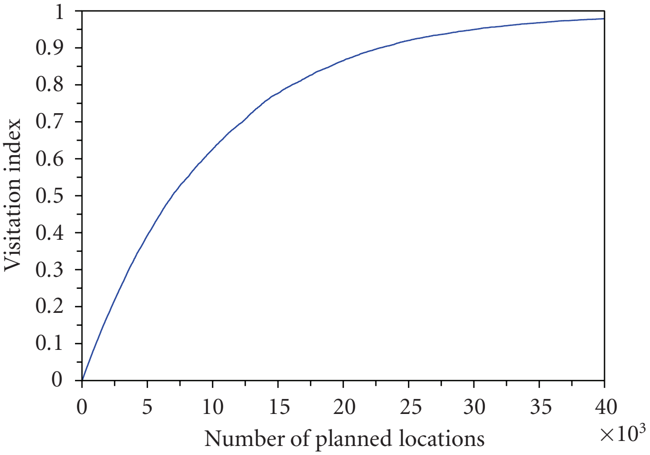


Figure 2.2. Index of terrain covering (visited cells portion) for 40 000 iterations, with map gain $K = 7$.

taken in account later in this paper. It is quite evident that a faster complete area covering could be obtained using a systemic scan without passing two or more times at one same terrain cell, but this classic strategy is absolutely predictable, and inadequate to the patrolling mission.

Another strategy to plan these points could be defined using a random numbers generator. Considering a uniform distribution random sequence, we obtain a very similar terrain space covering. The results for this alternative planning strategy are shown in Figures 2.3 and 2.4. Even if the appearance and the terrain covering are similar, the planning nature is quite different to the conservative standard map. We will discuss this fundamental difference in Section 4.

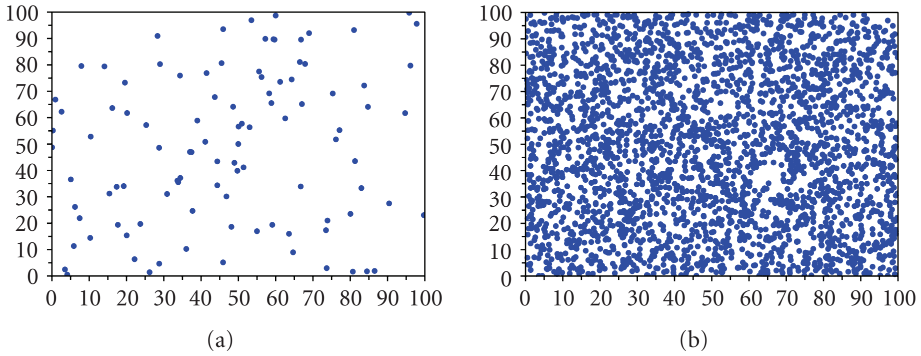


Figure 2.3. Terrain covering by subgoals planned points using a random numbers generator after 100 and 3000 planning iterations.

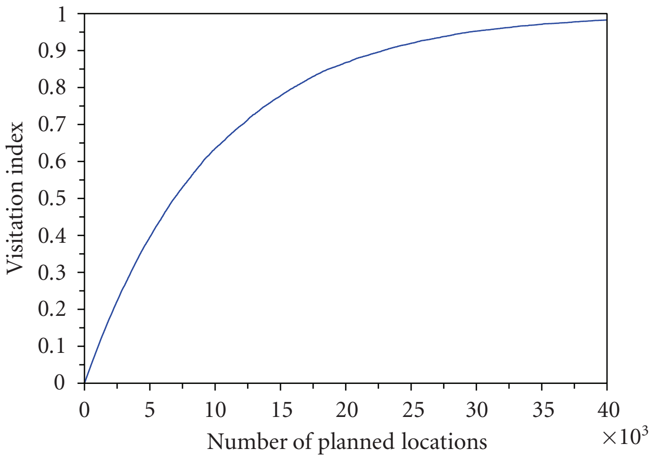


Figure 2.4. Terrain covering by subgoals planned points using a random numbers generator (40 000 planning iterations).

3. Kinematic control of the mobile robot

The mobile robot considered in this work is a typical differential motion robot with two degrees of freedom, composed by two active, parallel, and independent wheels, a third passive wheel with exclusive equilibrium functions (a sort of free steered standard wheel), and proximity sensors capable of obstacles detection. The active wheels are independently controlled on velocity and sense of turning. The sensors provide short-range distances to obstacles. For instance, these sensors could be sonar or infrared devices commonly used in mobile robots, with adequate accuracy. Additionally, the robot is assumed to be equipped with specific sensors for detection and recognition of the searched objects or intruders. This robot model represents an interesting compromise between control simplicity and degrees of freedom that allow the robot to accomplish mobility requirements

[21], and it is largely adopted in several researches on mobile robotics, for example, [22–24].

The robot chassis is considered as a rigid body operating on the horizontal plane. Its motion is obtained by driving the active wheels. The resultant motion is described in terms of linear velocity $v(t)$ and direction $\theta(t)$, representing an instantaneous linear motion of the medium point of the wheel axis and a rotational motion (rotational velocity $\omega(t)$) of the robot body over this same point.

The robot motion control can be done providing the wheels velocities, $\omega_l(t)$ and $\omega_r(t)$, or equivalently the body linear and angular velocities, $v(t)$ and $\omega(t)$, called input or control variables. The mathematical model of this kinematics problem considers these two input variables and also three state variables: the robot position and orientation ($x(t)$, $y(t)$, $\theta(t)$) [21]:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix}. \tag{3.1}$$

These equations constitute a nonholonomic dynamical system. The control of this system has been studied extensively by various research groups, and diverse solutions are available, for example, [22, 23]. The motion control strategy adopted in this work involves a state feedback controller, proposed in [24], which is an appropriate approach to produce a desired trajectory described by a sequence of coordinates (x_p, y_p). This means that the path-planning task is given by a specialized robot module, independent of the motion control module, that sets intermediate positions lying on the requested path.

The adopted control law considers the geometric situation shown in Figure 3.1. In this figure, the robot is placed at an arbitrary configuration (position x, y and orientation θ), and a desired position (the target x_p, y_p: the origin of frame $X_G Y_G$) is defined by the robot path-planner. In the robot reference frame $X_R Y_R$, the configuration error vector is defined by $e = [\rho\ \varphi]^T$, where ρ and φ localize the target position, and provide a coordinate change:

$$\begin{aligned} \rho &= \left(\Delta x^2 + \Delta y^2\right)^{1/2}, \\ \varphi &= 180 + \theta - \psi. \end{aligned} \tag{3.2}$$

The robot kinematics model is described by (3.1), where $\dot{x}(t)$ and $\dot{y}(t)$ are the linear velocity components in the absolute reference frame (fixed on the workspace). We define the angle φ between the X_R axis of body reference frame and the vector connecting the robot center and the desired position. The other configuration variables ρ and ψ represent, respectively, the distance between present and desired positions, and the angle between the direction to the target and the axis X_0.

The description of the motion in the new coordinates becomes

$$\begin{bmatrix} \dot{\rho} \\ \dot{\varphi} \end{bmatrix} \begin{bmatrix} -\cos\varphi & 0 \\ \dfrac{1}{\rho}\sin\varphi & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix}. \tag{3.3}$$

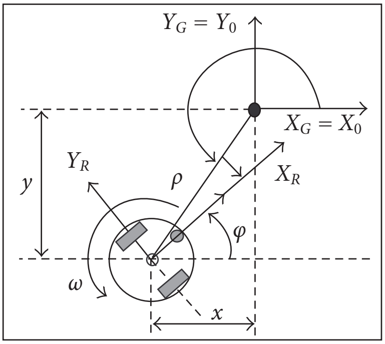


Figure 3.1. The robot control problem configuration, where the position/orientation error (ρ, φ), the linear and angular velocities (v, ω), the robot frame $(X_R Y_R)$, and the desired position frame $(X_G Y_G \equiv X_0 Y_0)$ can be seen.

Concerning these polar coordinates system descriptions, it is necessary to remark that the coordinate transformation is not defined at $x = y = 0$, that is, when the robot achieves the goal location. The adopted control law [24], in terms of error feedback (ρ, φ) to determine the value of system inputs v, ω, is given by

$$\begin{aligned} v &= k_1 \rho \cos\varphi, \\ \omega &= -k_1 \sin\varphi \cos\phi - k_2 \varphi. \end{aligned} \tag{3.4}$$

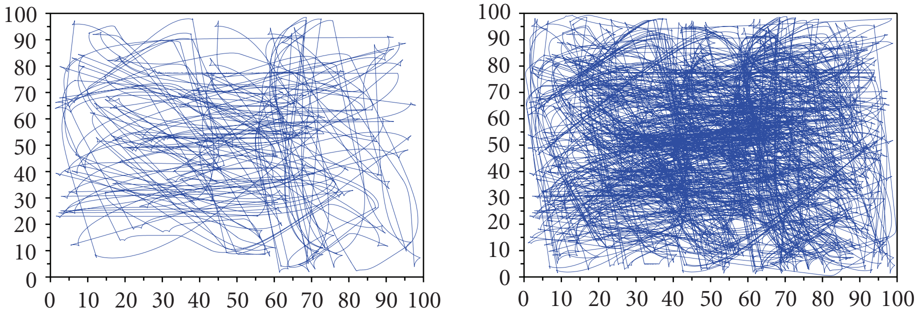
Using the Lyapunov stability analysis, we can verify the robot kinematic system with the application of this nonlinear control law [25]. Composing a quadratic Lyapunov function with the state variables ρ and φ given by $V = (1/2)(\rho^2 + \varphi^2)$, the time derivative of this Lyapunov function, considering the control law, is

$$\dot{V} = \rho(-v\cos\varphi) + \varphi\left(\omega + \frac{v}{\rho}\sin\varphi\right) = -k_1(\rho\cos\varphi)^2 - k_2\varphi^2. \tag{3.5}$$

The constant control gains k_1 and k_2 are exclusively positive. As a consequence, the value of $\dot{V}$ is negative for all nonnull value of (ρ, φ), and null at the origin of state space. That is a sufficient condition for the asymptotic convergence of (3.3).

If an obstacle is found on the trajectory, a specific navigation competence, obstacle avoidance, must be used to drive the robot from this obstacle. Is this work, the obstacle avoidance problem is not treated, nevertheless a simple solution could be implemented, for example, using the algorithm proposed in [26].

Using this feedback control law, we intend to validate the proposed path planning, examining the trajectory unpredictability, the terrain covering by the robot motion, and the general characteristics of the obtained trajectories. Nevertheless, we note that any stable control law can provide adequate closed-loop motion respecting the planned sequence of passage points, and obtaining the chaotic robot motion.



(a)

(b)

Figure 4.1. Two instants of the mobile robot trajectory evolution, after 200 and 1000 planned passage locations, using the adopted continuous control law (considering a terrain with dimension 100 × 100 in a normalized measurement unit).

4. Numerical simulations

To test our mobile robot patrolling approach, we have simulated the robot kinematic motion applying the closed-loop control law discussed in the previous section to track a sequence of planned objective points, provided by the standard map. The adopted control law obtains smooth trajectories between two subsequent locations, reinforcing the apparently erratic nature of the movement, which constitute an interesting feature for patrol missions. The trajectory results of the application of this control law are shown in Figure 4.1. For any other terrain shape, the planning process can fit the area of interest inside a square standard map, ignoring the points planned outside the terrain but ensuring the desired fast workspace scan.

We analyze the terrain covering of this executed (or effective) trajectory using the same performance index proposed for the planned sequence of points (Section 2), that is, the terrain is divided in 10 000 square unit cells, and the index represents the visited cells percentage (index = 1 represents 100%). The time evolution of the terrain covering index is shown in Figure 4.2. We do not take in to account here the extra area covered by the sensors perception range region that certainly could augment the area covering indexes. The terrain scan analysis presented here is based on a worst case in which the sensor perception field is considered as a single point.

We can see that the central part of the terrain will be visited more often than the regions close to the borders. This is a consequence of frequent changes of motion direction and it is in accordance with the strategy which is to cover completely the patrolled terrain all the time so that an intruder appearing in an arbitrary time/location inside the terrain is quickly detected by the rover. Evidently, the complete terrain covering will be obtained only after a long-term execution of the robot planning/motion procedure.

Considering the basic requirements of patrol missions, and the main ideas of our approach, we can discuss different ways to obtain the construction of mobile robot trajectories by combining the adoption of a motion control law and a strategy for the determination of the regions sequence to be inspected by the robot.

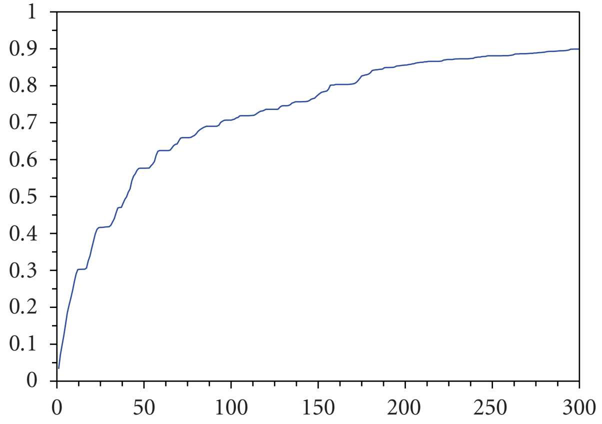


Figure 4.2. The time evolution of the index of terrain covering for the executed robot trajectory (300 simulation time intervals).

In the example presented, we adopt a continuous control law that provides smooth trajectories and minimizes unnecessary maneuvers and consequent control switches. Another very different but intuitive control strategy could be adopted: a discontinuous control law based on an initial rotational maneuver around the robot wheel axis center to orientate the robot towards the next desired position, followed by a straight trajectory to the objective.

In Section 2, we mentioned an alternative approach to be used as our path-planning generator module: to use a series of random locations, uniformly distributed in the patrolled space. This sort of random planning strategy results in a similar terrain covering to the chaotic planning. However, the nature of locations planning is quite different: the chaotic one is deterministic.

We can conclude our proposed approach by analyzing the advantages and disadvantages of these aspects of the patrolling scenario: the type of control law and the inspection planning strategy. Firstly, we compare the two options of control laws presented here, continuous and discontinuous, and we can verify that the continuous one offers advantages, because the smoother trajectories save control switches and maneuvers; moreover, they contribute to perform unpredictable trajectories for external observers. On the other hand, the discontinuous control produces a sequence of piecewise predictable straight trajectories (an example can be seen in Figure 4.3). However, in terms of time of terrain scanning, the advantageous control approach is the discontinuous one, because it will evidently cover faster the terrain than the discontinuous control when using straight trajectories.

If we compare the two options of scanning plan, chaotic and random approaches, both of them provide very similar results in terms of terrain scan appearance and patrol covering. In spite of that, chaotic trajectories have an important advantage, they are based on a deterministic sequence of objective points. This means that the behavior of the rover can be predicted in advance for the system designer. In terms of navigation expertise, this

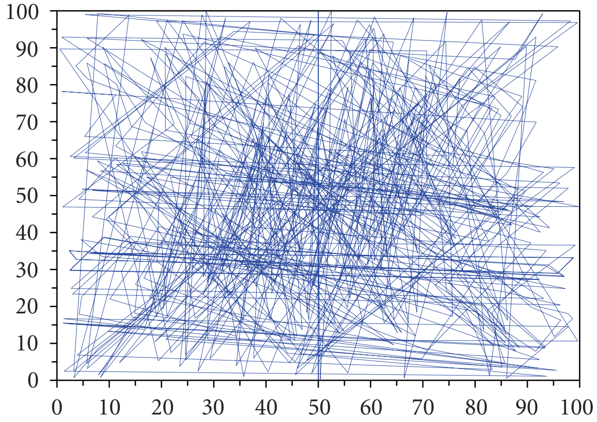


Figure 4.3. An example of trajectories obtained using the discontinuous control law (500 planned passage locations, standard map planning).

path-planning determinism represents an important advantage over navigation based on a sort of random walk trajectory. This feature can facilitates the frequent robot localization procedure, which is a crucial function because the knowledge of the robot position with appropriate precision constitutes very necessary information for the robot itself and also for the mission operation center. This determinism can be also advantageous to other mission aspects, for example, executed trajectory supervision, terrain's scanning information, and precise localization of intruders or targets.

5. Conclusion

The presented strategy to deal with terrain exploration missions for mobile robots can achieve adequately the main requirements for patrol missions: we are able to achieve very efficient, opportunistic, and proper crafted trajectory that fits the desired requirements for patrol missions. Imparting chaotic motion behavior to the robot motion through the utilization of an area-preserving chaotic map as a path planner ensures high unpredictability of robot trajectories, resembling a nonplanned erratic motion from external observers' point of view. Validation tests, based on numerical simulations of closed-loop motion control to follow the sequence of objective points on the robot trajectory, confirm that the chaotic planning procedure can obtain adequate results. The advantageous property of the proposed chaotic motion planning over unplanned or randomly planned motion resides mainly in the deterministic nature of chaotic behavior, which can be useful for important functions of the robot motion control, for example, the robot localization and the terrain mapping.

This study shows that the application of dynamical behaviors of nonlinear systems to solutions for mobile robots control problems represents an interesting interdisciplinary interface for researchers of both scientific domains, opening promising perspectives of future works including experimental realizations.

Acknowledgments

The authors acknowledge the financial support of *Conselho Nacional de Desenvolvimento Científico e Tecnológico* - CNPq, Brazil, *Fundações de Amparo à Pesquisa dos Estados de Minas Gerais* and *São Paulo* (FAPEMIG) (Minas Gerais, Brazil), and FAPESP (São Paulo, Brazil), and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), Brazil.

References

[1] J. Palacin, J. A. Salse, I. Valganon, and X. Clua, "Building a mobile robot for a floor-cleaning operation in domestic environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 5, pp. 1418–1424, 2004.

[2] J. H. Suh, Y. J. Lee, and K. S. Lee, "Object-transportation control of cooperative AGV systems based on virtual-passivity decentralized control algorithm," *Journal of Mechanical Science and Technology*, vol. 19, no. 9, pp. 1720–1730, 2005.

[3] B. Rooks, "Robotics outside the metals industries," *Industrial Robot*, vol. 32, no. 3, pp. 205–208, 2005.

[4] MobileRobots Inc., "PatrolBot Brochure," Amherst, NH, USA, January 2006, http://www.MobileRobots.com/.

[5] E. Krotkov and J. Blitch, "The defense advanced research projects agency (DARPA) tactical mobile robotics program," *The International Journal of Robotics Research*, vol. 18, no. 7, pp. 769–776, 1999.

[6] H. R. Everett and D. W. Gage, "From laboratory to warehouse: security robots meet the real-world," *The International Journal of Robotics Research*, vol. 18, no. 7, pp. 760–768, 1999.

[7] H. R. Everett, "Robotic security systems," *IEEE Instrumentation & Measurement Magazine*, vol. 6, no. 4, pp. 30–34, 2003.

[8] F. Capezio, A. Sgorbissa, and R. Zaccaria, "GPS-based localization for a surveillance UGV in outdoor areas," in *Proceedings of the 5th IEEE International Workshop on Robot Motion and Control (RoMoCo '05)*, pp. 157–162, Dymaczewo, Poland, June 2005.

[9] D. M. Carroll, C. Nguyen, H. R. Everett, and B. Frederick, "Development and testing for physical securtiy robots," in *Unmanned Ground Vehicle Technology VII*, vol. 5804 of *Proceedings of SPIE*, pp. 550–559, Orlando, Fla, USA, March 2005.

[10] C. Choi, S.-G. Hong, J.-H. Shin, I.-K. Jeong, and J.-J. Lee, "Dynamical path-planning algorithm of a mobile robot using chaotic neuron model," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '95)*, vol. 2, pp. 456–461, Pittsburgh, Pa, USA, August 1995.

[11] M. Islam and K. Murase, "Chaotic dynamics of a behavior-based miniature mobile robot: effects of environment and control structure," *Neural Networks*, vol. 18, no. 2, pp. 123–144, 2005.

[12] U. Nehmzow, "Quantitative analysis of robot-environment interaction-towards "scientific mobile robotics"," *Robotics and Autonomous Systems*, vol. 44, no. 1, pp. 55–68, 2003.

[13] Y. Nakamura and A. Sekiguchi, "The chaotic mobile robot," *IEEE Transaction on Robotics and Automation*, vol. 17, no. 6, pp. 898–904, 2001.

[14] A. Jansri, K. Klomkarn, and P. Sooraksa, "On comparison of attractors for chaotic mobile robots," in *Proceedings of the 30th IEEE Annual Conference of Industrial Electronics Society (IECON '04)*, vol. 3, pp. 2536–2541, Busan, Korea, November 2004.

[15] Y. Bae, "Target searching method in the chaotic mobile robot," in *Proceedings of the 23rd IEEE Digital Avionics Systems Conference (DASC '04)*, vol. 2, pp. 12.D.7–12.1–9, Salt Lake City, Utah, USA, October 2004.

[16] L. S. Martins-Filho, R. F. Machado, R. Rocha, and V. S. Vale, "Commanding mobile robots with chaos," in *ABCM Symposium Series in Mechatronics*, J. C. Adamowski, E. H. Tamai, E. Villani, and P. E. Miyagi, Eds., vol. 1, pp. 40–46, ABCM, Rio de Janeiro, Brazil, 2004.

[17] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Studies in Nonlinearity, Westview Press, Boulder, Colo, USA, 2nd edition, 2003.

[18] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion*, vol. 38 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1983.

[19] J. D. Meiss, "Symplectic maps, variational principles, and transport," *Reviews of Modern Physics*, vol. 64, no. 3, pp. 795–848, 1992.

[20] B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," *Physics Reports*, vol. 52, no. 5, pp. 263–379, 1979.

[21] R. Siegwart and I. R. Nourbakhsh, *Introduction to Autonomous Mobile Robots*, MIT Press, Cambridge, Mass, USA, 2004.

[22] A. Astolfi, "Exponential stabilization of a mobile robot," in *Proceedings of the 3rd European Control Conference (ECC '95)*, pp. 3092–3097, Rome, Italy, September 1995.

[23] C. Canudas de Wit and O. J. Sørdalen, "Exponential stabilization of mobile robots with nonholonomic constraints," *IEEE Transactions on Automatic Control*, vol. 37, no. 11, pp. 1791–1797, 1992.

[24] S.-O. Lee, Y.-J. Cho, M. Hwang-Bo, B.-J. You, and S.-R. Oh, "A stable target-tracking control for unicycle mobile robots," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '00)*, vol. 3, pp. 1822–1827, Takamatsu, Japan, October-November 2000.

[25] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice- Hall, Upper Saddle River, NJ, USA, 1991.

[26] V. J. Lumelsky and T. Skews, "Incorporating range sensing in the robot navigation function," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 5, pp. 1058–1069, 1990.

Luiz S. Martins-Filho: Departamento de Computação (DECOM), Universidade Federal de Ouro Preto (UFOP), Campus Morro do Cruzeiro, 35400-000 Ouro Preto, MG, Brazil
Email address: luizm@iceb.ufop.br

Elbert E. N. Macau: Laboratório Associado de Computação e Matemática Aplicada (LAC), Instituto Nacional de Pesquisas Espaciais (INPE), P.O. Box 515, 12227-010 São José dos Campos, SP, Brazil
Email address: elbert@lac.inpe.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 83893, 16 pages
doi:10.1155/2007/83893

Research Article

Inductorless Chua's Circuit: Experimental Time Series Analysis

R. M. Rubinger, A. W. M. Nascimento, L. F. Mello, C. P. L. Rubinger, N. Manzanares Filho, and H. A. Albuquerque

Received 8 September 2006; Revised 6 November 2006; Accepted 11 February 2007

Recommended by José Manoel Balthazar

We have implemented an operational amplifier inductorless realization of the Chua's circuit. We have registered time series from its dynamical variables with the resistor R as the control parameter and varying from 1300 Ω to 2000 Ω. Experimental time series at fixed R were used to reconstruct attractors by the delay vector technique. The flow attractors and their Poincaré maps considering parameters such as the Lyapunov spectrum, its subproduct the Kaplan-Yorke dimension, and the information dimension are also analyzed here. The results for a typical double scroll attractor indicate a chaotic behavior characterized by a positive Lyapunov exponent and with a Kaplan-Yorke dimension of 2.14. The occurrence of chaos was also investigated through numerical simulations of the Chua's circuit set of differential equations.

1. Introduction

Chaotic electronic circuits [1] have been widely studied during the last few decades due to their easy implementation, robustness, reproducibility of results, and also as a test platform for synchronization [2–4], chaos control [4–6], signal encryption [7], and secure communications [8, 9]. Also it is easy, through Kirchhoff's laws, to obtain the circuit described by a set of differential equations and carry on simulations which in most times, present good agreement with experimental data. The Chua's circuit [1, 10] is one of the most famous circuits on the literature and the reasons, among others, are:

(1) Chua's circuit has a quite simple construction characterized by four passive linear elements and one of them with nonlinear $i(V)$ characteristic represented by a piecewise linear equation, as shown in Figure 1.1;

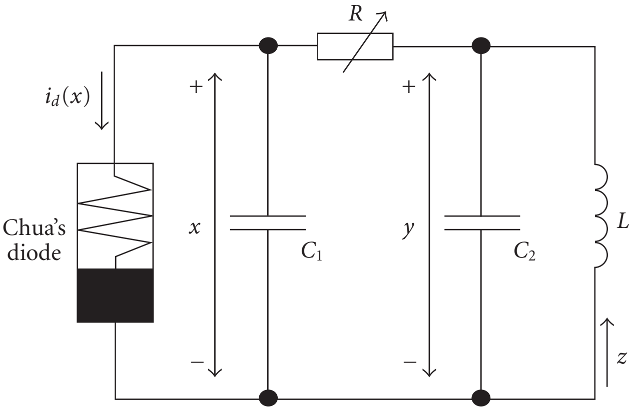


Figure 1.1. Chua's circuit. The dynamical variables are x, y, and z corresponding to the voltage across capacitor C_1, the voltage across capacitor C_2, and the current through the inductor, respectively. The nonlinear element is the Chua's diode and the nonlinearity is presented through $i_d(x)$ characteristics.

(2) it exhibits a number of distinct routes to chaos and multistructural chaotic attractors [11];
(3) attractors that occur in Chua's circuit arise from very complex homoclinic tangencies and loops of a saddle focus [11];
(4) many opened questions on the system's behavior and the lack of a possibility to fully describe Chua's circuit from its equations [11].

The Chua's circuit has been the object of study of hundreds of papers, where its topological, numerical, physical, and dynamical characterizations are deeply investigated. See [12–15] and references therein.

Point (4) suggests that numerical analysis such as that carried on this work could provide some contributions to understand Chua's circuit dynamical behavior. Chua's circuit dynamical equations are given by

$$\begin{aligned}\frac{dx}{dt} &= f_1(x,y,z) = \left(\frac{y-x}{RC_1}\right) - \frac{i_d(x)}{C_1},\\ \frac{dy}{dt} &= f_2(x,y,z) = \left(\frac{x-y}{RC_2}\right) + \frac{z}{C_2},\\ \frac{dz}{dt} &= f_3(x,y,z) = -\frac{y}{L} - z\left(\frac{r_L}{L}\right),\\ i_d(x) &= m_0x + \frac{1}{2}(m_1 - m_0)\{|x+B_P| - |x-B_P|\},\end{aligned} \tag{1.1}$$

where R, C_1, C_2, and L are passive linear elements, r_L is the inductor's resistance, i_d is the current through Chua's diode with m_0, m_1 and B_p as parameters.

Chua's diode and the active component "inductor" were implemented according to Tôrres and Aguirre [16]. This inductor implementation turns easy and compact to construct the Chua's circuit. It has another advantage since it can be designed as resistance free as have been carried on this work.

This paper is organized as follows: Section 2 is devoted to a detailed description of the parameters used to build and analyze Chua's circuit. A brief study of the equilibrium points of Chua's differential equations and the existence of a homoclinic loop is presented in Section 3. This was carried in order to identify the possible dynamical behavior for the chosen parameters of the circuit and to support the analyses carried in Section 4. In Section 4, we present the time series analysis of some illustrative experimental time series obtained from the Chua's circuit implementation. This section is the core of our work. Our aim is to characterize attractors obtained from this particular implementation of Chua's circuit with respect to its sensitivity to initial conditions and its dimension on the state space. Finally, concluding remarks are presented in Section 5.

2. Experimental details

Chua's circuit was constructed in a single face circuit board with the same scheme of [16] but with all capacitors 1000 times lower. This way $C_1 = 23.5\,\mathrm{nF}$, $C_2 = 235\,\mathrm{nF}$, and $L = 42.3\,\mathrm{mH}$. These values were obtained from the combination of passive components and measured with a digital multimeter with a 3% precision. We evaluate the oscillation main frequency as a rough approximation by $1/(2\pi(LC_2)^{1/2})$ which gives about 1600 Hz. This oscillation frequency allowed us to store large time series for data analysis. Other parameters were experimentally determined. From Chua's diode $i(V)$ characteristics linear fittings as $B_p = 1.8\,\mathrm{V}$, $m_1 = -0.758\,\mathrm{mS}$, and $m_0 = -0.409\,\mathrm{mS}$ with the significant digits limited by the fitting accuracy. Here S stands for inverse resistance unity. The resistor R, used as the control parameter, was a precision multiturn potentiometer and kept in the range of 1300 Ω to 2000 Ω.

A data acquisition (DAQ) interface with 16 bit resolution, maximum sampling rate of 200 k samples/s, and adjustable voltage range of maximal peak voltage of 10 V was applied for data storage. The Chua's circuit oscillations were measured at the x point depicted in Figure 1.1 after passing through an active buffer. Also Labview® was used to develop data acquisition software and analysis [17, 18]. A Keithley 237 voltage/current source in series with the Chua's diode was applied to obtain the $i(V)$ data. For each time series the potentiometer R was detached from the circuit for resistance measurements with a 3(1/2) digit multimeter.

Four representative attractors obtained with R as 1480 Ω, 1560 Ω, 1670 Ω, and 1792 Ω will be presented in Section 4 with the respective analyses. Particular attention will be given to the double scroll attractor.

3. Differential equation analysis

Considering a resistance free inductor, that is, $r_L = 0$, we have determined the operating points which coincide with the equilibrium points of (1.1), that is, its solution for

$$\dot{x} = \dot{y} = \dot{z} = 0. \tag{3.1}$$

Here the dot over the variables stands for time derivatives. The solutions correspond to the state space points $(-Ri_d(x), 0, i_d(x))$, which coincide with the interception of the load line with the graph of i_d in the plane $y = 0$. The load line is a straight line with slope $-1/R$ determined by the Kirchhoff's laws applied to the circuit composed by R and the Chua's diode. One of these equilibrium points will always be the origin $(0,0,0)$.

For

$$1300\,\Omega \le R < -\frac{1}{m_1} \approx 1319.26\,\Omega, \tag{3.2}$$

(1.1) presents only the equilibrium point at the origin. The origin is a saddle focus point, since the Jacobian matrix of (1.1) at $(0,0,0)$ has one negative real eigenvalue and two complex eigenvalues with positive real parts. Here $f(x,y,z) = (f_1(x,y,z), f_2(x,y,z), f_3(x,y,z))$ is defined by (1.1). For $1318.93\,\Omega < R < 1319.26\,\Omega$, the origin is a (1-2)-saddle point, that is, the Jacobian matrix $Jf(0,0,0)$ has three real eigenvalues, being one negative and two positives.

For $R = 1319.26\,\Omega$, (1.1) presents a line segment of equilibrium points. In fact, all points $(x, 0, m_1 x)$, $-B_p \le x \le B_p$, are equilibrium points of (1.1).

For

$$1319.26\,\Omega < R \le 2000\,\Omega < -\frac{1}{m_0} \approx 2444.99\,\Omega, \tag{3.3}$$

(1.1) presents three equilibrium points

$$\begin{aligned} p_0 &= (0,0,0), \\ p_1 &= \left(\frac{R(m_0 - m_1)B_p}{Rm_0 + 1}, 0, \frac{(m_1 - m_0)B_p}{Rm_0 + 1}\right), \\ p_2 &= \left(\frac{R(m_1 - m_0)B_p}{Rm_0 + 1}, 0, \frac{(m_0 - m_1)B_p}{Rm_0 + 1}\right). \end{aligned} \tag{3.4}$$

For $1319.26\,\Omega < R < 1323.93\,\Omega$, p_0 is a (2-1)-saddle point, and for $1323.93\,\Omega \le R \le 2000\,\Omega$, the equilibrium point p_0 is of saddle-focus type, since the Jacobian matrix $Jf(p_0)$ has one real positive eigenvalue λ_{00} and two complex eigenvalues, λ_{01} and λ_{02}, with negative real parts. Therefore, p_0 has a 1-dimensional unstable manifold and a 2-dimensional stable manifold. The equilibrium points p_1 and p_2 are of saddle-focus type too, but their stable manifolds are 1-dimensional and their unstable manifolds are 2-dimensional, since the Jacobian matrix $Jf(p_1) = Jf(p_2)$ has one real negative eigenvalue and two complex eigenvalues with positive real parts.

The presence of homoclinic loops connecting p_0 to itself, that is, p_0 possesses a 2-dimensional stable manifold and a 1-dimensional unstable manifold which intersect nontransversely, for some value of the parameter R, plays a fundamental role in the existence of chaos in (1.1).

The existence of a homoclinic loop at p_0 is now outlined, according to [19]. Equation (1.1) can be written in dimensionless form

$$\begin{aligned}\frac{d\overline{x}}{d\tau} &= \alpha(\overline{x}-\overline{y})+i(\overline{x}),\\ \frac{d\overline{y}}{d\tau} &= 0.1\big(\alpha(\overline{y}-\overline{x})\big)-\overline{z},\\ \frac{d\overline{z}}{d\tau} &= 3.321\overline{y},\end{aligned} \tag{3.5}$$

where

$$i(\overline{x}) = \begin{cases}\overline{x}-0.853 & \text{if } \overline{x} \le -1 \text{(region I)},\\ 1.853\overline{x} & \text{if } |\overline{x}| \le 1 \text{(region II)},\\ \overline{x}+0.853 & \text{if } \overline{x} \ge 1 \text{(region III)},\end{cases} \tag{3.6}$$

and the dimensionless variables and parameters are given by

$$\begin{gathered}\overline{x} = \frac{1}{B_p}x, \qquad \overline{y} = \frac{1}{B_p}y, \qquad \overline{z} = \frac{1}{m_0B_p}z,\\ \tau = \frac{m_0}{C_1}t, \qquad i(\overline{x}) = \frac{1}{m_0B_p}i_d(x), \qquad \alpha = \frac{1}{m_0R}.\end{gathered} \tag{3.7}$$

For $\alpha = -1.64042$ ($R = 1490.46\,\Omega$), (3.5) has the equilibrium points

$$\begin{gathered}q_0 = (0,0,0), \qquad q_1 = (-1.33193,0,-2.18493),\\ q_2 = (1.33193,0,2.18493).\end{gathered} \tag{3.8}$$

The eigenvalues of the Jacobian matrix of (3.5) at q_0 are 0.406522 and $-0.178994 \pm i0.376325$, with the respective eigenvectors

$$\begin{aligned}e_0 &= (0.716695,0.0847343,0.69222),\\ f_0 &= (-0.32928,0.0500556,-0.928716),\\ g_0 &= (0.124423,-0.105239,0).\end{aligned} \tag{3.9}$$

It follows that the unstable line at q_0 is generated by e_0 while the stable plane π_0 is generated by $\{f_0,g_0\}$. Let N_1 be the intersection of the plane $\overline{x} = 1$ and the unstable line at q_0. Thus $N_1 = (1,0.118229,0.96585)$. Let $X(\tau) = (\overline{x}(\tau),\overline{y}(\tau),\overline{z}(\tau))$ be the solution of (3.5) in the region III with the initial condition N_1. If $\tau = 8.2870398$ then $N_2 = X(8.2870398) = (1,-0.249007,2.42616)$ belongs to intersection of the plane $\overline{x} = 1$ and the stable plane π_0 since $\det[N_2,f_0,g_0] = 0$. Therefore a homoclinic loop at q_0 can be defined by the trajectory along the unstable eigenvector e_0. By symmetry of (3.5), there is another homoclinic loop at q_0 defined by the trajectory of the unstable eigenvector $-e_0$.

The chaotic nature of the Chua's (1.1) was proved by establishing the existence of a homoclinic loop of the saddle focus at the origin and by applying the Shil'nikov condition

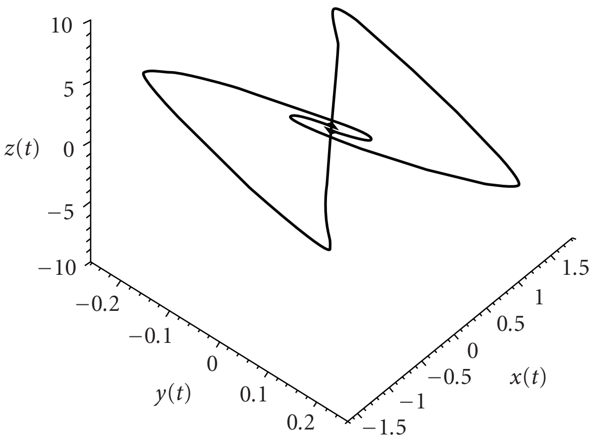


Figure 3.1. Homoclinic loops. They were obtained solving (3.5) with initial conditions $N_1 = (1, 0.118229, 0.96585)$ and $M_1 = (-1, -0.118229, -0.96585)$ and $\tau \in [-10, 30]$.

$\lambda_{00} > -\text{Re}(\lambda_{01}) > 0$ [11]. In this work, Shil'nikov saddle-focus condition is satisfied by $1334.94\,\Omega \leq R \leq 2000\,\Omega$. Figure 3.1 presents a draft of the homoclinic loop found at $\alpha = -1.64042$ corresponding to $R = 1490.46\,\Omega$.

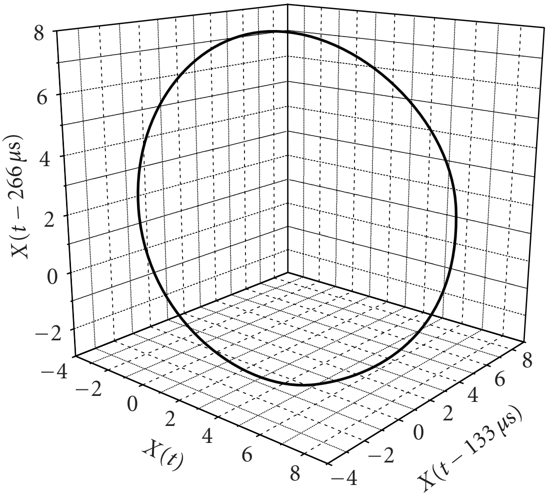
In Figure 3.1 it is possible to identify the stable and unstable manifolds associated with it. The value of R for the homoclinic loops is near of the value found for the experimental measurements of the cycle-one attractor obtained with $R = 1480\,\Omega$ as will be presented in the next section. It should be pointed out that the nominal values of capacitors and resistors used in this implementation were selected by measurements with digital multimeters which are subjected to experimental errors between 1% and 3%. Thus the value of R for the occurrence of the homoclinic loops is compatible with our experimental results.

4. Experimental results and discussion

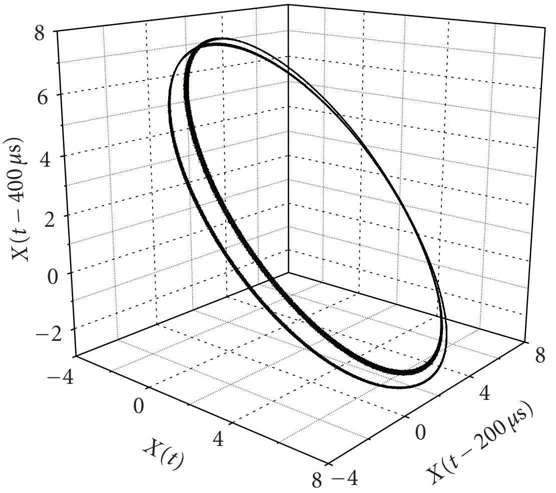
For this work we have carried out time series measurements of the variable $x(t)$ for some R values and proceeded as described in Section 2. Figures 4.1 and 4.2 present the four selected attractors obtained from time series with R as 1480 Ω, 1560 Ω, 1670 Ω, and 1792 Ω. They correspond to a cycle one, cycle two, chaotic-like in one region and the double scroll, respectively. For attractor reconstruction (see Figures 4.1 and 4.2) proper time delay [20] and the embedding dimension [21] were determined.

Figure 4.3 presents the mutual information for attractor 4. The first minimum corresponds to the optimal time delay for the delayed vectors. For the double scroll attractor it is of 7-time steps of 33 μs.

The false nearest neighbors algorithm was applied to verify if the time series is sensitive to noise [22]. Since Chua's system is a three-variable system, it turns out that false nearest neighbors should indicate the embedding dimensions as three. A higher than three embedding dimension for this system would mean significant noise contamination [17, 18].



(a) Attractor 1

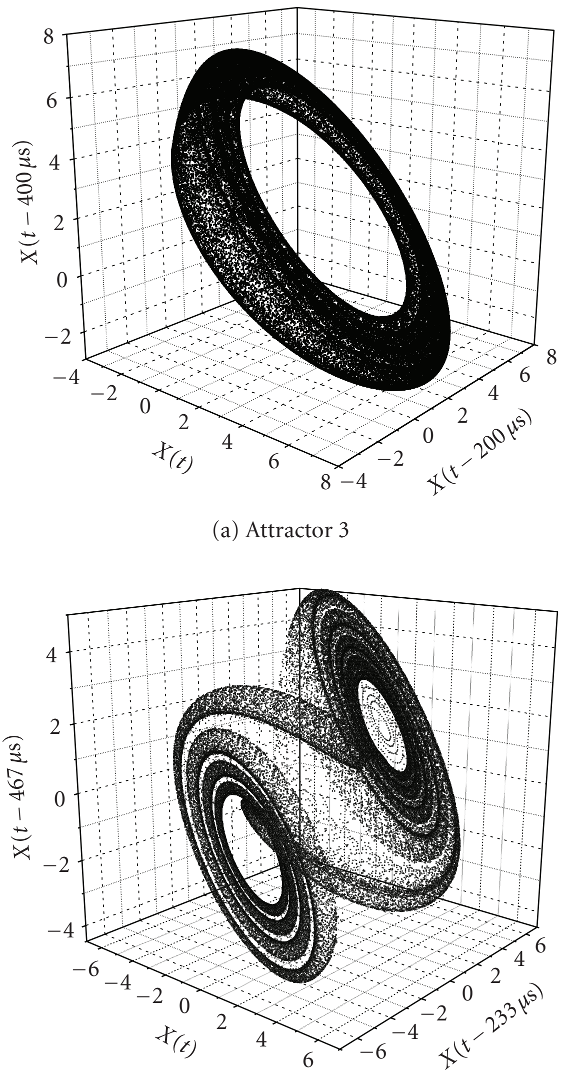


(b) Attractor 2

Figure 4.1. Periodic attractors obtained from delayed coordinates of the *x* variable. (a) Was obtained from a time series with $R = 1480\,\Omega$ and is a cycle one attractor. (b) Was obtained from a time series with $R = 1560\,\Omega$ and is a cycle two attractor.

Since our results indicate no false nearest neighbors for embedding dimensions above 3, we can neglect noise contribution for the geometric invariants that will be presented in the following.

Figure 4.4 presents the false nearest neighbor plot for attractor 4. As can be seen, the proper embedding dimension is 3. In Figure 4.5, we present the Poincaré section for the



(a) Attractor 3

(b) Attractor 4

Figure 4.2. Chaotic attractors obtained from delayed coordinates of the x variable. (a) Was obtained from a time series with $R = 1670\,\Omega$ and occupies one state space region. (b) Was obtained from a time series with $R = 1792\,\Omega$ and is the double scroll attractor.

periodic attractors presented in Figure 4.1. In Figure 4.5(a) we have a fixed point obtained from attractor 1. In Figure 4.5(b) we have a period two pair of points obtained from attractor 2.

In Figure 4.6 we present the Poincaré section for the chaotic attractors presented in Figure 4.2. In Figure 4.6(a) we have the Poincaré section for attractor 3 represented by a

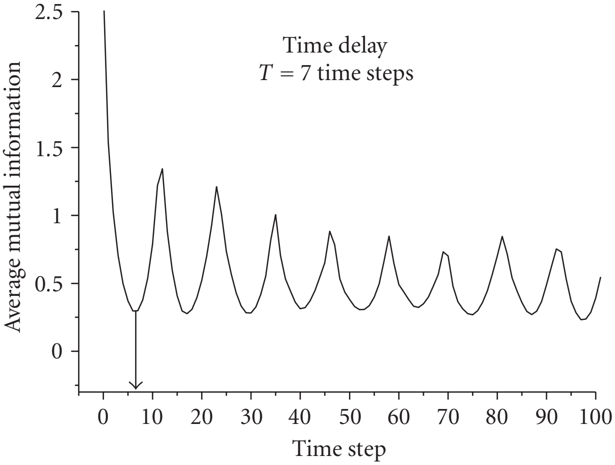


Figure 4.3. Average mutual information for the attractor 4. The first minimum corresponds to the optimal time delay.

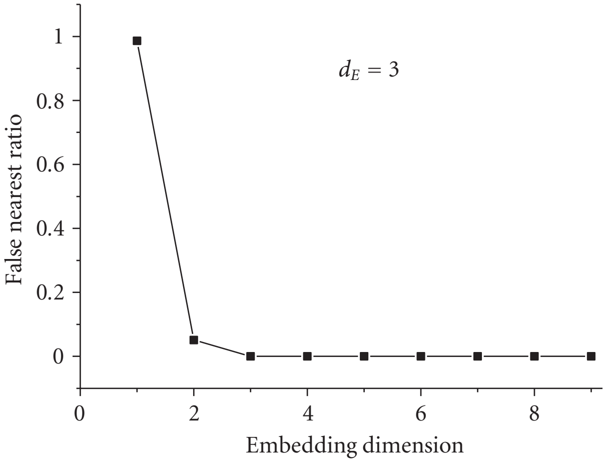
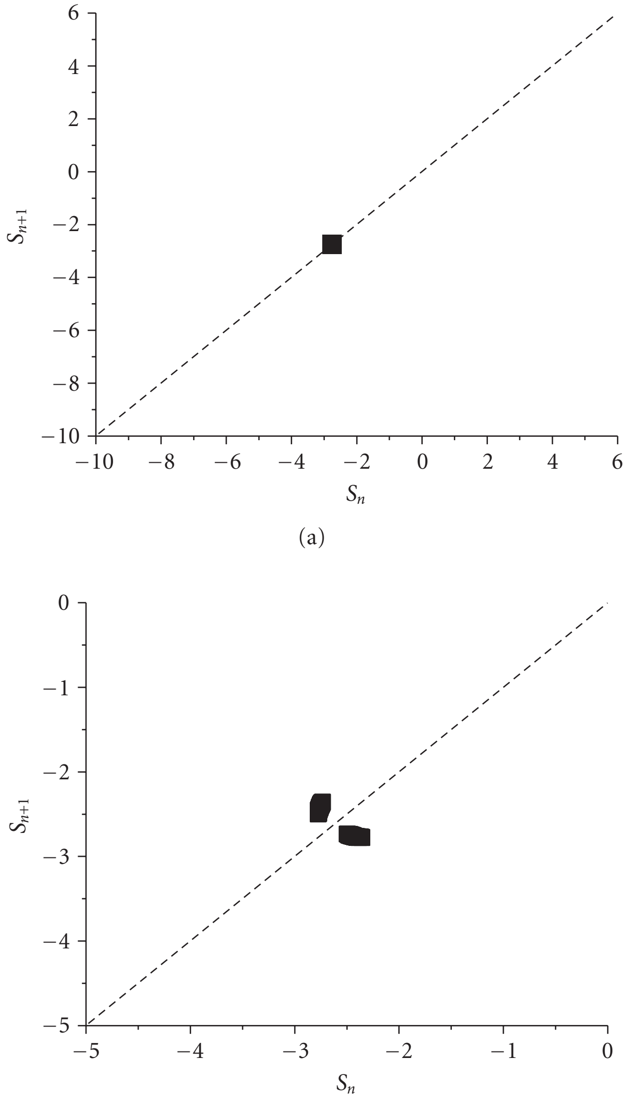


Figure 4.4. False nearest neighbor ratio as a function of the embedding dimension. The false nearest neighbors become negligible after $d_E = 3$. This confirms that the Chua's circuit is a 3-variable system.

continuous curve crossing the $y = x$ line. In Figure 4.6(b) we have a more complex pattern obtained for attractor 4. It is basically composed by two curves, one corresponding to each side of the "scroll" of the flow attractor.

Time series analyses were carried for all attractors. The estimated parameters were the Lyapunov spectrum [23] with its subproduct the Kaplan-Yorke dimension (D_{KY}) [24] and the information dimension (D_1) [25]. D_1 was measured for both flow and map representations. We will present detailed analysis for attractor 4 and summarize the information for all attractors in a table that will follow.



(a)

(b)

Figure 4.5. Poincaré section obtained from the extrema sequence of the attractors 1(a) and 2(b) presented in Figure 4.1. The dashed line is $y = x$, which shows that in (a) we have a fixed point and in (b) a period 2 points.

Figure 4.7 presents the Lyapunov spectrum for the attractor 4, obtained by using the method described in [23] and implemented in [17, 18]. It is characterized by a positive, a null, and a negative Lyapunov exponent. This configuration is a characteristic of chaotic attractors. The Kaplan-Yorke dimension for this attractor is evaluated as $D_{\mathrm{KY}} = 2.14$.

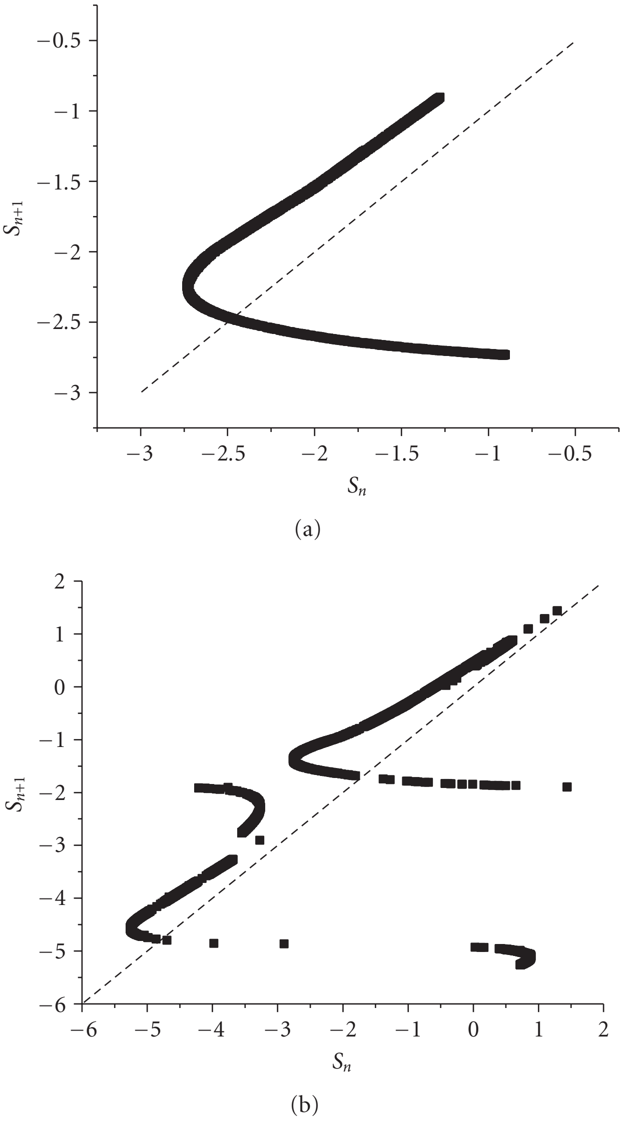


Figure 4.6. Poincaré section obtained from the extrema sequence of the attractors 3(a) and 4(b) presented in Figure 4.3. The dashed line is $y = x$, which shows that in both cases the attractors resemble chaotic.

Dimension analysis gives complementary information since it is common to find strange attractors with fractal shape.

D_{KY} is considered as equivalent to D_1 [26]. Considering this we present in Figure 4.8 the D_1 for attractor 4. In Figure 4.8(a) we present the results for the D_1 measured for the flow attractor and in Figure 4.8(b) for its Poincaré map. D_1 is characterized by a region

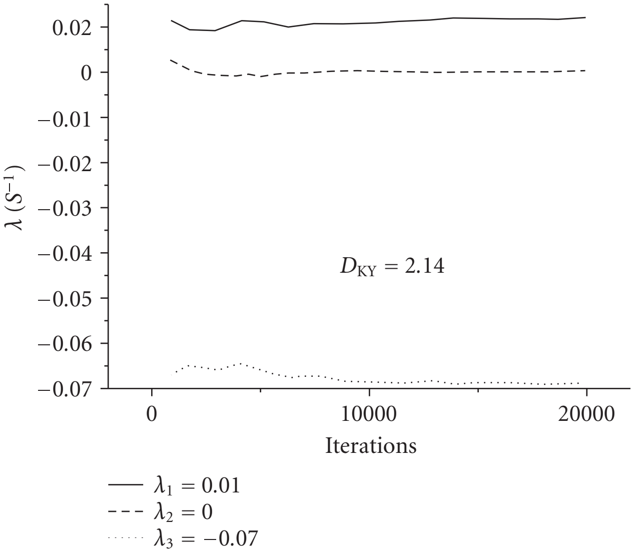


Figure 4.7. Lyapunov spectrum for attractor 4. Each Lyapunov exponent corresponds to a state space direction. The positive Lyapunov exponent is an evidence of chaotic behavior. D_{KY} is evaluated as $D_{\text{KY}} = 2.14$.

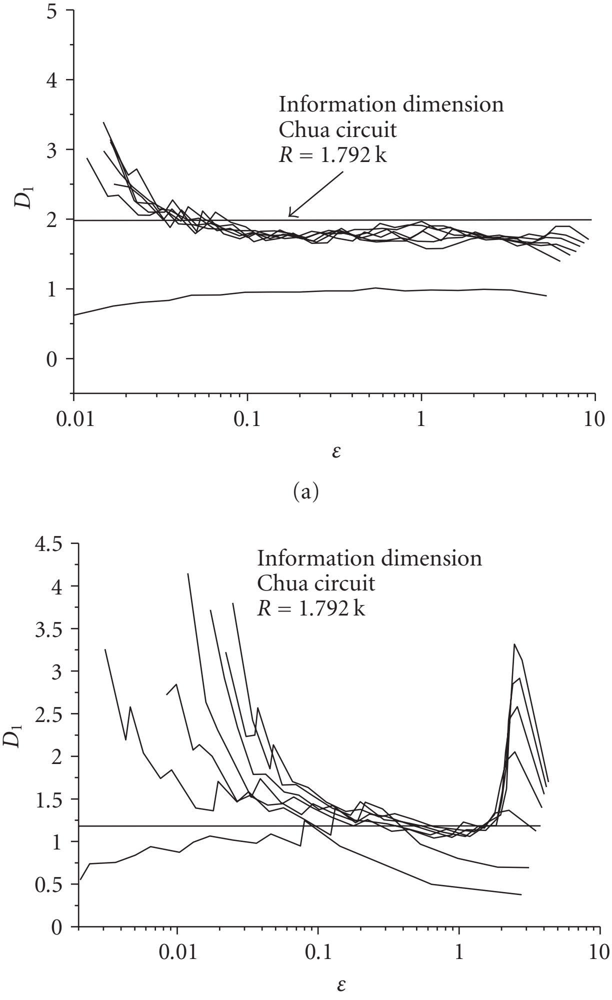
of zero slope independent of the embedding dimensions above the proper one (i.e., 3 for Chua's circuit). In Figure 4.8(a) D_1 is estimated as 1.8 ± 0.1 and in Figure 4.8(b) 1.2 ± 0.1.

According to [26] the dimension of a map attractor is related to the dimension of its flow attractor by a difference of one unity. This occurs because the map is obtained by eliminating the flow direction which is related to the null Lyapunov exponent. Since the null Lyapunov exponent is associated to a dimension of one, the map information dimension (D_{1M}) must be related to the flow information dimension (D_{1F}) by $D_{1M} = D_{1F} - 1$.

Considering that $D_{\text{KY}} \sim D_{1F}$ we can infer that our measurements of D_{1F} are underestimated and that $D_{1M} + 1$ is compatible with the corresponding values of D_{KY}. The reason for the low value of D_{1F} is yet unknown but certainly it is related to the direction of the flow and thus to the null Lyapunov exponent.

Table 4.1 summarizes the results for the four presented attractors. The first column, assigned as #, indicates the number of the attractor as defined in the text. 1, 2, 3, and 4 correspond to the attractors obtained with R in ohms defined in column 2. The third column is D_{1M}, measured for the Poincaré maps and the fourth column presents D_{1F}, measured for the flow attractor. The fifth column is the Kaplan-Yorke dimension. The sixth column is the minimal embedding dimension obtained from the false nearest neighbor algorithm. The last column lists the three Lyapunov exponents in decreasing order.

Both periodic attractors presented three negative Lyapunov exponents, but the first two can be considered as null when compared with the third value. Considering this,



(a)

(b)

Figure 4.8. Information dimension D_1 for attractor 4. In (a) we present the result for the flow attractor and in (b) for its Poincaré map. In (a) the straight line is a guide that indicates that the dimension is below 2.0. In (b) the dimension is evaluated at 1.2.

D_{KY} is estimated as 1.0 for attractors 1 and 2. This is in agreement with the value of 1.0 obtained for D_{1F}.

Attractors 3 and 4 presented one positive, one null, and one negative Lyapunov exponent. The sum of the exponents is negative, which means that attractors contract volume in state space. The Kaplan-Yorke dimension for them is above 2.0, whilst the D_1 was determined as 1.8 for the flow representation of the attractors.

Table 4.1. Analysis results for the four presented attractors.

#	$R\Omega$	D_{1M}	D_{1F}	D_{KY}	FNN	Lyap. Exp.
1	1480	0.0	1.0	1.0	2	−0.01 −0.02 −0.15
2	1560	0.0	1.0	1.0	3	−0.02 −0.02 −0.15
3	1670	1.3	1.8	2.19	3	0.01 0.00 −0.08
4	1792	1.2	1.8	2.14	3	0.01 0.00 −0.07

As discussed above the latter value is underestimated. Two facts corroborate for this assumption. One is that D_1 measured for the Poincaré maps of the attractors does not differ by one unity from the measurement carried on the flow attractors, but they do differ by approximately one unity from the Kaplan-Yorke dimension.

The other fact is also related to the Poincaré map of the attractors. The visual inspection of the Poincaré maps presented in Figure 4.6 indicates that they are objects with dimension greater than 1. Thus, the flow attractor must be an object with a dimension greater than 2, since by adding 1 to a number between 1 and 2 the resulting number must be between 2 and 3.

5. Summary

We have implemented experimentally an operational amplifier inductorless realization of the Chua's circuit.

A homoclinic loop was found by numerical analysis of normalized Chua's differential equations at a parameter corresponding to $R = 1490.46\,\Omega$. Indeed, bifurcations were observed experimentally in the vicinity of R for the homoclinic loop.

We selected four representative attractors obtained with R as 1480 Ω, 1560 Ω, 1670 Ω, and 1792 Ω to present in this work. They correspond to a cycle one, cycle two, chaotic-like in one region, and the double scroll, respectively.

Considering the double scroll, that is, for $R = 1792\,\Omega$, the information dimension of a three-dimensional delay vector reconstruction of the attractor (D_{1F}) and of its Poincaré map (D_{1M}) are 1.8 and 1.2, respectively. Also the Lyapunov spectrum gives positive, null, and negative exponents with a Kaplan-Yorke dimension as 2.14 characterizing the attractor as chaotic. This indicates that the flow attractor dimension has been underestimated and that the Kaplan-Yorke dimension is better suited for this attractor.

Acknowledgments

The authors acknowledge the Brazilian Agencies CNPq, CAPES, and FAPEMIG for financial support. R. M. Rubinger and L. F. Mello are partially supported by FAPEMIG, under the projects EDT 1968/03 and EDT 1929/03, respectively.

References

[1] L. O. Chua, “Chua’s circuit: ten years later,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E77-A, no. 11, pp. 1811–1822, 1994.

[2] G.-P. Jiang, W. K.-S. Tang, and G. Chen, “A simple global synchronization criterion for coupled chaotic systems,” *Chaos, Solitons and Fractals*, vol. 15, no. 5, pp. 925–935, 2003.

[3] J. Zhang, C. Li, H. Zhang, and J. Yu, “Chaos synchronization using single variable feedback based on backstepping method,” *Chaos, Solitons and Fractals*, vol. 21, no. 5, pp. 1183–1193, 2004.

[4] M. T. Yassen, “Adaptive control and synchronization of a modified Chua’s circuit system,” *Applied Mathematics and Computation*, vol. 135, no. 1, pp. 113–128, 2003.

[5] T. Wu and M.-S. Chen, “Chaos control of the modified Chua’s circuit system,” *Physica D: Nonlinear Phenomena*, vol. 164, no. 1-2, pp. 53–58, 2002.

[6] C.-C. Hwang, H.-Y. Chow, and Y.-K. Wang, “A new feedback control of a modified Chua’s circuit system,” *Physica D: Nonlinear Phenomena*, vol. 92, no. 1-2, pp. 95–100, 1996.

[7] K. Li, Y. C. Soh, and Z. G. Li, “Chaotic cryptosystem with high sensitivity to parameter mismatch,” *IEEE Transactions on Circuits and Systems I*, vol. 50, no. 4, pp. 579–583, 2003.

[8] Z. Li, K. Li, C. Wen, and Y. C. Soh, “A new chaotic secure communication system,” *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1306–1312, 2003.

[9] E. P. dos Santos, M. S. Baptista, and I. L. Caldas, “Dealing with final state sensitivity for synchronous communication,” *Physica A: Statistical Mechanics and Its Applications*, vol. 308, no. 1–4, pp. 101–112, 2002.

[10] L. O. Chua, “Chua’s circuit: an overview ten years later,” *Journal of Circuits Systems and Computers*, vol. 4, no. 2, pp. 117–159, 1994.

[11] L. P. Shil’nikov, “Chua’s circuit: rigorous results and future problems,” *IEEE Transactions on Circuits and Systems I*, vol. 40, no. 10, pp. 784–786, 1993.

[12] C. Letellier, G. Gouesbet, and N. F. Rulkov, “Topological analysis of chaos in equivariant electronic circuits,” *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 6, no. 12B, pp. 2531–2555, 1996.

[13] D. M. Maranhão and C. P. C. Prado, “Evolution of chaos in the Matsumoto-Chua circuit: a symbolic dynamics approach,” *Brazilian Journal of Physics*, vol. 35, no. 1, pp. 162–169, 2005.

[14] S. Kahan and A. C. Sicardi-Schifino, “Homoclinic bifurcations in Chua’s circuit,” *Physica A: Statistical Mechanics and Its Applications*, vol. 262, no. 1-2, pp. 144–152, 1999.

[15] T. Matsumoto, L. O. Chua, and K. Ayaki, “Reality of chaos in the double scroll circuit: a computer-assisted proof,” *IEEE Transactions on Circuits and Systems*, vol. 35, no. 7, pp. 909–925, 1988.

[16] L. A. B. Tôrres and L. A. Aguirre, “Inductorless Chua’s circuit,” *Electronics Letters*, vol. 36, no. 23, pp. 1915–1916, 2000.

[17] http://www.mpipks-dresden.mpg.de/~tisean/.

[18] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 1997.

[19] A. F. Gribov and A. P. Krishchenko, “Analytical conditions for the existence of a homoclinic loop in Chua circuits,” *Computational Mathematics and Modeling*, vol. 13, no. 1, pp. 75–80, 2002.

[20] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[21] 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.

[22] T. Wu and M.-S. Chen, "Chaos control of the modified Chua's circuit system," *Physica D: Nonlinear Phenomena*, vol. 164, no. 1-2, pp. 53–58, 2002.

[23] M. Sano and Y. Sawada, "Measurement of the Lyapunov spectrum from a chaotic time series," *Physical Review Letters*, vol. 55, no. 10, pp. 1082–1085, 1985.

[24] J. Kaplan and J. Yorke, "Chaotic behavior of multidimensional difference equations," in *Functional Differential Equations and Approximation of Fixed Points*, H. O. Peitgen and H. O. Walther, Eds., Springer, New York, NY, USA, 1987.

[25] R. Radii and A. Politi, "Statistical description of chaotic attractors: the dimension function," *Journal of Statistical Physics*, vol. 40, no. 5-6, pp. 725–750, 1985.

[26] J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, pp. 617–656, 1985.

R. M. Rubinger: Instituto de Ciências Exatas, Universidade Federal de Itajubá,
37500-903 Itajubá, MG, Brazil
Email address: rero@unifei.edu.br

A. W. M. Nascimento: Instituto de Ciências Exatas, Universidade Federal de Itajubá,
37500-903 Itajubá, MG, Brazil
Email address: valdisejnsantos@uol.com.br

L. F. Mello: Instituto de Ciências Exatas, Universidade Federal de Itajubá,
37500-903 Itajubá, MG, Brazil
Email address: lfmelo@unifei.edu.br

C. P. L. Rubinger: Instituto de Ciências Exatas, Universidade Federal de Itajubá,
37500-903 Itajubá, MG, Brazil
Email address: carla@fis.ua.pt

N. Manzanares Filho: Instituto de Engenharia Mecânica, Universidade Federal de Itajubá,
37500-903 Itajubá, MG, Brazil
Email address: nelson@unifei.edu.br

H. A. Albuquerque: Departamento de Física, Centro de Ciências Tecnológicas,
Universidade do Estado de Santa Catarina, 89223-100 Joinville, SC, Brazil
Email address: dfi2haa@joinville.udesc.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 86852, 10 pages
doi:10.1155/2007/86852

Research Article

Chaos Synchronization Criteria and Costs of Sinusoidally Coupled Horizontal Platform Systems

Jianping Cai, Xiaofeng Wu, and Shuhui Chen

Received 24 September 2006; Revised 12 December 2006; Accepted 11 February 2007

Recommended by José Manoel Balthazar

Some algebraic sufficient criteria for synchronizing two horizontal platform systems coupled by sinusoidal state error feedback control are derived by the Lyapunov stability theorem for linear time-varying system and Sylvester's criterion. The state variables are restricted in a subregion in order to obtain easily verified criteria. The validity of these algebraic criteria is illustrated with some numerical examples. A new concept, synchronization cost, is introduced based on a measure of the magnitude of the feedback control. The minimal synchronization cost as well as optimal coupling strength is calculated numerically. The results are meaningful in engineering application.

1. Introduction

Horizontal platform devices are widely used in offshore engineering and earthquake engineering. Mechanical model for a horizontal platform system with an accelerometer is depicted in Figure 1.1. The platform can freely rotate about the horizontal axis, which penetrates its mass center. When the platform deviates from horizon, the accelerometer will give an output signal to the torque generator, which generates a torque to inverse the rotation of the platform about rotational axis. The equation governing this system is

$$A\ddot{y} + D\dot{y} + rg\sin y - \frac{3g}{R}(B - C)\cos y \sin y = F\cos\omega t, \tag{1.1}$$

where y denotes the rotation of the platform relative to the earth, A, B, and C are respectively the inertia moment of the platform for axis 1, 2, and 3, D is the damping coefficient,

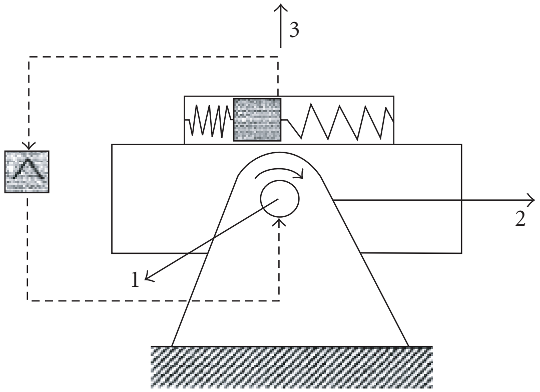


Figure 1.1. Mechanical model for a horizontal platform system with an accelerometer.

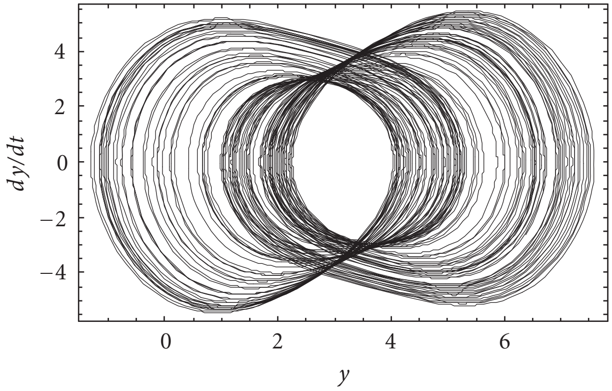


Figure 1.2. Double-scroll attractor of the horizontal platform system.

r is the proportional constant of the accelerometer, g is the acceleration constant of gravity, R is the radius of the earth, and $F\cos\omega t$ is harmonic torque. More details about this model can be found in [1, 2]. Such horizontal platform systems can reduce the swing of moving devices and keep the system close to horizontal position. They are used in modelling offshore platforms and earthquake-proof devices. As shown in Figure 1.2, the horizontal platform system has a double-scroll attractor when its parameter values are $A = 0.3$, $B = 0.5$, $C = 0.2$, $D = 0.4$, $r = 0.1155963$, $R = 6378000$, $g = 9.8$, $F = 3.4$, and $\omega = 1.8$. It was numerically verified in [1] that two identical horizontal platform systems coupled by a linear, sinusoidal, or exponential state error feedback control can achieve chaos synchronization. Analytic criteria for chaos synchronization have the advantage over numerical ones because they can reveal the relationship between the criteria and system parameters, and then they are convenient for design and analysis of the coupling controller [3–11]. Algebraic sufficient criteria for synchronizing the driving-response horizontal platform systems via linear state error feedback control were obtained in [12].

In this paper, some sufficient criteria for synchronizing the horizontal platform systems coupled by sinusoidal state error feedback control are further derived by the Lyapunov stability theory and the Sylvester's criterion. In order to obtain easily verified algebraic criteria, the state variables are restricted in a subregion, which is different from [12]. Furthermore, a new concept of synchronization cost is introduced based on a measure of the magnitude of the feedback control. The minimal synchronization cost, as well as optimal coupling strength is calculated numerically. Minimal cost means the lowest energy input, which is meaningful in engineering application.

2. Algebraic sufficient synchronization criteria

Let $x_1 = y, x_2 = \dot{y}$, and $x = (x_1, x_2)^T$, and rewrite the governing equation in form of vector

$$\dot{x} = Mx + f(x) + m(t) \tag{2.1}$$

with

$$M = \begin{pmatrix} 0 & 1 \\ 0 & -a \end{pmatrix}, \qquad f(x) = \begin{pmatrix} 0 \\ -b\sin x_1 + c\cos x_1 \sin x_1 \end{pmatrix}, \qquad m(t) = \begin{pmatrix} 0 \\ h\cos\omega t \end{pmatrix},$$
$$a = \frac{D}{A} > 0, \qquad b = \frac{rg}{A} > 0, \qquad c = \frac{3g}{RA}(B - C), \qquad h = \frac{F}{A} > 0. \tag{2.2}$$

A driving-response synchronization scheme for two identical platform systems coupled by a sinusoidal state error feedback controller is constructed as follows:

$$\text{driving system: } \dot{x} = Mx + f(x) + m(t), \tag{2.3}$$
$$\text{response system: } \dot{y} = My + f(y) + m(t) + u(t), \tag{2.4}$$
$$\text{controller: } u(t) = \left(k_1 \sin(x_1 - y_1), k_2 \sin(x_2 - y_2)\right)^T, \tag{2.5}$$

where $y = (y_1, y_2)^T$, T means transpose, and k_1 and k_2 are constant coupling coefficients. Defining an error variable $e = x - y$, or $(e_1, e_2) = (x_1 - y_1, x_2 - y_2)$, we can obtain an error dynamical system

$$\dot{e} = M(x - y) - u(t) + f(x) - f(y) = (M - K(t) + N(t))e \tag{2.6}$$

with

$$K(t) = \begin{pmatrix} k_1 s_1(t) & 0 \\ 0 & k_2 s_2(t) \end{pmatrix}, \qquad s_1(t) = \frac{\sin(x_1 - y_1)}{x_1 - y_1}, \qquad s_2(t) = \frac{\sin(x_2 - y_2)}{x_2 - y_2},$$
$$N(t) = \begin{pmatrix} 0 & 0 \\ q(t) & 0 \end{pmatrix}, \qquad q(t) = \frac{-b(\sin x_1 - \sin y_1) + c(\sin x_1 \cos x_1 - \sin y_1 \cos y_1)}{x_1 - y_1}. \tag{2.7}$$

Our object is to select suitable coupling coefficients k_1 and k_2 such that $x(t)$ and $y(t)$ satisfy

$$\lim_{t\to+\infty} ||x(t)-y(t)|| = \lim_{t\to+\infty} ||e(t)|| = 0, \tag{2.8}$$

where $\|x(t)-y(t)\| = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$ denotes the Euclidean norm of vector. By the theory of stability, chaos synchronization of systems (2.3) and (2.4) in the sense of (2.8) is equivalent to asymptotic stability of the error system (2.6) at the origin $e=0$.

Taking a quadratic Lyapunov function $V(e)=e^TPe$ with P a symmetric positive definite constant matrix, then the derivative of $V(e)$ with respect to time along the trajectory of system (2.6) is

$$\dot{V}(e) = \dot{e}^TPe + e^TP\dot{e} = e^T\left[P(M-K(t)+N(t)) + (M-K(t)+N(t))^TP\right]e. \tag{2.9}$$

By the Lyapunov stability theorem for linear time-varying system (see [13, Theorem 4.1]), a sufficient condition that the error system (2.6) is asymptotically stable at the origin is that the following matrix

$$Q(t) = P(M-K(t)+N(t)) + (M-K(t)+N(t))^TP \tag{2.10}$$

is negative definite, denoting it by

$$Q(t) < 0. \tag{2.11}$$

For simplicity, we choose $P=\operatorname{diag}\{p_1,p_2\}$ with $p_1>0$ and $p_2>0$, then

$$Q(t) = \begin{pmatrix} -2p_1k_1s_1(t) & p_1+p_2q(t) \\ p_1+p_2q(t) & -2p_2(k_2s_2(t)+a) \end{pmatrix}. \tag{2.12}$$

By the Sylvester's criterion, $Q(t)<0$ is equivalent to the following inequalities:

$$p_1k_1s_1(t) > 0, \qquad 4p_1p_2k_1s_1(t)(k_2s_2(t)+a) > (p_1+p_2q(t))^2. \tag{2.13}$$

Note that $s_1(t)>0$ and $s_2(t)>0$ if (x_1,x_2) and (y_1,y_2) are limited in the region $G=\{|x_1-y_1|<\pi, |x_2-y_2|<\pi\}$. So we conclude that under condition (2.13) the error system (2.6) is locally asymptotically stable at the origin in the region G. In order to get an easily verified algebraic condition, we further restrict the variables in the subregion $G_0=\{|x_1-y_1|\le 3\pi/4, |x_2-y_2|\le 3\pi/4\}$, then we have $2\sqrt{2}/3\pi \le s_1(t) \le 1$ and $2\sqrt{2}/3\pi \le s_2(t) \le 1$. Now, a simple algebraic sufficient criterion for synchronizing the systems (2.3) and (2.4) can be obtained from (2.13) as

$$k_1 > 0, \qquad k_2 > \frac{9\pi^2(p_1+p_2(b+|c|))^2}{32p_1p_2k_1} - a, \tag{2.14}$$

in which the inequality $|q(t)| < b+|c|$ has been used as in [12].

The synchronization criterion obtained here only renders a sufficient but not necessary condition. It is natural to expect that a sharp criterion can provide more choices of the

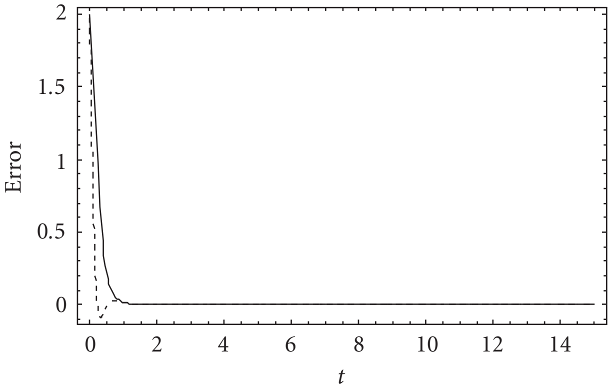


Figure 2.1. Error between the driving-response horizontal platform systems (2.3)–(2.5) with the coupling coefficients $k_1 = 5.6$ and $k_2 = 6.2$, solid curve for $x_1 - y_1$ and dashed curve for $x_2 - y_2$, initial conditions $(x_1(0), x_2(0)) = (1,1)$ and $(y_1(0), y_2(0)) = (-1,-1)$.

coupling coefficients. To this end, we can minimize the lower bound of k_2 in inequality (2.14) by choosing $p = \operatorname{diag}\{(b+|c|)p_2, p_2\}$ and obtain a sharper criterion

$$k_1 > 0, \qquad k_2 > \frac{9\pi^2(b+|c|)}{8k_1} - a. \tag{2.15}$$

Similarly, if the controller is chosen as $u(t) = (k_1 \sin(x_1 - y_1), 0)^T$, the sufficient criteria associated with inequalities (2.14) and (2.15) become, respectively,

$$k_1 > \frac{3\pi\left(p_1 + p_2(b+|c|)\right)^2}{8\sqrt{2}p_1 p_2 a}, \tag{2.16}$$

$$k_1 > \frac{3\pi(b+|c|)}{2\sqrt{2}a}. \tag{2.17}$$

The theoretical sufficient criteria are illustrated with the following examples. If we choose $p_2 = 1$ and $p_1 = (b+|c|)p_2 = 3.776615$, it is easy to verify that the coupling coefficients $k_1 = 5.6$ and $k_2 = 6.2$ satisfy inequalities (2.15). For this choice, the two coupled horizontal platform systems (2.3) and (2.4) can be asymptotically synchronized. The parameter values are chosen such that the system is in a state of chaos: $A = 0.3$, $B = 0.5$, $C = 0.2$, $D = 0.4$, $r = 0.1155963$, $R = 6378000$, $g = 9.8$, $F = 3.4$, and $\omega = 1.8$. The result is shown in Figure 2.1 with initial values $(x_1(0), x_2(0)) = (1,1)$ and $(y_1(0), y_2(0)) = (-1,-1)$, which are chosen arbitrarily in the region G_0. In this paper, software *Mathematica* is applied to implement relative calculations and plots.

For the controller $u(t) = (k_1 \sin(x_1 - y_1), 0)^T$, inequality (2.17) should be $k_1 > 9.43706$. Chaos synchronization for $k_1 = 9.5$ is illustrated in Figure 2.2, where p_1, p_2, and other parameter values are the same as above.

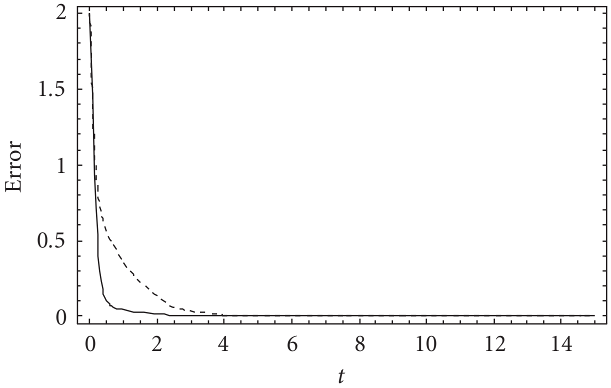


Figure 2.2. Error between the driving-response horizontal platform systems (2.3)–(2.5) with the coupling coefficients $k_1 = 9.5$ and $k_2 = 0$, solid curve for $x_1 - y_1$ and dashed curve for $x_2 - y_2$, initial conditions $(x_1(0), x_2(0)) = (1, 1)$ and $(y_1(0), y_2(0)) = (-1, -1)$.

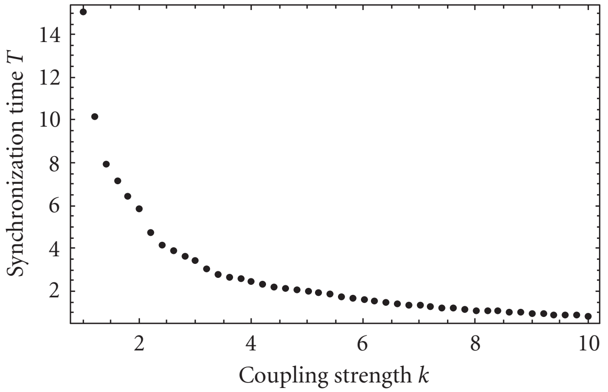


Figure 3.1. Synchronization time of systems (2.3) and (2.4) with sinusoidal controller $u(t) = (k\sin(x_1 - y_1), k\sin(x_2 - y_2))^T$, synchronization error measure $d < 0.001$, $L = 1000$, initial conditions $(x_1(0), x_2(0)) = (1,1)$ and $(y_1(0), y_2(0)) = (-1,-1)$.

3. Synchronization time and cost

Firstly, we numerically investigate the behavior of synchronization time T_{syn} as a function of coupling strength k_1 and/or k_2. The synchronization time is defined as the initial time when the error measure $d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} < \varepsilon$ is satisfied and maintains in a long enough time interval $[T_{\text{syn}}, T_{\text{syn}} + L]$, where ε is the precision of the synchronization, and L is a sufficiently large positive constant. As shown in Figures 3.1 and 3.2, the synchronization time T_{syn} gradually decreases with the increase of coupling strength, and approaches an asymptotic minimal value. This is a very interesting phenomenon,

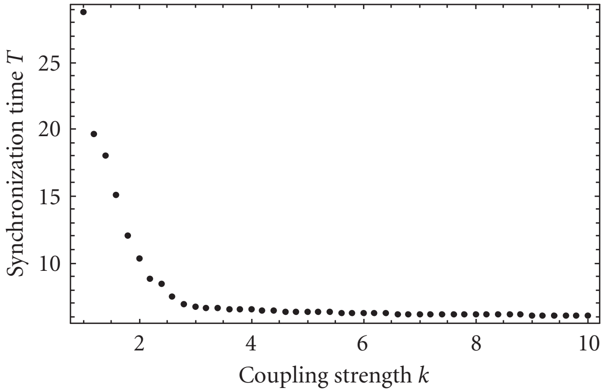


Figure 3.2. Synchronization time of systems (2.3) and (2.4) with sinusoidal controller $u(t) = (k\sin(x_1 - y_1), 0)^T$, synchronization error measure $d < 0.001$, $L = 1000$, initial conditions $(x_1(0), x_2(0)) = (1,1)$ and $(y_1(0), y_2(0)) = (-1,-1)$.

since one might think that the synchronization could be led as fast as desired if coupling strength is large enough. Figures 3.1 and 3.2 confirm that very large values of coupling strength are not necessary to ensure the synchronization with approximately the minimum T_{syn}. Such phenomenon also occurred in synchronization scheme of single-well Duffing oscillators [14]. Generally, synchronizing two chaotic systems is not cost-free. In order to evaluate what price must be paid to achieve synchronization, a new concept of synchronization cost for scheme (2.3)–(2.5) is introduced as follows:

$$\int_0^\infty k_1 \,|\sin(x_1 - y_1)|\, dt + \int_0^\infty k_2 \,|\sin(x_2 - y_2)|\, dt. \tag{3.1}$$

The meaning of this definition refers to the cost to achieve a certain degree of synchronization in the sense of (2.8). Note that the magnitude of $|x_i - y_i|$ is very small once synchronization is nearly achieved. So a good approximation of cost should be

$$\int_0^{T_{\text{syn}}} k_1 \,|\sin(x_1 - y_1)|\, dt + \int_0^{T_{\text{syn}}} k_2 \,|\sin(x_2 - y_2)|\, dt, \tag{3.2}$$

which will be adopted in the following simulations. Another definition of synchronization cost adopted in [15] for linear control is

$$\lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau k_i \,|x_i - y_i|\, dt, \quad i = 1,2, \tag{3.3}$$

which refers to the cost per unit time required to keep the synchronization going. The meaning is different from ours.

From the viewpoint of preventing from a useless increase of coupling strength, that is, from an unavailing waste of input energy, the calculation of minimal synchronization cost, as well as optimal coupling strength, is of great practical interest. Synchronization

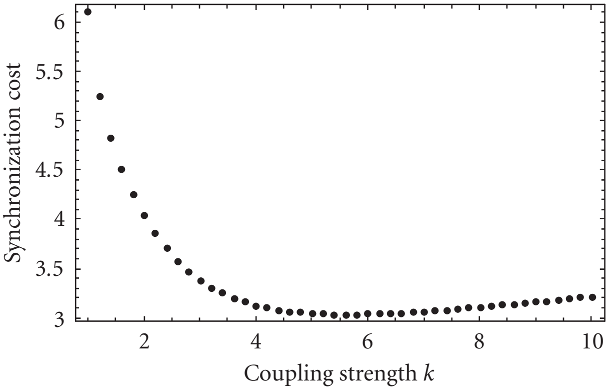


Figure 3.3. Synchronization cost of systems (2.3) and (2.4) with sinusoidal controller $u(t) = (k\sin(x_1 - y_1), k\sin(x_2 - y_2))^T$, initial conditions $(x_1(0), x_2(0)) = (1,1)$ and $(y_1(0), y_2(0)) = (-1,-1)$.

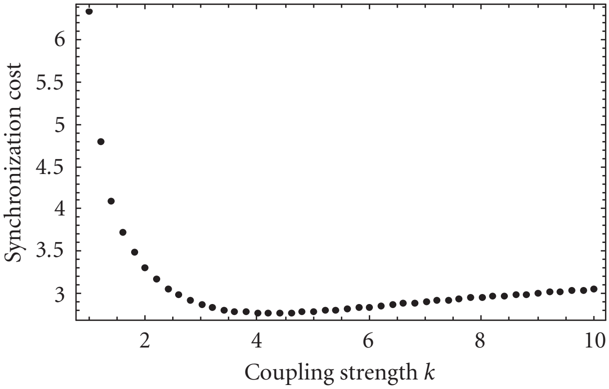


Figure 3.4. Synchronization cost of systems (2.3) and (2.4) with sinusoidal controller $u(t) = (k\sin(x_1 - y_1), 0)^T$, initial conditions $(x_1(0), x_2(0)) = (1,1)$ and $(y_1(0), y_2(0)) = (-1,-1)$.

cost versus coupling strength is simulated in Figures 3.3 and 3.4 with different controllers. From these figures we can see that the synchronization cost decreases rapidly at first, then reaches a minimal value and increases slowly with the increase of coupling strength at last. The explanation of this phenomenon is in agreement with the simulations of synchronization time shown in Figures 3.1 and 3.2. The critical coupling strength with the minimal synchronization cost can be chosen as the optimal coupled strength in the sense of consumed energy. The optimal coupling strength and minimal synchronization cost are 5.6 and 3.03922 in Figure 3.3, 4.2 and 2.77078 in Figure 3.4, respectively. Although double-variable-coupled configuration (x- and y-coupled) can lead to fast synchronization, its minimal synchronization cost is larger than that of single-variable-coupled configuration (x-coupled).

4. Conclusions

Some algebraic sufficient criteria for synchronizing driving-response horizontal platform systems coupled by sinusoidal state error feedback control are derived and their validity is illustrated with some numerical examples. Numerical simulations show that the synchronization time approaches an asymptotic minimal value with the increase of coupling strength. The concept of synchronization cost is introduced and the minimal synchronization cost as well as optimal coupling strength is calculated numerically. The minimal synchronization cost refers to the lowest-energy input, which is of great practical interest.

Acknowledgments

The authors are grateful to the anonymous referees for constructive comments. This research is supported by the National Natural Science Foundation of China under Grants no. 10672193, no. 60674049, and no. 10571184, and the Foundation of Advanced Research Center of Zhongshan University under Grant no. 06M13.

References

[1] Z.-M. Ge, T.-C. Yu, and Y.-S. Chen, “Chaos synchronization of a horizontal platform system,” *Journal of Sound and Vibration*, vol. 268, no. 4, pp. 731–749, 2003.

[2] C.-L. Huang, “Nonlinear dynamics of the horizontal platform,” M.S. thesis, National Chiao Tung University, Hsinchu, Taiwan, 1996.

[3] J. A. K. Suykens, P. F. Curran, and L. O. Chua, “Master-slave synchronization using dynamic output feedback,” *International Journal of Bifurcation and Chaos*, vol. 7, no. 3, pp. 671–679, 1997.

[4] J. Lü, T. Zhou, and S. Zhang, “Chaos synchronization between linearly coupled chaotic systems,” *Chaos, Solitons and Fractals*, vol. 14, no. 4, pp. 529–541, 2002.

[5] G.-P. Jiang, W. K.-S. Tang, and G. Chen, “A simple global synchronization criterion for coupled chaotic systems,” *Chaos, Solitons and Fractals*, vol. 15, no. 5, pp. 925–935, 2003.

[6] E. M. Elabbasy, H. N. Agiza, and M. M. El-Dessoky, “Global synchronization criterion and adaptive synchronization for new chaotic system,” *Chaos, Solitons and Fractals*, vol. 23, no. 4, pp. 1299–1309, 2005.

[7] J. Sun and Y. Zhang, “Some simple global synchronization criterions for coupled time-varying chaotic systems,” *Chaos, Solitons and Fractals*, vol. 19, no. 1, pp. 93–98, 2004.

[8] Y. Lei, W. Xu, J. Shen, and T. Fang, “Global synchronization of two parametrically excited systems using active control,” *Chaos, Solitons and Fractals*, vol. 28, no. 2, pp. 428–436, 2006.

[9] J. H. Park, “Stability criterion for synchronization of linearly coupled unified chaotic systems,” *Chaos, Solitons and Fractals*, vol. 23, no. 4, pp. 1319–1325, 2005.

[10] J.-G. Wang and Y. Zhao, “Chaotic synchronization of the master slave chaotic systems with different structures based on BANG-BANG control principle,” *Chinese Physics Letters*, vol. 22, no. 10, pp. 2508–2510, 2005.

[11] J.-H. Shen, S. Chen, and J. Cai, “Chaos synchronization criterion and its optimizations for a nonlinear transducer system via linear state error feedback control,” *Chinese Physics Letters*, vol. 23, no. 6, pp. 1406–1409, 2006.

[12] X. Wu, J. Cai, and M. Wang, “Master-slave chaos synchronization criteria for the horizontal platform systems via linear state error feedback control,” *Journal of Sound and Vibration*, vol. 295, no. 1-2, pp. 378–387, 2006.

[13] J. Slotine and W. P. Li, *Applied Nonlinear Control*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1991.

[14] Y. C. Kouomou and P. Woafo, "Stability and optimization of chaos synchronization through feedback coupling with delay," *Physics Letters, Section A*, vol. 298, no. 1, pp. 18–28, 2002.

[15] C. Sarasola, F. J. Torrealdea, A. D'Anjou, and M. Graña, "Cost of synchronizing different chaotic systems," *Mathematics and Computers in Simulation*, vol. 58, no. 4–6, pp. 309–327, 2002.

Jianping Cai: Department of Applied Mechanics and Engineering, Zhongshan University,
Guangzhou 510275, China
Email address: stscip@mail.sysu.edu.cn

Xiaofeng Wu: Center for Control and Optimization, South China University of Technology,
Guangzhou 510640, China
Email address: wuxiaof@21cn.com

Shuhui Chen: Department of Applied Mechanics and Engineering, Zhongshan University,
Guangzhou 510275, China
Email address: stscsh@mail.sysu.edu.cn

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 31267, 20 pages
doi:10.1155/2007/31267

Research Article

Stabilizability and Motion Tracking Conditions for Mechanical Nonholonomic Control Systems

Elżbieta Jarzębowska

Received 28 September 2006; Revised 5 February 2007; Accepted 17 April 2007

Recommended by José Manoel Balthazar

This paper addresses formulation of stabilizability and motion tracking conditions for mechanical systems from the point of view of constraints put on them. We present a new classification of constraints, which includes nonholonomic constraints that arise in both mechanics and control. Based on our classification we develop kinematic and dynamic control models of systems subjected to these constraints. We demonstrate that a property of being a "hard-to-control" nonholonomic system may not be related to the nature of the constraints. It may result from the formulation of control objectives for a system. We examine two control objectives which are stabilization to the target equilibrium by a continuous static state feedback control and motion tracking. Theory is illustrated with examples of control objective formulations for systems with constraints of various types.

1. Introduction

A control design project does not begin when a control engineer is handed a model of a system. It begins at the onset of the model formulation. The paper is focused on the formulation of kinematic and dynamic control models of constrained systems and a subsequent specification of control objectives for them. We provide a new classification of constraints, which is a basis for the formulation of the models. We consider nonholonomic constraints, which may be of two types: material and non-material. Equations, which specify the non-material constraints, may be differential equations of high-order with respect to time derivatives of coordinates.

Dynamic models of mechanical systems with first-order nonholonomic constraints can be developed using classical methods of analytical mechanics, for example, Lagrange's equations with multipliers and their modifications. For systems with second-order constraints Appell's equations are available [1–4]. Recently, a method of the derivation of equations of motion of systems with nonholonomic constraints of high-order has been developed. This is a generalized programmed motion equations (GPME) method [5, 6]. The high-order constraints are referred to as programmed, since they are put by a designer to specify tasks that systems have to perform or they may arise from design and control objectives [5–7]. They are non-material constraints in contrast to materials that are given by nature. Also, an equation that specifies the angular momentum conservation is meant as a non-material nonholonomic constraint [1]. Constraints that arise from underactuation in a control system are non-material nonholonomic and second-order [8, 9].

Nonholonomic control systems are a class of nonlinear control systems, which are not amenable to methods of linear control theory even locally and they are not transformable into linear control problems in any meaningful way. They require different control approaches than other nonlinear control systems due to the presence of the nonholonomic constraints. Moreover, control systems in which the high-order constraints are present require different control approaches than systems with first-order constraints [1–3, 10–12].

Nonholonomic control systems can be presented in a general form [2]:

$$\dot{x} = F(x, u), \tag{*}$$

where $x \in M$, and M is a smooth n-dimensional manifold referred to as the state space, $u \in U$, $u(t)$ is a time-dependent map from the nonnegative reals $\mathbb{R}^+$ to a constraint set $\Sigma \subset \mathbb{R}^m$, F is assumed to be C^∞ (smooth) or C^ω (analytic) and is taken from $M \times \mathbb{R}^m$ into TM such that for each fixed u, F is a vector field on M. The map u is assumed to be piecewise smooth or piecewise analytic, that is, it is admissible. There are many generalizations and specializations of this definition, for example, for Hamiltonian and Lagrangian control systems; see [2] and references there. For the scope of this paper we may consider affine nonlinear control systems in the form [2, 11]

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i, \tag{**}$$

where f is the drift vector field, g_i, $i = 1, \dots, m$, are the control vector fields, and both are smooth on M. We assume that the constraint set Σ contains an open neighborhood of the origin in $\mathbb{R}^m$.

In this paper, we make connections between the control models (**) for systems with material and non-material nonholonomic constraints, and control objectives stated for them. We demonstrate that nonholonomically constrained systems are not "hard to control" when proper control objectives and strategies are employed.

We select two control objectives, that is, stabilization (local asymptotic stabilizability) by a continuous static-state feedback strategy and motion tracking.

The system (**) is said to be LAS if there exists a feedback $u(x)$ defined on a neighborhood of 0 such that $0 \in M$ is an asymptotically stable equilibrium of the closed-loop system. A feedback controller $u(x)$ is said to be a static-state feedback when it is a continuous map $u : M \to U : x \to u(x)$, $u(0) = 0$, such that the closed-loop system (**) has a unique solution $x(t, x(0))$, $t \geq 0$, for sufficiently small initial state $x(0)$. The asymptotic stabilizability of the target equilibrium holds only if the dimension of the equilibria set including the target is equal to the number of control inputs [13, 14]. This result is equivalent to Brockett's necessary condition for feedback stabilization [2, 15]. Based on Brockett's condition, control models of nonholonomic systems are not asymptotically stabilizable even locally. However, we can still formulate control objectives for some control problems that make them LAS.

Motion tracking consists in tracking a desired motion specified by algebraic or differential equations of constraints [16, 17]. This extended definition of tracking includes trajectory tracking as a peculiar case for which a trajectory is specified by an algebraic equation. Usually in nonlinear control, motion tracking means trajectory tracking. It is achieved using two kinds of models. One considers velocities of a system as control inputs and uses a kinematic model, and ignores a system dynamics (see [2, 18, 19] and references therein). The second uses the system dynamics, where control forces and torques as well as velocities can be control inputs [10–12]. For a nonholonomic system with first-order constraints, kinematic and dynamic control models are usually integrated. A control strategy developed in such a way has a two-level architecture. The lower control level operates within the kinematic model to stabilize the system motion to a desired trajectory. The upper control level uses the dynamic model and stabilizes feedback obtained on the lower control level [12]. Trajectory tracking for nonholonomic systems with first-order constraints can also be achieved using controllers based on dynamic models in a reduced-state form [3, 12, 20]. For a control objective other than trajectory tracking these nonlinear control strategies are not applicable and new strategies have to be pursued. A model reference tracking control strategy for programmed motion is a tool developed for tracking motions specified by equations of constraints of arbitrary order [6, 7, 16, 17]. The strategy uses only dynamic models of a system, both derived by the GPME method. This strategy can also be applied to underactuated systems additionally subjected to programmed constraints [21]. For this reason we do not have to distinguish the underactuated systems as a special class of systems with second-order nonholonomic constraints as it is usually done.

Contributions of the paper consist of the presentation of the new classification of nonholonomic constraints, formulation of control models and control objectives for systems with such constraints, and design control strategies to realize these objectives.

The paper is organized as follows. In Section 2 we present the classification of constrained systems. In Sections 3 and 4 we address kinematic and dynamic control models for them. In Sections 5 and 6, based on examples, we review these models with respect to the stabilizability conditions and possibilities of motion tracking. Examples are illustrated with simulation results. The paper closes with conclusions and a list of references.

2. Classification of nonholonomic systems

Classifications of nonholonomic constraints known to the author capture material constraints and non-material that arise from the conservation law and underactuation, for example, [1, 8, 9, 12, 14, 22]. In Table 2.1 we present a new classification, which includes nonholonomic non-material high-order constraints. The high-order constraints enable specification of many tasks, control objectives, and motion requirements that are usually considered “side conditions” not the constraints. In the new classification they are treated in the same way as other constraints on systems, provided that they can be specified by algebraic or differential equations. An example is an equation of a desired trajectory, which we treat as a constraint.

We do not consider high-order constraints in biomechanical systems herein, see, for example, [23].

The most common constrained systems are these with first-order material constraints (group 1). They arise from the condition that vehicle wheels or fingers of multifinger hands grasping objects do not slip. There is a subgroup of the wheeled systems for which their wheels are not powered. These are a snakeboard [24, 25], a roller-racer [26, 27], a roller-blader [28], a roller-walker [29], or snake-like robots [30, 31]. All these systems can move their bodies due to the relative motion of their joints. This motion is referred to as snake-like motion. Control properties of systems with powered and idle wheels significantly differ.

The constraints from group 2 originate from the conservation law and have the form of nonholonomic constraint equations of first-order. They play the same role as the material constraints do, that is, they specify conditions, which system velocities have to satisfy. Usually, they are distinguished as the “conservation laws” not the constraints *per se* [1]. They arise for space vehicles and robots, for a falling cat [1], for a sportsman performing a summersault [32], and for an astronaut on a space walk [3]. Some of these systems may be underactuated; then we assign them to group 5.

Underactuated systems from group 3 are defined as systems for which the dimension of the configuration space exceeds that of the control input space. Dynamic models of underactuated systems are classified as second-order nonholonomic system models (see (2.3a) in Table 2.1). This is due to equations that represent unactuated degrees of freedom, which are second-order nonholonomic and nonintegrable in general [9].

The underactuated systems may be wheeled mobile robots, underactuated vehicles and manipulators with unactuated joints or space robots without jets or momentum wheels [8]. Sometimes specific properties of these systems are utilized to facilitate control design, for example, equipping unactuated joints with breaking mechanisms or including gravity terms make linearization of system models about equilibrium controllable (see (2.3b) in Table 2.1).

The constraints form group 4 are programmed and they are specified by (4). We assume that they are ideal constraints. Equations (4) may specify both material and non-material constraints on a system and for this reason they are referred to as a unified constraint formulation. We state the following proposition.

Proposition 2.1. *The unified constraint formulation* $B(t,q,\dot{q},\dots,q^{(p-1)})q^{(p)} + s(t,q,\dot{q},\dots,q^{(p-1)}) = 0$ *may specify both material and non-material constraints on mechanical systems.*

Proof. The proof is based upon the reasoning that the type of a constraint equation does not influence the derivation of equations of motion of a system subjected to this constraint. The only concern is the constraint order and whether it is ideal. Indeed, when

Table 2.1. Classification of nonholonomic constraints.

Kind of constraints	Systems/constraint equations	Number of degrees of freedom (m), number of control inputs (l)	LAS	Tracking
(1) First-order, material nonholonomic.	Car-like vehicles, mobile platforms with powered wheels, multifingered hands, nonholonomic manipulators, dexterous manipulation. $B_1(q,\dot{q}) = 0 \qquad (2.1)$ B_1 is a $(k \times n)$ full rank matrix, $n > k$.	$m = n - k$; $m = l$	−	+
	Wheeled vehicles with idle wheels, nonholonomic toys, snake-like robots and manipulators. Constraints have the form (2.1), $n > k$.	$m = n - k$; $m \geq l$	−	+
(2) First-order, non-material nonholonomic (conservation law).	Space vehicles and robots, sportsman, falling cat. $B_2(q)\dot{q} + b_2(q) = 0 \qquad (2.2)$ B_2 is a $(k \times n)$ full rank matrix, $n > k$	$m = n - k$; $m \geq l$	May be	+
(3) Second-order, non-material nonholonomic, (underactuated).	Manipulators, space systems, underwater vehicles. $M_{11}(q)\ddot{q}_1 + M_{12}(q)\ddot{q}_2 + C_1(q,\dot{q}) = T_1(q)\tau,$ $M_{21}(q)\ddot{q}_1 + M_{22}(q)\ddot{q}_2 + C_2(q,\dot{q}) = 0, \qquad (2.3a)$	No gravity is present: $m = n$, $m > l$	−	+
	$M_{11}(q)\ddot{q}_1 + M_{12}(q)\ddot{q}_2 + C_1(q,\dot{q}) + D_1(q) = T_1(q)\tau,$ $M_{21}(q)\ddot{q}_1 + M_{22}(q)\ddot{q}_2 + C_2(q,\dot{q}) + D_2(q) = 0, \qquad (2.3b)$	Gravity is present: $m = n$, $m > l$	+	+

Table 2.1. Continued.

(4) High-order, non-material nonholonomic (programmed).	Task specifications for any system: $B(t,q,\dot{q},\ldots,q^{(p-1)})q^{(p)} + s(t,q,\dot{q},\ldots,q^{(p-1)}) = 0, \qquad (2.4)$ B is a $(k \times n)$ full rank matrix, $n \geq k$, s is a $(k \times 1)$ vector.	$m = n - k$, $m \geq l$	May be	+
(5) Different types of constraints put on a system.	Underactuated vehicles with idle wheels, manipulators and other systems with material and programmed constraints. The unified constraint (4), $n \geq k$.	$m = n - k$, $m \geq l$	May be	+

$p = 0$ we get a position constraint, which may be a material constraint that describes, for example, a constant distance between link ends or be a programmed constraint that specifies a desired trajectory. When $p = 1$, a constraint equation is in the form (2.1) or (2.2). It can be a material constraint, a specification of the conservation law, or a programmed constraint that specifies a desired velocity. For all examples of constraints of order $p = 1$, equations of motion are generated in the same way provided that constraints are ideal. Material constraints are of orders $p = 0$ or $p = 1$ and can be presented by (2.1). Equations for the conservation law are of order $p = 1$ and are specified by (2.2). Constraint equations for $p > 1$ are of the non-material type. Two or more constraint equations, each of a different type, may be listed in (4). The constraint (4) can be used then to specify constraint equations of any order and type. □

It should be emphasized that the constraint equations which have been investigated so far in nonlinear control were mostly in the so-called Chaplygin form, they were mostly driftless and differentially flat, and could be transformed into the power or chained forms or to their extensions [2, 3, 33]. A trajectory tracking control design for such systems can be considered a solved problem, at least theoretically [2, 11, 19]. Systems with both material and programmed constraints may be, in general, non-Chaplygin and may not be transformable into any special control form [34].

For the unified constraint formulation (4) we introduce a definition.

Definition 2.2. The equations of constraints (4) are completely nonholonomic if they cannot be integrated with respect to time, that is, constraint equations of a lower-order cannot be obtained.

If we can integrate (4) $(p - 1)$ or less times, that is, we can obtain nonholonomic constraints of first-orders or orders lower than p, we say that (4) are partially integrable. If (4) can be integrated completely, we say that they are holonomic.

We assume that (4) are completely nonholonomic. Then they do not restrict positions $q(t)$ and their time derivatives up to $(p-1)$th-order. Our definition is an extension of a definition of completely nonholonomic first-order constraints [2] and completely nonholonomic second-order constraints [9]. Necessary and sufficient integrability conditions for differential constraints of arbitrary order such as (4) are formulated in [35].

The constraint equations (4) may be of different orders. From the point of view of a control strategy design they may be differentiated. For the numerical simulation the differentiated constraint equations have to be stabilized; for more details see [17].

Finally, the constraints belong to group 5 when they are of different types and also arise form underactuation in a system.

3. Kinematic control models of constrained systems

Kinematic control models of systems with the material constraints (2.1) have a form of driftless state equations [1, 2]

$$\dot{x} = \sum_{i=1}^{n-k} g_i(x)u_i, \tag{3.1}$$

where g_i, $i = 1,\dots n-k$, are control vector fields smooth on M. The vector $x \in M$, and M is a smooth n-dimensional manifold referred to as the state space, $u(t)$ is a time-dependent map from the nonnegative reals $\mathbb{R}^+$ to a constraint set $\Sigma \subset \mathbb{R}^{n-k}$, which contains an open neighborhood of the origin in $\mathbb{R}^{n-k}$. For systems from group 1 stabilizability conditions and trajectory tracking algorithms at kinematic and dynamic control levels are well established [1–3, 11, 18]. Nonholonomic systems with the constraints (2.1) are not LAS [13–15]. A trajectory tracking formulated as an asymptotic stabilization of a tracking error is LAS for them [14]. The same holds for motion tracking [6]. For some vehicles with idle wheels subjected to the constraints (2.1) no kinematic control models can be developed [26, 27].

Kinematic control models of systems with the constraints (2.2) and (4) may have the form

$$\dot{x} = f(x) + \sum_{i=1}^{n-k} g_i(x)u_i, \tag{3.2}$$

where f is the drift vector field smooth on M.

For these systems a trajectory tracking and motion tracking control formulated as an asymptotic stabilization of a tracking error is LAS. In Section 5 we show that we can select a control objective that may make (3.2) stabilizable at some equilibrium by a continuous static-state feedback.

For the unified constraints (4) we formulate the following theorem.

Theorem 3.1. *The unified constraint formulation (4) can be presented in the state space control form (3.2).*

Proof. Let us take a new p-vector $x = (x_1,\dots,x_p)$ such that $x_1 = q$, $\dot{x}_1 = x_2,\dots,\dot{x}_{p-1} = x_p$. If time t is present explicitly in (4), we reorder coordinates, assigning $x_0 = t$. With the new vector x (4) can be written as $(p-1+k)$ first-order equations

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= x_3,\\ \vdots \quad & \quad \vdots\\ \vdots \quad & \quad \vdots\\ \dot{x}_{p-1} &= x_p,\\ B(x_1,\dots,x_p)\dot{x}_p &= -s(x_1,\dots,x_p)\end{aligned} \tag{3.3a}$$

or in a matrix form

$$C(x)\dot{x} = b(x), \tag{3.3b}$$

where C is a $(p-1+k)\times p$ matrix and b is a $(p-1+k)$-dimensional vector. Let $f(x)$ be a particular solution of (3.3b) so that $C(x)f(x) = b(x)$. Let $g(x)$ be a full-rank matrix, whose column space is in the null space of $C(x)$, that is, $C(x)g(x) = 0$. Then, the solution of (3.3b) is given by $\dot{x} = f(x) + g(x)u(t)$ for any smooth vector $u(t)$. □

In the control models (3.1) or (3.2) the number of equations is less than the number of degrees of freedom of a system, to which they are related, that is, $n > k$. When constraints are programmed, we say that the program is partly specified. When the number of equations (4) and (2.1) or (2.2) is equal to the number of degrees of freedom, that is, $n = k$, a system motion is fully specified provided that the constraints are not mutually exclusive [7, 36]. In this paper, we consider partly specified programs.

4. Dynamic control models of constrained systems

Motions specified by equations of programmed constraints have to be controlled at a dynamic level. There are important reasons to formulate a motion tracking control problem at the dynamic level. The first reason, significant from the perspective of this paper, is that we consider constraints of high-order, which specify dynamic properties of systems. Secondly, this is the level at which control takes place in practice. Designing controllers at the dynamic level usually leads to significant improvements in performance and implementability, and can help in the early identification and resolution of difficulties. Finally, unmodeled dynamics, friction, and disturbances can be taken into account at that level. Also, for massive wheeled robots that operate at high speeds, dynamics-based control strategies are necessary to obtain realistic control results [19]. It is interesting to consider tracking for holonomically constrained systems in this regard; the kinematic control problem is trivial, but the dynamic control problem is still quite challenging [37]. For wheeled vehicles that perform the snake-like motion, control at the dynamic level is only possible. The reason is that we cannot determine their global motions by just the shape variations, since they do not posses a sufficient number of nonholonomic constraint equations for this [26, 27]. Dynamic control models of such systems consist of (2.3a) and (2.1). For underactuated systems dynamic control models are (2.3a) or (2.3b). For systems with the constraints (4) dynamic models can be derived by the GPME method only.

The GPME method can also be used to derive the dynamic models (2.3a) and (2.3b), and dynamic models of systems with the constraints (2.1) or (2.2). To demonstrate this, recall that dynamic control models used in control theory are mostly based on Lagrange's equations with multipliers [2, 7], that is,

$$\begin{gathered} M(q)\ddot{q}+C(q,\dot{q})+D(q)=J(q)^T\lambda+E(q)\tau, \\ J(q)\dot{q}=0, \end{gathered} \tag{4.1}$$

where q is a n-vector of generalized coordinates, $M(q)$ is a $(n\times n)$ positive definite symmetric inertia matrix, $C(q,\dot{q})$-vector of centripetal and Coriolis forces, $D(q)$-vector of gravitational forces, $E(q)$-vector of an input transformation, $J(q)$ is a full-rank $(k\times n)$ matrix of the constraint equations, $2\le n-k<n$, λ is a k-vector of Lagrange multipliers, $E(q)\tau$ is a vector of generalized forces applied to a system, and τ is an r-vector of control inputs. For control applications, the dynamic control model (4.1) has to be transformed to the reduced-state form [2, 20, 32]. The reduced-state equations characterize the control dependent motion on the constraint manifold. The reduction procedure consists in the elimination of the constraint reaction forces. To this end let $q=(q_1,q_2)$ be a partition of the configuration variables corresponding to the partitioning of the matrix function $J(q)$ as $J(q)=[J_1(q),J_2(q)]$, $\det J_1(q)\neq 0$, and $q_1\in\mathbb{R}^k$, $q_2\in\mathbb{R}^{n-k}$. The second time derivative of a vector of dependent coordinates q_1 extracted from the constraint equations and inserted into the first of (4.1) yields equations of motion decoupled into two sets, from which one is used to design a control strategy

$$\begin{gathered} M_{22}(q)\ddot{q}_2+C_{22}(q,\dot{q}_2)\dot{q}_2+D_2(q)=E_2\tau, \\ \dot{q}_1=-J_1^{-1}(q)J_2(q)\dot{q}_2, \end{gathered} \tag{4.2a}$$

and the second when one wishes to retrieve the constraint reaction forces

$$M_{12}(q)\ddot{q}_2+C_{12}(q,\dot{q}_2)\dot{q}_2+D_1(q)=E_1\tau+J_1^T\lambda. \tag{4.2b}$$

The dynamic control model (4.2a) can be written in the extended kinematic control form [2]

$$\dot{q}=g_1(q)v_1+\cdots+g_{n-k}(q)v_{n-k},\qquad i=1,\dots,n-k,\ 2\le n-k<n, \tag{4.3a}$$

$$v_i^{r_i}=u_i, \tag{4.3b}$$

where $r_i,\dots,r_m$ denote an order of time differentiation and v is the output of a linear system consisting of chains of integrators. Equations (4.3a), (4.3b) form a dynamic model, since in applications from mechanics $r_i=1$, $i=1,\dots,n-k$, controls are typically generalized forces and the model consists of the constraint (4.3a) and the equations of motion (4.3b), which reduce to $\dot{v}=u$.

The dynamic control model (4.2a) is applicable to systems with the constraints (2.1) and for trajectory tracking. A desired trajectory is specified by $q_{2p}=q_{2p}(t)$, where "p" stands for "program." It is enough then to control $q_2(t)$ and $q_1(t)$ is also controlled, since it satisfies the constraint equations. The resulting tracking is state tracking. Using the same reduced-state dynamics (4.2a), the input-output decoupling procedure can be applied for output tracking [10]. In what follows we address state tracking strategies. For systems with the constraints (4) a new tracking control strategy is designed, that is, the model reference tracking control strategy for programmed motion. All details about this strategy can be found in [5, 6, 17] and here we report it briefly. Its architecture consists of three blocks. One is a control law block with feedback and the two are dynamic models. The first one is a reference dynamic model for programmed motion. It is a constrained dynamics that incorporates effects of all constraints on a system and has the form

$$\begin{gathered} M(q)\ddot{q}+V(q,\dot{q})+D(q)=Q(t,q,\dot{q}), \\ B\left(t,q,\dot{q},\ldots,q^{(p-1)}\right)q^{(p)}+s\left(t,q,\dot{q},\ldots,q^{(p-1)}\right)=0. \end{gathered} \tag{4.4}$$

The matrix $M(q)$ is a $(n-k\times n)$ matrix, $B(t,q,\dot{q},\ldots,q^{(p-1)})$ is a full-rank $(k\times n)$ matrix. $V(q,\dot{q})$, $D(q)$, and $Q(t,q,\dot{q})$ are all $(n-k\times 1)$ vectors and they stand, respectively, for centripetal, Coriolis and friction forces, for gravitational forces, and for other external forces applied to a system. Equations (4.4) form a reference block that plans a programmed motion.

The second dynamic model in the strategy is a dynamic control model, which incorporates effects of material constraints and conservation laws only, that is,

$$\begin{gathered} M_c(q)\ddot{q}+V_c(q,\dot{q})\dot{q}+D_c(q)=E_c(q)\tau, \\ B_1(q)\dot{q}=0. \end{gathered} \tag{4.5}$$

Equations (4.5) consist of $(n-k)$ equations of motion and k equations of the constraints. They form a "plant" block in the strategy. Both models are derived by the GPME so they are in the reduced-state form. Outputs $q_{ip}(t)$, $i=1,\ldots,n$, of (4.4) are inputs to the control law τ in (4.5). We can demonstrate that (4.5) are equivalent to (4.2a).

Theorem 4.1. *The dynamic control model (4.5) is equivalent to the reduced-state dynamic control model (4.2a).*

Proof. The reduction procedure that results in (4.2a) can be accomplished in several ways [3, 20]. We start from Lagrange's equations with multipliers (4.1), which we write as

$$\begin{gathered} \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}}\right)-\frac{\partial T}{\partial q}=J^T(q)\lambda+Q(q,\dot{q}), \\ J(q)\dot{q}=0, \end{gathered} \tag{4.6}$$

where we assume that $Q(q,\dot{q})$ stands for all external forces applied to a system.

To eliminate constraint forces from (4.6) we project these equations onto the linear subspace generated by the null space of $J(q)$. Since $(J^T(q)\lambda)\cdot\delta q=0$, Lagrange's equations become

$$\left[\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}}\right)-\frac{\partial T}{\partial q}-Q\right]\cdot\delta q=0, \tag{4.7}$$

where $\delta q\in\mathbb{R}^n$ and satisfies $J(q)\delta q=0$. We partition the coordinate vector q and the $J(q)$ matrix such that $q=(q_1,q_2)\in\mathbb{R}^k\times\mathbb{R}^{n-k}$, and $J=[J_1(q)J_2(q)]$, $J_1(q)\in\mathbb{R}^{k\times k}$ is invertible.

Then the relation $\delta q_1 = -J_1^{-1}(q)J_2(q)\delta q_2$ holds. Inserting it to (4.7) we obtain

$$J_1^{-1}J_2\left[\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_1}\right) - \frac{\partial T}{\partial q_1} - Q_1\right] - \left[\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_2}\right) - \frac{\partial T}{\partial q_2} - Q_2\right] = 0. \tag{4.8}$$

Equations (4.8) are second-order differential equations in terms of q. They can be simplified by reusing the constraint equation $\dot{q}_1 = -J_1^{-1}(q)J_2(q)\dot{q}_2$ to eliminate $\dot{q}_1$ and $\ddot{q}_1$. The evolution of q_1 can be retrieved by reapplication of the constraint equations. Equations (4.8) are equivalent to Nielsen's equations in Maggi's form [7], that is,

$$J_1^{-1}J_2\left[\frac{\partial \dot{T}}{\partial \dot{q}_1} - 2\frac{\partial T}{\partial q_1} - Q_1\right] - \frac{\partial \dot{T}}{\partial \dot{q}_2} - 2\frac{\partial T}{\partial q_2} - Q_2 = 0 \tag{4.9a}$$

which are the GPME for $p = 1$, that is, they are (4.5). It is enough to show that

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_\sigma}\right) = \frac{\partial^2 T}{\partial \dot{q}_\sigma \partial t} + \sum_{\rho=1}^{n} \frac{\partial^2 T}{\partial \dot{q}_\sigma \partial q_\rho}\dot{q}_\rho + \sum_{\rho=1}^{n} \frac{\partial^2 T}{\partial \dot{q}_\sigma \partial \dot{q}_\rho}\ddot{q}_\rho, \tag{4.10a}$$

$$\dot{T} = \frac{\partial T}{\partial t} + \sum_{\rho=1}^{n} \frac{\partial T}{\partial q_\rho}\dot{q}_\rho + \sum_{\rho=1}^{n} \frac{\partial T}{\partial \dot{q}_\rho}\ddot{q}_\rho. \tag{4.10b}$$

Based on (4.10b) we have

$$\frac{\partial \dot{T}}{\partial \dot{q}_\sigma} = \frac{\partial^2 T}{\partial t \partial \dot{q}_\sigma} + \sum_{\rho=1}^{n} \frac{\partial^2 T}{\partial q_\rho \partial \dot{q}_\sigma}\dot{q}_\rho + \sum_{\rho=1}^{n} \frac{\partial^2 T}{\partial \dot{q}_\rho \partial \dot{q}_\sigma}\ddot{q}_\rho + \frac{\partial T}{\partial q_\sigma} \tag{4.11}$$

and comparing (4.10a) and (4.11) we obtain that

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_\sigma}\right) = \frac{\partial \dot{T}}{\partial \dot{q}_\sigma} - \frac{\partial T}{\partial q_\sigma}. \tag{4.12}$$

Relations (4.12) inserted into (4.8) for q_1 and q_2 yield that terms in brackets in (4.8) are equal to $(\partial \dot{T}/\partial \dot{q}_\sigma - 2(\partial T/\partial q_\sigma))$, $\sigma = 1,2$, and (4.8) are equivalent to (4.9a), that is, equivalent to the GPME for $p = 1$. □

Theorem 4.2. *There exists a static-state feedback $U(\dot{q}_1, q, u) : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ such that the dynamics (4.5) can be transformed to the state space control formulation (4.3a), (4.3b).*

Proof. First, transform (4.5) to the state space control formulation. To this end, present the constraint equation as

$$\ddot{q} = G(q)\ddot{q}_1 + \dot{G}(q)\dot{q}_1, \tag{4.13}$$

where partition of the vector q is $q = (q_1, q_2)$ and $q_1 \in \mathbb{R}^{n-k}$, $q_2 \in \mathbb{R}^k$, $m = n - k$, and q_1, q_2 are the vectors of independent and dependent coordinates, respectively. Columns of the matrix $G(q)$ span the right null space of $B_1(q)$. It is the $(n\times m)$ matrix of the form

$$G=\begin{bmatrix} I_{(m\times m)} \\ -B_{12}^{-1}(q)B_{11}(q)\end{bmatrix}, \tag{4.14}$$

where I is a $(m\times m)$ identity matrix, $B_{12}^{-1}(q)B_{11}(q)$ is a locally smooth $(k\times m)$ matrix function, and the matrix $B_1(q)$ is expressed as $B_1 = [B_{11}(q), B_{12}(q)]$, and $B_{11}(q)$ is a $k\times(n-k)$ matrix function, and $B_{12}(q)$ is a $(k\times k)$ locally nonsingular matrix function. Elimination of second-order derivatives of dependent coordinates from the first of (4.5) yields

$$\begin{gathered} M_c(q)G(q)\ddot{q}_1+[M_c(q)\dot{G}(q)+V_c(q,\dot{q})G(q)]\dot{q}_1+D_c(q)=E_c(q)\tau, \\ \dot{q}=G(q)\dot{q}_1. \end{gathered} \tag{4.15}$$

Equations (4.15) are exactly the reduced-state dynamic model of a nonholonomic system [4, 20].

Now, introduce in (4.15) a new state variable vector $x=(q,\dot{q}_1)=(x_1,x_2)$ such that $\dot{x}_1=\dot{q}=(\dot{q}_1,\dot{q}_2)$, $\dot{x}_2=\ddot{q}_1$ and $x_1\in\mathbb{R}^n$, $x_2\in\mathbb{R}^m$. Then, (4.15) takes the form

$$\begin{gathered} M_c(x_1)G(x_1)\dot{x}_2+[M_c(x_1)\dot{G}(x_1)+V_c(x_1,\dot{x}_1)G(x_1)]x_2+D_c(x_1)=E_c(x_1)\tau, \\ \dot{x}_1=G(x_1)x_2. \end{gathered} \tag{4.16}$$

Now, select for the dynamics (4.16) a static-state feedback $U(x_2,x_1,u):\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^m$ defined by the relation $M_c(x_1)G(x_1)u+[M_c(x_1)\dot{G}(x_1)+V_c(x_1,\dot{x}_1)G(x_1)]x_2+D_c(x_1)=E_c(x_1)\tau$. Application of this static-state feedback to (4.16) transforms it to the form

$$\begin{gathered} \dot{x}_1=G(x_1)x_2, \\ \dot{x}_2=u, \end{gathered} \tag{4.17}$$

which is a desired state space control formulation with $f(x)=(G(x_1),0)$ and $g(x)=(0,e_i)$, and e_i is the standard basis vector in $\mathbb{R}^{n-k}$. □

The first of (4.17) is the constraint equation. The second is the motion equation, which transforms immediately to the linear controllable dynamics [11]

$$\frac{d}{dt}\begin{bmatrix} q_1 \\ \dot{q}_1\end{bmatrix}=\begin{bmatrix} 0 & I_m \\ 0 & 0\end{bmatrix}\begin{bmatrix} q_1 \\ \dot{q}_1\end{bmatrix}+\begin{bmatrix} 0 \\ I_m\end{bmatrix}u. \tag{4.18}$$

Equations (4.17) can be transformed to the normal form equivalent to the one obtained for instance in [2]. Taking new state variables $z_1=q_1$, $z_2=q_2$, $z_3=\dot{q}_1$, which are related to x_1 and x_2 such that $\dot{x}_1=(\dot{z}_1,\dot{z}_2)$, $\dot{z}_3=\dot{x}_2$, (4.17) can be written as

$$\begin{gathered} \dot{z}_1=z_3, \\ \dot{z}_2=G^*(z_1,z_2)z_3, \\ \dot{z}_3=u. \end{gathered} \tag{4.19}$$

This form is equivalent to the control form (3.2) where $f(z) = (z_3, G^*(z_1,z_2)z_3, 0), g_i = (0,0,e_i)$ and e_i is the standard basis vector in $\mathbb{R}^{n-k}$. The matrix G^* in (4.19) is a $(k \times n-k)$ submatrix of the matrix G defined in (4.14).

We demonstrated that the dynamic control model derived with the aid of the GPME can be presented in a standard state space representation (4.17) or (4.19). This allows us to reformulate for our dynamics (4.5) all theoretical control results obtained for the classical control models [1–3, 11, 12, 19, 20, 33].

A main motivation to design the model reference tracking control strategy for programmed motion is that a variety of equations of the non-material constraints (4) disables designing a general algorithm for a tracking controller. Instead, we separate programmed constraints from material and conservation laws. All constraint equations on a system, that is, (2.1), (2.2), and (4) are merged into the reference dynamic model (4.4). Material constraints and conservation laws are merged into the dynamic control model (4.5). This separation yields that (4.5) can be derived once for a given system and different reference dynamic models (4.4) that specify different programmed motions can be plugged into (4.5) each time. Also, this separation makes motion tracking analog to trajectory tracking and enables application of controllers originally dedicated to holonomic systems. This latter property of the tracking strategy significantly increases its scope of applications.

5. Stabilizability conditions for constrained systems

The control model (3.1) is not LAS due to Brockett's condition [2, 13, 14]. For the control model (3.2) we may formulate a control objective, for which we may design a continuous static-state feedback that makes (3.2) LAS. To show this, consider a model of a free-floating space robot presented in Figure 5.1. The angular momentum conservation yields the constraint equation

$$[J+(m_1+m_2)l_1^2+m_2l_2^2]\dot{\phi}+[(m_1+m_2)l_1^2+m_2l_2^2]\dot{\theta}_1+m_2l_2^2\dot{\theta}_2+m_2l_1l_2\cos\theta_2(2\dot{\phi}+2\dot{\theta}_1+\dot{\theta}_2)=K_0 \tag{5.1}$$

which can be written as

$$B_{1\varphi}\dot{\varphi}+B_{1\theta 1}\dot{\theta}_1+B_{1\theta 2}\dot{\theta}_2=K_0, \tag{5.2}$$

where K_o is the initial angular momentum that may or may not be zero and

$$\begin{aligned} B_{1\varphi} &= J+(m_1+m_2)l_1^2+m_2l_2^2+2m_2l_1l_2\cos\theta_2, \\ B_{1\theta 1} &= (m_1+m_2)l_1^2+m_2l_2^2+2m_2l_1l_2\cos\theta_2, \\ B_{1\theta 2} &= m_2l_2^2+m_2l_1l_2\cos\theta_2. \end{aligned} \tag{5.3}$$

In (5.1) J is the inertia of the base body, and m_1, m_2-masses of links concentrated at their ends. We assume that no external forces act on the space robot model. Let us select $\dot{\varphi} = u_1$, $\dot{\theta}_2 = u_2$ as controls and introduce a state vector $x \in \mathbb{R}^3$ such that $x_1 = \varphi - \varphi_p$, $x_2 = \theta_1 - \theta_{1p}$, $x_3 = \theta_2 - \theta_{2p}$. It quantifies the error between current values $(\varphi, \theta_1, \theta_2)$ and

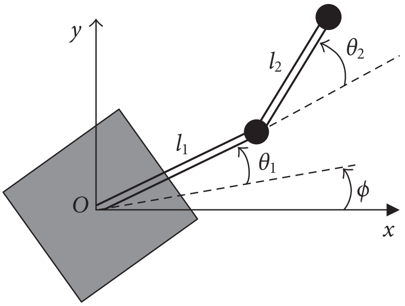


Figure 5.1. Free-floating space robot.

desired values $(\varphi_p, \theta_{1p}, \theta_{2p})$ of the coordinates. Then the control model (3.2) for the space robot becomes

$$\frac{dx}{dt} = \begin{bmatrix} 0 \\ \overline{K}_o(x) \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -\overline{K}_1(x) & -\overline{K}_2(x) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{5.4}$$

with $\overline{K}_o(x) = (K_o(x))/(B_{1\theta 1}(x))$, $\overline{K}_1(x) = (B_{1\varphi}(x))/(B_{1\theta 1}(x))$, $\overline{K}_2(x) = (B_{1\theta 2}(x))/(B_{1\theta 1}(x))$. When K_o is zero it seems natural to formulate a control objective as to asymptotically stabilize the equilibrium $x = 0$. Then the system (5.4) is driftless and the number of states $n = 3$ and $n - k = 2$. The equilibrium is not LAS. When K_o is not zero, the drift term never vanishes and $x = 0$ is not an equilibrium. It implies that asymptotic stabilization of $x = 0$ is not an appropriate control objective.

Instead, we can formulate a control problem as follows: make a system achieve $x(t_p) = 0$ for a given initial time $t = 0$ and some final time $t = t_p$. For this formulation of the control goal, we can apply the following time-varying transformation. Select $\xi \in \mathbb{R}^3$, $\xi = (\xi_1, \xi_2, \xi_3)$ such that

$$\begin{gathered} \xi_1 = x_1, \\ \xi_2 = x_2 + \overline{K}_1(0)x_1 + \overline{K}_2(0)x_3 - \overline{K}_o(0)(t - t_p), \\ \xi_3 = x_3. \end{gathered} \tag{5.5}$$

In the new coordinates the control model (5.4) has the form

$$\frac{d\xi}{dt} = \begin{bmatrix} 0 \\ \overline{K}_o(\xi) - \overline{K}_o(0) \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -[\overline{K}_1(\xi) - \overline{K}_1(0)] & -[\overline{K}_2(\xi) - \overline{K}_2(0)] \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{5.6}$$

and the system has equilibrium at $\xi = 0$. It can be verified that the dimension of the equilibrium set is 2 and $n - k = 2$. Then the system is stabilizable by continuous static-state feedbacks.

We conclude that the control problem formulation is significant as well as the structure of the system and constraints on it.

6. Motion tracking conditions for constrained systems

Tracking a desired motion becomes a control objective in a case of the programmed constraints put on a system. Trajectory tracking can be formulated as asymptotic stabilization of a tracking error and the tracking error dynamics are LAS [13, 14]. Consider an example of a system subjected to the high-order constraint (4). Take a two-link planar manipulator model presented in Figure 6.1 [16, 36]. We put a constraint on the manipulator end effector, which specifies the rate of change of the curvature $\Phi(t)$ of its trajectory. In the joint coordinates the constraint has the form

$$F_2\dddot{\Theta}_1 + \dddot{\Theta}_2 - F_1 = 0, \tag{6.1}$$

where

$$\begin{gathered}
F_1 = \frac{A_\phi - A_1 - A_2 a_o}{a_2 + a_4 a_o}, \qquad F_2 = \frac{a_1 + a_2 + a_o(a_3 + a_4)}{a_2 + a_4 a_o}, \qquad a_o = \frac{a_5}{a_6}, \\
A_\phi = \frac{-\Phi(a_5^2 + a_6^2)^2\left[\dot{\Phi}(a_5^2 + a_6^2) + 3\Phi(a_5a_7 + a_6a_8)\right]}{a_6(a_5a_8 - a_7a_6)}, \\
A_1 = 3a_3\dot{\Theta}_1\ddot{\Theta}_1 + 3a_4(\ddot{\Theta}_1 + \ddot{\Theta}_2)(\dot{\Theta}_1 + \dot{\Theta}_2) - a_1\dot{\Theta}_1^3 - a_2(\dot{\Theta}_1 + \dot{\Theta}_2)^3, \\
A_2 = 3a_3\dot{\Theta}_1\ddot{\Theta}_1 + 3a_2(\ddot{\Theta}_1 + \ddot{\Theta}_2)(\dot{\Theta}_1 + \dot{\Theta}_2) + a_3\dot{\Theta}_1^3 + a_4(\dot{\Theta}_1 + \dot{\Theta}_2)^3, \\
a_1 = -l_1\sin\Theta_1, \qquad a_3 = -l_1\cos\Theta_1, \\
a_2 = -l_2\sin(\Theta_1 + \Theta_2), \qquad a_4 = -l_2\cos(\Theta_1 + \Theta_2), \\
a_5 = a_1\dot{\Theta}_1 + a_2(\dot{\Theta}_1 + \dot{\Theta}_2), \qquad a_7 = a_1\dot{\Theta}_1 + a_3\dot{\Theta}_1^2 + a_2(\ddot{\Theta}_1 + \ddot{\Theta}_2) + a_4(\dot{\Theta}_1 + \dot{\Theta}_2)^2, \\
a_6 = -a_3\dot{\Theta}_1 - a_4(\dot{\Theta}_1 + \dot{\Theta}_2), \qquad a_8 = -a_3\ddot{\Theta}_1 + a_1\dot{\Theta}_1^2 - a_4(\ddot{\Theta}_1 + \ddot{\Theta}_2) + a_2(\dot{\Theta}_1 + \dot{\Theta}_2)^2.
\end{gathered} \tag{6.2}$$

For this constraint $n = 2$, $n - k = 1$. The kinematic control model (3.2) generated for (6.1) has a drift that does not vanish. One option is to look for one control input that can steer a system to the desired motion consistent with (6.1). The other is to apply the model reference tracking control for programmed motion based on (4.4) and (4.5). The reference dynamic model of the manipulator subjected to the third-order constraint (6.1) and developed by the GPME is

$$\begin{gathered}
(b_1 - b_2F_2)\ddot{\Theta}_1 + (b_2 - \delta F_2)\ddot{\Theta}_2 + c = 0, \\
\dddot{\Theta}_2 = F_1 - F_2\dddot{\Theta}_1,
\end{gathered} \tag{6.3}$$

where $\alpha = I_{z1} + I_{z2} + m_1r_1^2 + m_2(l_1^2 + r_2^2)$, $\beta = m_2l_1r_2$, $\delta = I_{z2} + m_2r_2^2$, $b_1 = \alpha + 2\beta\cos\Theta_2$, $b_2 = \delta + \beta\cos\Theta_2$, and $c = -\beta\dot{\Theta}_2(\dot{\Theta}_2 + 2\dot{\Theta}_1)\sin\Theta_2 - 4/3\beta\dot{\Theta}_1^2F_2\sin\Theta_2$.

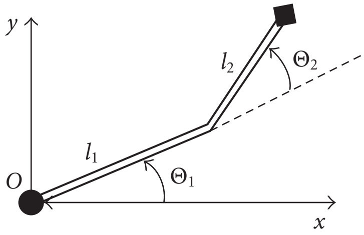


Figure 6.1. Two-link planar manipulator.

The parameters above consist of inertia and geometric data for the manipulator model. The dynamic control model of the manipulator is as follows:

$$\begin{bmatrix} \alpha+2\beta\cos\Theta_2 & \delta+\beta\cos\Theta_2 \\ \delta+\beta\cos\Theta_2 & \delta \end{bmatrix}\begin{bmatrix}\ddot{\Theta}_1\\ \ddot{\Theta}_2\end{bmatrix} + \begin{bmatrix} -\dot{\Theta}_2\beta\sin\Theta_2 & -\beta\sin\Theta_2(\dot{\Theta}_1+\dot{\Theta}_2) \\ \dot{\Theta}_1\beta\sin\Theta_2 & 0 \end{bmatrix}\begin{bmatrix}\dot{\Theta}_1\\ \dot{\Theta}_2\end{bmatrix} = \begin{bmatrix}\tau_1\\ \tau_2\end{bmatrix}, \tag{6.4}$$

since the manipulator with no programmed constraints is holonomic.

The reference dynamics (6.3) produces programmed outputs Θ_{1p}, Θ_{2p}, and their derivatives, which are inputs to the control dynamics (6.4). Two control inputs $\tau = (\tau_1, \tau_2)$ are torques, which have to be applied at manipulator joints to track the desired motion specified by (6.1). Furthermore, they can be static-state feedbacks designed in the same way as for any holonomic system, specifically for any manipulator [37, Chapter 3.4]. Indeed, when to select computed torque controllers τ_1, τ_2, and the PD controller for the outer loop, the tracking error is asymptotically stable as long as the PD controller gains are all positive. Specifically, we have

$$\tau = M_c(\Theta)u + V_c(\Theta,\dot{\Theta})\dot{\Theta}, \tag{6.5}$$

where $\Theta = (\Theta_1,\Theta_2)$, $M_c(\Theta)$, $V_c(\Theta,\dot{\Theta})$ are matrices that furnish (6.4), and u is a new input. The PD controller can be defined as $u = \ddot{\Theta}_p - 2\sigma\dot{e} - \sigma^2 e$ and a vector of a tracking error as $e(t) = \Theta(t) - \Theta_p(t)$. The tracking error satisfies the equation $\ddot{e} + 2\sigma\dot{e} + \sigma^2 e = 0$ in which σ is a convergence rate diagonal matrix. It converges to zero exponentially, that is, the end-effector motion converges to the programmed motion.

In a general case of a dynamic control model of a nonholonomic system, according to Theorem 4.2, the computed torque applied to (4.5) results in (4.17) that can be written as

$$\begin{gathered} \ddot{q}_1 = u, \\ \ddot{q}_2 = -B_{12}^{-1}(q)B_{11}(q)\ddot{q}_1 - \frac{d}{dt}\left[B_{12}^{-1}(q)B_{11}(q)\right]\dot{q}_1. \end{gathered} \tag{6.6}$$

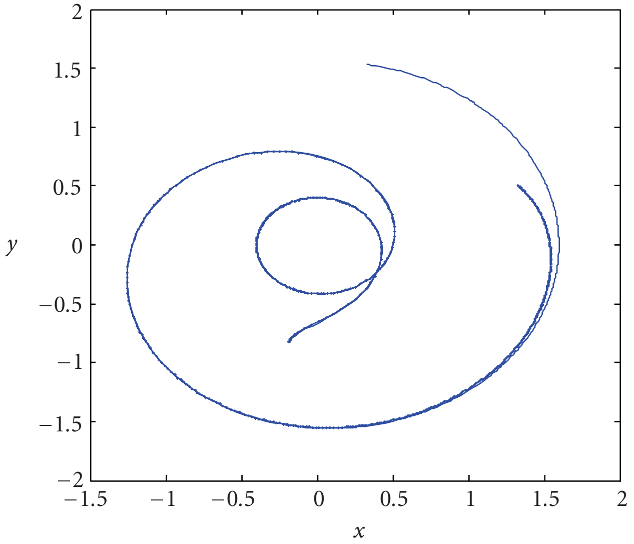


Figure 6.2. Programmed motion tracking by the PD controller.

A vector of a new input is u and it can be selected as

$$u = \ddot{q}_{1p} - 2\sigma\dot{\tilde{q}} - \sigma^2\tilde{q}, \tag{6.7}$$

where $\tilde{q} = q_1 - q_{1p}$ is a position tracking error. The tracking error satisfies the equation $\ddot{\tilde{q}} + 2\sigma\dot{\tilde{q}} + \sigma^2\tilde{q} = 0$ and converges to zero exponentially. This simple sample of a controller design illustrates the philosophy of the application of the reference dynamic model in the model reference tracking control strategy for programmed motion.

Simulation results for tracking the programmed motion specified by (6.1) by the PD controller and tracking errors are presented in Figures 6.2 and 6.3. Position and velocity errors are denoted by $e_1 = \Theta_1 - \Theta_{1p}$, $e_2 = \Theta_2 - \Theta_{2p}$, and e_3 and e_4 for the angle time derivatives, respectively.

This tracking strategy can be employed in the same way with the application of other static-state feedback controllers [17].

7. Conclusions

In this paper, we have presented the new constraint classification with respect to kinds of constraints put on mechanical systems. This classification reflects the extended constraint concept that includes non-material nonholonomic constraints of high-order. The general form of equations of constraints referred to as the unified constraint formulation follows this classification. For systems subjected to the unified high-order constraints kinematic and dynamic control models have been developed and examined from the point of view of stabilizability and motion tracking conditions. We have demonstrated that constrained systems are not "hard to control" when appropriate control objectives are formulated

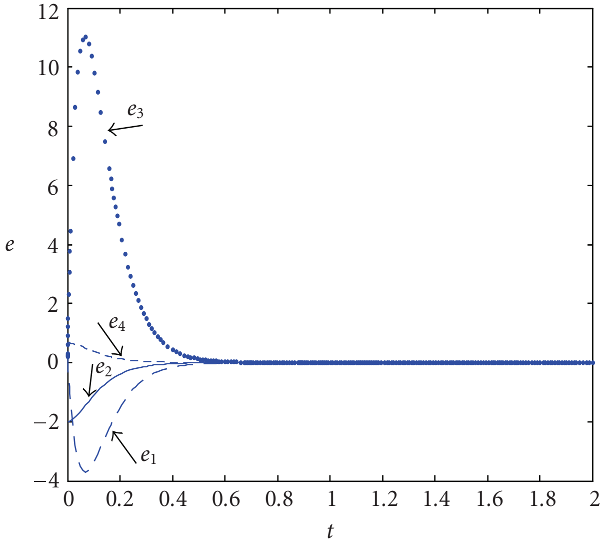


Figure 6.3. Position and velocity tracking errors versus time.

and control strategies are applied. In this paper, we applied the model reference tracking control strategy for programmed motion to track motions specified by the constraint equations.

References

[1] A. M. Bloch, P. S. Krishnaprasad, J. E. Marsden, and R. M. Murray, "Nonholonomic mechanical systems with symmetry," Tech. Rep. CIT/CDS 94-013, California Institute of Technology, Pasadena, Calif, USA, 1994.

[2] A. M. Bloch, *Nonholonomic Mechanics and Control*, vol. 24 of *Interdisciplinary Applied Mathematics*, Springer, New York, NY, USA, 2003.

[3] R. N. Murray, Z. X. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, Fla, USA, 1994.

[4] J. G. Papastavridis, *Analytical Mechanics: A Comprehensive Treatise on the Dynamics of Constrained Systems; for Engineers, Physicists, and Mathematicians*, Oxford University Press, New York, NY, USA, 2002.

[5] E. Jarzębowska, "On derivation of motion equations for systems with nonholonomic high-order program constraints," *Multibody System Dynamics*, vol. 7, no. 3, pp. 307–329, 2002.

[6] E. Jarzębowska, "Dynamics modeling of nonholonomic mechanical systems: theory and applications," *Nonlinear Analysis*, vol. 63, no. 5–7, pp. e185–e197, 2005.

[7] E. Jarzębowska, "Model-based control strategies for systems with constraints of the program type," *Communications in Nonlinear Science and Numerical Simulation*, vol. 11, no. 5, pp. 606–623, 2006.

[8] S. Martinez, J. Cortez, and F. Bullo, "Motion planning and control problems for underactuated robots," in *Control Problems in Robotics*, A. Bicchi, H. I. Christensen, and D. Prattichizzo, Eds., Springer Tracts in Advanced Robotics, pp. 59–74, Springer, New York, NY, USA, 2003.

[9] M. Reyhanoglu, A. van der Schaft, N. H. McClamroch, and I. Kolmanovsky, "Nonlinear control of a class of underactuated systems," in *Proceedings of the 35th IEEE Conference on Decision and Control*, vol. 2, pp. 1682–1687, Kobe, Japan, December 1996.

[10] B.-S. Chen, T.-S. Lee, and W.-S. Chang, "A robust H^∞ model reference tracking design for nonholonomic mechanical control systems," *International Journal of Control*, vol. 63, no. 2, pp. 283–306, 1996.

[11] A. Isidori, *Nonlinear Control Systems*, Communications and Control Engineering Series, Springer, Berlin, Germany, 2nd edition, 1989.

[12] X. Yun and N. Sarkar, "Unified formulation of robotic systems with holonomic and nonholonomic constraints," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 4, pp. 640–650, 1998.

[13] M. Ishikawa and M. Sampei, "Necessary conditions for feedback stabilizability based on qualities of equilibria set," in *Proceedings of the 37th IEEE Conference on Decision and Control (CDC '98)*, vol. 4, pp. 4600–4601, Tampa, Fla, USA, December 1998.

[14] M. Ishikawa and M. Sampei, "Classification of nonholonomic systems from mechanical and control-theoretical viewpoints," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '00)*, vol. 1, pp. 121–126, Takamatsu, Japan, October-November 2000.

[15] R. W. Brockett, "Asymptotic stability and feedback stabilization," in *Differential Geometric Control Theory (Houghton, Mich, 1982)*, R. W. Brockett, R. S. Milman, and H. J. Sussmann, Eds., vol. 27 of *Progr. Math.*, pp. 181–191, Birkhäuser, Boston, Mass, USA, 1983.

[16] E. Jarzębowska, "Robot motion planning in the presence of program constraints," in *Proceedings of the 7th IFAC Symposium on Robot Control (SYROCO '03)*, vol. 1, pp. 117–122, Elsevier, Wroclaw, Poland, September 2003.

[17] E. Jarzębowska, "Control oriented dynamic formulation for robotic systems with program constraints," *Robotica*, vol. 24, no. 1, pp. 61–73, 2006.

[18] J.-P. Laumond, Ed., *Robot Motion Planning and Control*, vol. 229 of *Lecture Notes in Control and Information Sciences*, Springer, London, UK, 1998.

[19] G. Oriolo, A. De Luca, and M. Vendittelli, "WMR control via dynamic feedback linearization: design, implementation, and experimental validation," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 6, pp. 835–852, 2002.

[20] M. Giergiel, et al., *Modeling and Control of Wheeled Robots*, PWN, Warsaw, Poland, 2002.

[21] E. Jarzębowska, "Tracking control design for underactuated constrained systems," *Robotica*, vol. 24, no. 5, pp. 591–593, 2006.

[22] W. Chung, *Nonholonomic Manipulators*, vol. 13 of *Springer Tracts in Advanced Robotics*, Springer, Berlin, Germany, 2004.

[23] N. Hogan, "An organizing principle for a class of voluntary movements," *The Journal of Neuroscience*, vol. 4, no. 11, pp. 2745–2754, 1984.

[24] F. Bullo and A. D. Lewis, "Kinematic controllability and motion planning for the snakeboard," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 3, pp. 494–498, 2003.

[25] J. Ostrowski, J. P. Desai, and V. Kumar, "Optimal gait selection for nonholonomic locomotion systems," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '97)*, vol. 1, pp. 786–791, Albuquerque, NM, USA, April 1997.

[26] E. Jarzębowska and R. Lewandowski, "Modeling and control design using the Boltzmann-Hamel equations: a roller-racer example," in *Proceedings of the 8th International IFAC Symposium on Robot Control (SYROCO '06)*, Bologna, Italy, September 2006.

[27] P. S. Krishnaprasad and D. P. Tsakiris, "Oscillations, SE(2)-snakes and motion control: a study of the roller racer," *Dynamical Systems*, vol. 16, no. 4, pp. 347–397, 2001.

[28] S. Chitta, F. W. Heger, and V. Kumar, “Design and gait control of a rollerblading robot,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA ’04)*, vol. 4, pp. 3944–3949, New Orleans, La, USA, April-May 2004.

[29] S. Hirose and H. Takeuchi, “Study on roller-walk (basic characteristics and its control),” in *Proceedings of the 13th IEEE International Conference on Robotics and Automation (ICRA ’96)*, vol. 4, pp. 3265–3270, Minneapolis, Minn, USA, April 1996.

[30] H. Date, Y. Hoshi, and M. Sampei, “Locomotion control of a snake-like robot based on dynamic manipulability,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS ’00)*, vol. 3, pp. 2236–2241, Takamatsu, Japan, October-November 2000.

[31] S. Hirose and A. Morishima, “Design and control of a mobile robot with an articulated body,” *International Journal of Robotics Research*, vol. 9, no. 2, pp. 99–114, 1990.

[32] L. S. Crawford and S. S. Sastry, “Biological motor control approaches for a planar diver,” in *Proceedings of the 34th IEEE Conference on Decision and Control (CDC ’95)*, vol. 4, pp. 3881–3886, New Orleans, La, USA, December 1995.

[33] P. Rouchon, P. Martin, and R. M. Murray, “Flat systems, equivalence and trajectory generation,” Tech. Rep. CDS 2003-008, California Institute of Technology, Pasadena, Calif, USA, 2003.

[34] E. Jarzębowska and N. H. McClamroch, “On nonlinear control of the Ishlinsky system as an example of a nonholonomic non-Chaplygin system,” in *Proceedings of the American Control Conference*, vol. 5, pp. 3249–3253, Chicago,Ill, USA, June 2000.

[35] T.-J. Tarn, M. Zhang, and A. Serrani, “New integrability conditions for differential constraints,” *Systems & Control Letters*, vol. 49, no. 5, pp. 335–345, 2003.

[36] E. Jarzębowska, “Model-based reference tracking with kinematic specifications of programmed motion,” in *Proceedings of the 11th IEEE International Conference on Methods and Models in Automation and Robotics (MMAR ’05)*, pp. 603–608, Miedzyzdroje, Poland, August-September 2005.

[37] F. L. Lewis, C. T. Abdallah, and D. M. Dawson, *Control of Robot Manipulators, Theory and Practice*, Marcel Dekker, New York, NY, USA, 2004.

Elżbieta Jarzębowska: Institute of Aircraft Engineering and Applied Mechanics,
Warsaw University of Technology, Nowowiejska 24 str., 00-665 Warsaw, Poland
Email address: elajarz@meil.pw.edu.pl

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 57238, 19 pages
doi:10.1155/2007/57238

Research Article

Stabilization and Observability of a Rotating Timoshenko Beam Model

Alexander Zuyev and Oliver Sawodny

Received 30 September 2006; Revised 12 February 2007; Accepted 13 May 2007

Recommended by José Manoel Balthazar

A control system describing the dynamics of a rotating Timoshenko beam is considered. We assume that the beam is driven by a control torque at one of its ends, and the other end carries a rigid body as a load. The model considered takes into account the longitudinal, vertical, and shear motions of the beam. For this distributed parameter system, we construct a family of Galerkin approximations based on solutions of the homogeneous Timoshenko beam equation. We derive sufficient conditions for stabilizability of such finite dimensional system. In addition, the equilibrium of the Galerkin approximation considered is proved to be stabilizable by an observer-based feedback law, and an explicit control design is proposed.

1. Introduction

Control issues for several models of flexible manipulators have been intensively studied by many authors. A particular list of references in this area can be found in monographs [1, 2]. There are two common approaches to represent the motion of such manipulators. The first approach deals with systems of rigid bodies [3], Galerkin approximations [4, 5], or finite element methods [6] to derive mathematical models with finite degrees of freedom. The second approach treats a manipulator as a distributed parameter system. The majority of publications in this distributed parameter approach are concentrated on the Euler-Bernoulli beam model (see [7], [8, Chapter 10.8], [1, Chapter 4], [2, 9, 10]).

A possible extension of the Euler-Bernoulli model was proposed by Timoshenko [11]. From the engineering viewpoint, the Timoshenko beam has an advantage of describing the effects of rotary inertia and the deflection due to shear. Control of Timoshenko beams was studied in [12–17], [1, Chapter 5.1.2]. The motion of a payload, usually attached to a real manipulator, is neglected in all these publications. In [18], a clamped beam with an end mass is proved to be stabilizable by a feedback control applied to the tip. The author of [19] addresses the development of LQR techniques and computation algorithms for beams with controlling torques applied to the hub. A limitation of these results is that a knowledge of the full infinite dimensional space is required. In [5], a hybrid system of partial and ordinary differential equations, representing the oscillations of a flexible beam, has been studied for the case when the control is the acceleration at a point. We have considered a model for the vertical motion of a beam and estimated its physical parameters from measurements of modal frequencies in [20].

It should be emphasized that, in contrast to the above publications, we study here a rotating beam that carries a payload under the action of gravity, the control torque is applied at the hub, and the longitudinal motion is taken into account. The motivation for this study is to control the motion of a real flexible-link manipulator-turntable ladder. Such a turntable ladder has been described in [3], where a dynamical model with two rigid bodies (two degrees of freedom) has been used to represent the first mode of oscillations.

This paper is organized as follows. In Section 2, we derive the motion equations for a flexible beam with a load under the action of gravity and the control torque. Section 3 contains necessary details for computing the eigenvalues and eigenfunctions of the associated Sturm-Liouville problem. By using Galerkin's method, we approximate the dynamics by a system of ODEs in Section 4. In the derivation procedure, we exploit the variational form instead of taking the standard inner product in L^2. The order of approximation may be chosen arbitrarily. In Section 5, a state feedback control which stabilizes the equilibrium of the Galerkin approximation is obtained (Theorem 5.1). In order to justify a possibility of implementation of the controller proposed, we study the observability problem in Section 6. The closed-loop system is proved to be asymptotically stable, provided that the feedback is generated by a Luenberger-type observer (Theorem 6.2). The proof of Theorem 6.2 is based on the invariance principle. The main advantage of our approach is that the control design is done explicitly; all the parameters appearing in the feedback law and dynamical observer can be effectively computed through integral moments with respect to solutions of the Sturm-Liouville problem. It is also important that no derivatives of the input signals are needed for the state estimation.

2. Description of the model

Consider a flexible beam rotating in the vertical plane around the fixed point O (see Figure 2.1).

We assume that the beam is driven by a control torque M at one of its ends (the hub at O), and the other end (point C) carries a payload of mass m.

Let l be the length of the beam. We assume that the centerline of the beam in its undeformed reference configuration occupies the segment $[0,l]$ on the Ox-axis. Consider a particle P on the centerline and denote by x its coordinate in the reference configuration. At a given time t, let $(x+s(x,t),w(x,t))$ be the coordinates of the position vector for P in the Cartesian frame Oxy. We introduce the notation $\psi(x,t)$ for the rotation angle of the cross section area at P due to bending. By taking into account the longitudinal, vertical,

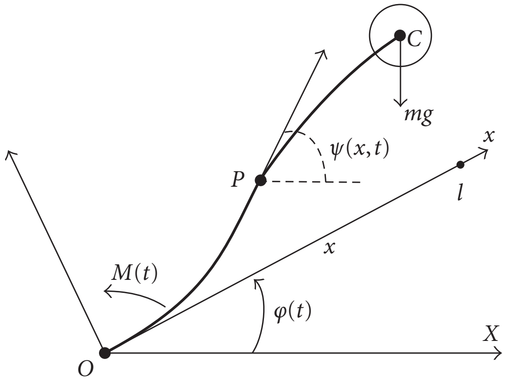


Figure 2.1. A rotating Timoshenko beam.

and shear motions, we derive the following expression for the kinetic energy of the system considered:

$$\begin{aligned}2T = &\int_0^l \{\rho(x)[(\dot{w}+x\dot{\varphi})^2+(w\dot{\varphi})^2+\dot{s}^2+2\dot{\varphi}(\dot{w}s-\dot{s}w)+(\dot{\varphi})^2(s+2x)s]+I_\rho(x)(\dot{\varphi}+\dot{\psi})^2\}\,dx\\ &+J_0(\dot{\varphi})^2+m\{(\dot{w}+x\dot{\varphi})^2+(w\dot{\varphi})^2+\dot{s}^2+2\dot{\varphi}(\dot{w}s-\dot{s}w)+(\dot{\varphi})^2(s+2x)s\}\,\big|_{x=l}\\ &+J_c\{\dot{\varphi}+\dot{\psi}\}^2|_{x=l},\end{aligned} \tag{2.1}$$

where $\varphi(t)$ is the angle between the moving axis Ox and the horizontal direction, $\rho(x)$ is the mass per unit length of the beam, $I_\rho(x)$ is the mass moment of inertia of the cross section, and J_0 is the hub moment of inertia. The mass distribution for the payload is characterized by the moment of inertia J_c with respect to its center of mass C.

In this paper, we use dots to denote derivatives with respect to time t, and primes to denote derivatives with respect to the space variable x.

Assuming that the beam is inextensible, we get the following relation on w and s:

$$s' = -\frac{1}{2}w'^2 + o(w'^2). \tag{2.2}$$

The integration of this relation, with the higher order terms being omitted, yields

$$s(x,t) = -\frac{1}{2}\int_0^x w'^2(\xi,t)\,d\xi. \tag{2.3}$$

We assume that the deformation of the beam is small and drop the terms of order higher than 2 relative to w when computing the Lagrangian of the system considered.

Following the Timoshenko beam model [11], [7, page 1142], and exploiting (2.1), (2.3), the Lagrangian takes the form

$$\begin{aligned} 2L = &\int_0^l \{\rho(x)((\dot{w}+x\dot{\varphi})^2+\dot{\varphi}^2w^2)-\rho_2(x)\dot{\varphi}^2w'^2+I_\rho(\dot{\varphi}+\dot{\psi})^2-K(\psi-w')^2-EI(\psi')^2\}\,dx \\ &+m\{\dot{\varphi}^2w^2(l,t)+(l\dot{\varphi}+\dot{w}(l,t))^2\}+J_c\{\dot{\varphi}+\dot{\psi}(l,t)\}^2+J_0\dot{\varphi}^2 \\ &-g\int_0^l \{(2\rho x-\rho_1 w'^2)\sin\varphi+2\rho w\cos\varphi\}\,dx-2mg\{l\sin\varphi+w(l,t)\cos\varphi\}, \end{aligned} \tag{2.4}$$

where

$$\rho_1(x)=\int_x^l \rho(\xi)\,d\xi+m, \qquad \rho_2(x)=\int_x^l \xi\rho(\xi)\,d\xi+ml. \tag{2.5}$$

Here E and I are Young's modulus and the moment of inertia of the cross section of the beam, respectively, g is the acceleration of gravity. The coefficient K is equal to kGA, where G is the modulus of elasticity in shear, A is the cross sectional area, and k is a constant depending on the shape of the cross section. We assume that ρ, I_ρ, EI, and K are all positive, differentiable functions of the space variable x.

If C^2-functions $(\varphi(t), w(x,t), \psi(x,t))$ define the motion of the system for the control torque $M(t)$ on a segment $t \in [t_1, t_2]$ then Hamilton's principle yields

$$\delta\left(\int_{t_1}^{t_2} L\,dt\right)+\int_{t_1}^{t_2} M(t)\delta\varphi(t)\,dt=0, \tag{2.6}$$

for any admissible variations $(\delta\varphi(t), \delta w(x,t), \delta\psi(x,t))$ satisfying the boundary conditions

$$\begin{gathered} \delta\varphi|_{t=t_1}=\delta\varphi|_{t=t_2}=0, \qquad \delta w|_{t=t_1}=\delta w|_{t=t_2}=0, \qquad \delta\psi|_{t=t_1}=\delta\psi|_{t=t_2}=0, \\ \delta w|_{x=0}=0, \qquad \delta\psi|_{x=0}=0. \end{gathered} \tag{2.7}$$

By computing the first variation in (2.6) and integrating by parts, we get

$$\int_{t_1}^{t_2} \left\{\left(M+\frac{\partial L}{\partial\varphi}-\frac{d}{dt}\frac{\partial L}{\partial\dot{\varphi}}\right)\delta\varphi(t)-\mu(\delta w(\cdot,t),\delta\psi(\cdot,t);\varphi,w,\psi)\right\}dt=0, \tag{2.8}$$

where the functional μ is linear with respect to δw and $\delta\psi$:

$$\begin{aligned} &\mu(\delta w(\cdot,t),\delta\psi(\cdot,t);\varphi,w,\psi) \\ &\quad=\int_0^l \delta w(x,t)\{(\ddot{w}+x\ddot{\varphi}-(\dot{\varphi})^2w+g\cos\varphi)\rho+(K(\psi-w')+(g\rho_1\sin\varphi-(\dot{\varphi})^2\rho_2)w')'\}\,dx \\ &\qquad+\int_0^l \delta\psi(x,t)\{I_\rho(\ddot{\psi}+\ddot{\varphi})+K(\psi-w')-(EI\psi')'\}\,dx+\delta\psi(l,t)\{J_c(\ddot{\varphi}+\ddot{\psi})+EI\psi'\}|_{x=l} \\ &\qquad+\delta w(l,t)\{K(w'-\psi)+m(\ddot{w}+l\ddot{\varphi}-(\dot{\varphi})^2w+g\cos\varphi)+m(l(\dot{\varphi})^2-g\sin\varphi)w'\}|_{x=l}. \end{aligned} \tag{2.9}$$

Thus, as (2.8) vanishes on each admissible variation satisfying (2.7), we get the following boundary value problem:

$$\begin{aligned}
&\ddot{w}+\frac{1}{\rho}\big(K(\psi-w')\big)'=-g\cos\varphi-x\ddot{\varphi}+\dot{\varphi}^2w+\frac{1}{\rho}\big((\rho_2\dot{\varphi}^2-g\rho_1\sin\varphi)w'\big)';\\
&\ddot{\psi}+\frac{K}{I_\rho}(\psi-w')-\frac{1}{I_\rho}(EI\psi')'=-\ddot{\varphi},\quad x\in(0,l);\\
&w|_{x=0}=\psi|_{x=0}=0;\\
&K(\psi-w')|_{x=l}=m\{\ddot{w}+l\ddot{\varphi}-\dot{\varphi}^2w+g\cos\varphi+(l\dot{\varphi}^2-g\sin\varphi)w'\}|_{x=l};\\
&-EI\psi'|_{x=l}=J_c(\ddot{\varphi}+\ddot{\psi}|_{x=l}),
\end{aligned}$$

$$\begin{aligned}
M(t)=\frac{d}{dt}\frac{\partial L}{\partial\dot{\varphi}}-\frac{\partial L}{\partial\varphi}&=\Big\{J_c+J_0+m[l^2+w^2(l,t)]+\int_0^l\big[I_\rho+(x^2+w^2)\rho-\rho_2w'^2\big]dx\Big\}\ddot{\varphi}\\
&+\int_0^l(\rho x\ddot{w}+I_\rho\ddot{\psi}+2\rho w\dot{\varphi}\dot{w}-2\rho_2w'\dot{\varphi}\dot{w}')dx+m(l\ddot{w}+2w\dot{\varphi}\dot{w})|_{x=l}+J_c\ddot{\psi}|_{x=l}\\
&+g\Big\{\int_0^l\Big(\rho x-\frac{1}{2}\rho_1w'^2\Big)dx+ml\Big\}\cos\varphi-g\Big\{\int_0^l\rho w\,dx+mw(l,t)\Big\}\sin\varphi.
\end{aligned}\tag{2.10}$$

Straightforward computations show that the above control system admits an equilibrium

$$\varphi(t)=\varphi_0,\qquad w(x,t)=w_0(x),\qquad \psi(x,t)=\psi_0(x),\qquad M(t)=M_0\tag{2.11}$$

if and only if the following conditions are satisfied:

$$\begin{aligned}
&\big(K(w_0'(x)-\psi_0(x))\big)'=g\{\rho\cos\varphi_0+(\rho_1w_0{}')'\sin\varphi_0\};\\
&\big(EI\psi_0'(x)\big)'+K(w_0'(x)-\psi_0(x))=0,\quad x\in(0,l);\\
&w_0(0)=\psi_0(0)=0;\qquad \psi_0'(l)=0;\qquad K(\psi_0(l)-w_0'(l))=mg(\cos\varphi_0-w_0{}'(l)\sin\varphi_0);\\
&\frac{M_0}{g}=\Big(\int_0^l\Big(\rho x-\frac{1}{2}\rho_1w_0'^2\Big)dx+ml\Big)\cos\varphi_0-\Big(\int_0^l\rho w_0\,dx+mw_0(l)\Big)\sin\varphi_0.
\end{aligned}\tag{2.12}$$

Our goal is to control the system (2.10) around its steady state (2.12).

3. Perturbed dynamics

Let (φ_0,w_0,ψ_0) be a solution of (2.12) with some M_0. Then plugging

$$\varphi=\varphi_0+\tilde{\varphi},\qquad w=w_0+\tilde{w},\qquad \psi=\psi_0+\tilde{\psi},\qquad M=M_0+\widetilde{M}\tag{3.1}$$

into the dynamical equations (2.10) yields the following control system:

$$\begin{gathered}
\ddot{\widetilde{\varphi}} = v; \\
\ddot{\widetilde{w}} + \frac{1}{\rho}\big(K(\widetilde{\psi} - \widetilde{w}')\big)' = -xv + g\widetilde{\varphi}\sin\varphi_0 - \frac{g}{\rho}\big((\widetilde{w}'\sin\varphi_0 + \widetilde{\varphi}w_0'\cos\varphi_0)\rho_1\big)' + \cdots; \\
I_\rho\ddot{\widetilde{\psi}} + K(\widetilde{\psi} - \widetilde{w}') - (EI\widetilde{\psi}')' = -I_\rho v; \\
\widetilde{w}|_{x=0} = \widetilde{\psi}|_{x=0} = 0; \\
\left(\frac{K}{m}(\widetilde{w}' - \widetilde{\psi}) + \ddot{\widetilde{w}} - g(\sin\varphi_0 + w_0'\cos\varphi_0)\widetilde{\varphi} - g\widetilde{w}'\sin\varphi_0 + \cdots\right)\Big|_{x=l} = -lv; \\
(EI\widetilde{\psi}' + J_c\ddot{\widetilde{\psi}})|_{x=l} = -J_c v,
\end{gathered} \tag{3.2}$$

where

$$\begin{aligned}
v = {}& \left(J_0 + \int_0^l \big[(w_0)^2\rho - (w_0')^2\rho_2\big]\,dx + mw_0^2(l)\right)^{-1} \\
&\times\Big\{\widetilde{M} + g\left(\int_0^l \Big[(\rho w_0 - \rho_1 w_0')\cos\varphi_0 - \frac{1}{2}\rho_1(w_0')^2\sin\varphi_0\Big]dx + mw_0(l)\cos\varphi_0\right)\widetilde{\varphi} \\
&\quad + g\int_0^l \big[(w_0'\cos\varphi_0 - \sin\varphi_0)\rho_1\widetilde{w}' + \rho\widetilde{w}\sin\varphi_0\big]dx + mg\widetilde{w}(l,t) - EI\widetilde{\psi}'(0,t)\Big\} + \cdots,
\end{aligned} \tag{3.3}$$

where the symbol "$\cdots$" denotes terms of order of smallness 2 or higher with respect to $\widetilde{\varphi}$, $\widetilde{w}$, $\widetilde{\psi}$ and their derivatives.

As, for each state $(\widetilde{\varphi}(t), \dot{\widetilde{\varphi}}(t), \widetilde{w}(\cdot,t), \dot{\widetilde{w}}(\cdot,t), \widetilde{\psi}(\cdot,t), \dot{\widetilde{\psi}}(\cdot,t))$, there is a one-to-one correspondence between $\widetilde{M}$ and v, we may treat v as a control in (3.2) and assume that it may take any value in $\mathbb{R}$.

3.1. Separation of variables. To derive a finite dimensional approximation, let us first study solutions of the control system (3.2) of a particular form

$$\widetilde{\varphi}(t) \equiv 0, \qquad \widetilde{w}(x,t) = \overline{w}(x)q(t), \qquad \widetilde{\psi}(x,t) = \overline{\psi}(x)q(t). \tag{3.4}$$

By substituting the above relations into (3.2), we get $\ddot{q}(t) = -\lambda q(t)$ together with the following Sturm-Liouville problem:

$$\begin{gathered}
\big(K(\overline{\psi} - \overline{w}') + g\rho_1\overline{w}'\sin\varphi_0\big)' - \lambda\rho\overline{w} = 0, \\
K(\overline{\psi} - \overline{w}') - (EI\overline{\psi}')' - \lambda I_\rho\overline{\psi} = 0, \quad x \in (0,l), \\
\overline{w}(0) = 0, \qquad \overline{\psi}(0) = 0, \\
K(\overline{w}'(l) - \overline{\psi}(l)) - mg\overline{w}'(l)\sin\varphi_0 - m\lambda\overline{w}(l) = 0, \\
EI\overline{\psi}'(l) - \lambda J_c\overline{\psi}(l) = 0,
\end{gathered} \tag{3.5}$$

where λ is a scalar parameter.

3.2. Eigenvalues of the Sturm-Liouville problem. Let

$$\mathcal{H} = \left\{ \begin{pmatrix} \overline{w} \\ \overline{\psi} \end{pmatrix} : \overline{w} \in H^1[0,l],\ \overline{\psi} \in H^1[0,l],\ \overline{w}(0) = \overline{\psi}(0) = 0 \right\}, \tag{3.6}$$

where $H^1[0,l]$ is the Sobolev space. Consider the following symmetric positive definite bilinear form on $\mathcal{H}$:

$$\left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}} = \int_0^l (\rho w_1 w_2 + I_\rho \psi_1 \psi_2)\, dx + m w_1(l) w_2(l) + J_c \psi_1(l) \psi_2(l). \tag{3.7}$$

A straightforward consequence of the above definition is the following.

Lemma 3.1. *Let (λ_1, w_1, ψ_1) and (λ_2, w_2, ψ_2) be nontrivial solutions of (3.5). Then*

$$\left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}} = 0 \quad \textit{if}\, \lambda_1 \neq \lambda_2. \tag{3.8}$$

Moreover, if $K(x) = \text{const}$ and

$$\frac{2(m + \int_0^l \rho\, dx) g \sin\varphi_0}{K} \leq 1, \qquad \frac{Kl^2}{EI} \leq 2 \tag{3.9}$$

then all eigenvalues λ of (3.5) are nonnegative real numbers.

Proof. If (λ_1, w_1, ψ_1) is a solution of (3.5) then

$$\begin{aligned} \lambda_1 \left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}} &= \left\langle \begin{matrix} \lambda_1 w_1 \\ \lambda_1 \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}} \\ &= \int_0^l \left(K(\psi_1 - w_1') + g\rho_1 w_1' \sin\varphi_0\right)' w_2\, dx + \int_0^l \left(K(\psi_1 - w_1') - (EI\psi_1')'\right) \psi_2\, dx \\ &\quad + \left(K(w_1'(l) - \psi_1(l)) - mg w_1'(l) \sin\varphi_0\right) w_2(l) + EI\psi_1'(l)\psi_2(l). \end{aligned} \tag{3.10}$$

Performing integration by parts in the above expression, we get

$$\lambda_1 \left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}} = \int_0^l \left\{K(w_1' w_2' + \psi_1 \psi_2 - w_1' \psi_2 - w_2' \psi_1) + EI\psi_1' \psi_2' - g\rho_1 w_1' w_2' \sin\varphi_0\right\} dx. \tag{3.11}$$

The permutation of arguments in (3.11) yields

$$\lambda_2 \left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}} = \lambda_2 \left\langle \begin{matrix} w_2 \\ \psi_2 \end{matrix}, \begin{matrix} w_1 \\ \psi_1 \end{matrix} \right\rangle_{\mathcal{H}} = \lambda_1 \left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_2 \\ \psi_2 \end{matrix} \right\rangle_{\mathcal{H}}. \tag{3.12}$$

Hence, $\left\langle \begin{smallmatrix} w_1 \\ \psi_1 \end{smallmatrix}, \begin{smallmatrix} w_2 \\ \psi_2 \end{smallmatrix} \right\rangle_{\mathcal{H}} = 0$ if $\lambda_1 \neq \lambda_2$. If $w_2 = w_1$ and $\psi_2 = \psi_1$ then (3.11) implies

$$\begin{aligned}
\lambda_1 \left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_1 \\ \psi_1 \end{matrix} \right\rangle_{\mathcal{H}} &= \int_0^l \left(K(w_1' - \psi_1)^2 + EI\psi_1'^2 - g\rho_1 w_1'^2 \sin\varphi_0\right)dx \\
&= \int_0^l \left(\frac{K}{2}w_1'^2 + EI\psi_1'^2 - K\psi_1^2 - g\rho_1 w_1'^2 \sin\varphi_0\right)dx + \frac{1}{2}\int_0^l K(w_1' - 2\psi_1)^2\, dx \\
&\geq \int_0^l \left(\left(\frac{K}{2} - g\rho_1 \sin\varphi_0\right)w_1'^2 + EI\psi_1'^2 - K{\psi_1}^2\right)dx.
\end{aligned} \tag{3.13}$$

The function $\psi_1(x)$ subject to the boundary condition $\psi_1(0) = 0$ satisfies Friedrichs' inequality of the following form (cf. [13, page 440]):

$$\int_0^l {\psi_1}^2(x)dx \leq \frac{l^2}{2}\int_0^l \psi_1'^2(x)dx. \tag{3.14}$$

Using this inequality in (3.13), we conclude that

$$\lambda_1 \left\langle \begin{matrix} w_1 \\ \psi_1 \end{matrix}, \begin{matrix} w_1 \\ \psi_1 \end{matrix} \right\rangle_{\mathcal{H}} \geq \int_0^l \left(\left(\frac{K}{2} - g\rho_1 \sin\varphi_0\right)w_1'^2 + \left(EI - \frac{Kl^2}{2}\right)\psi_1'^2\right)dx \geq 0, \tag{3.15}$$

provided that the conditions (3.9) are satisfied. This proves that all eigenvalues λ are nonnegative. □

For the rest of this section we assume that EI, I_ρ, K, and ρ are constants, and that $\sin\varphi_0 = 0$. The coefficients of the Sturm-Liouville problem are constant under this assumption, and, therefore, it is easy to find the general solution of the corresponding system of ODEs. This solution is needed for computing the coefficients of an approximate dynamical model in the sequel (formulae (4.5) define coefficients of the approximate system (4.4) through eigenvalues and eigenfunctions of (3.5)).

We introduce in (3.5) the following dimensionless functions:

$$\zeta\left(\frac{x}{l}\right) = \frac{\overline{w}(x)}{l}, \qquad \theta\left(\frac{x}{l}\right) = \overline{\psi}(x), \tag{3.16}$$

and parameters:

$$p_1 = \frac{\rho l^2}{K}, \qquad p_2 = \frac{Kl^2}{EI}, \qquad p_3 = \frac{I_\rho l^2}{EI}, \qquad p_4 = \frac{ml}{K}, \qquad p_5 = \frac{lJ_c}{EI}. \tag{3.17}$$

Then (3.5) is reduced to the following problem:

$$\frac{d}{d\tau}\begin{pmatrix}\zeta(\tau)\\ \zeta_\tau(\tau)\\ \theta(\tau)\\ \theta_\tau(\tau)\end{pmatrix}=\begin{pmatrix}0 & 1 & 0 & 0\\ -\lambda p_1 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ 0 & -p_2 & p_2-\lambda p_3 & 0\end{pmatrix}\times\begin{pmatrix}\zeta(\tau)\\ \zeta_\tau(\tau)\\ \theta(\tau)\\ \theta_\tau(\tau)\end{pmatrix},\quad \tau=\frac{x}{l}\in(0,1); \tag{3.18}$$

$$\zeta_\tau(1)-\theta(1)=\lambda p_4\zeta(1),\qquad \theta_\tau(1)=\lambda p_5\theta(1),\qquad \zeta(0)=\theta(0)=0, \tag{3.19}$$

where $\zeta_\tau(\tau)$ and $\theta_\tau(\tau)$ stand for derivatives with respect to τ. The eigenvalues and eigenvectors of the matrix in (3.18) are, respectively, given by

$$\mu_j=i\sigma_j,\quad v_j=\begin{pmatrix}4\mu_j(\lambda p_3-\sigma_j{}^2)\\ \lambda c_3+4\lambda p_1(\sigma_j{}^2-\lambda p_3)\\ \lambda c_3\\ \lambda c_3\mu_j\end{pmatrix},\quad j=1,2,3,4, \tag{3.20}$$

where

$$\begin{gathered}\sigma_1=-\sigma_2=\frac{\sqrt{2}}{2}\sqrt{c_1\lambda-\sqrt{c_2{}^2\lambda^2+c_3\lambda}},\qquad \sigma_3=-\sigma_4=\frac{\sqrt{2}}{2}\sqrt{c_1\lambda+\sqrt{c_2{}^2\lambda^2+c_3\lambda}},\\ c_1=p_1+p_3,\qquad c_2=p_1-p_3,\qquad c_3=4p_1p_2.\end{gathered} \tag{3.21}$$

The general solution of (3.18) therefore reads as

$$(\zeta,\zeta_\tau,\theta,\theta_\tau)^T(\tau)=C_1v_1e^{i\sigma_1\tau}+C_2v_2e^{-i\sigma_1\tau}+C_3v_3e^{i\sigma_3\tau}+C_4v_4e^{-i\sigma_3\tau}. \tag{3.22}$$

By substituting (3.22) into the boundary conditions (3.19), we get a system of linear algebraic equations with respect to (complex) variables C_1, C_2, C_3, C_4. That system has a nontrivial solution if

$$\varkappa(\lambda)=$$

$$\begin{vmatrix}e^{-i\sigma_1} & e^{i\sigma_1} & e^{-i\sigma_3} & e^{i\sigma_3}\\ \sigma_1(\sigma_1^2-\lambda p_3)e^{-i\sigma_1} & -\sigma_1(\sigma_1^2-\lambda p_3)e^{i\sigma_1} & \sigma_3(\sigma_3^2-\lambda p_3)e^{-i\sigma_3} & -\sigma_3(\sigma_3^2-\lambda p_3)e^{i\sigma_3}\\ (\sigma_1^2-\lambda p_3)(p_1+ip_4\sigma_1) & (\sigma_1^2-\lambda p_3)(p_1-ip_4\sigma_1) & (\sigma_3^2-\lambda p_3)(p_1+ip_4\sigma_3) & (\sigma_3^2-\lambda p_3)(p_1-ip_4\sigma_3)\\ i\sigma_1-\lambda p_5 & i\sigma_1+\lambda p_5 & i\sigma_3-\lambda p_5 & i\sigma_3+\lambda p_5\end{vmatrix}=0. \tag{3.23}$$

The roots of $\varkappa(\lambda)=0$ define the eigenvalues λ for the Sturm-Liouville problem (3.5) when its coefficients are constant. It is clear that the function $\varkappa(\lambda)$, given by (3.23), is analytic in its domain of definition. Then the uniqueness theorem for analytic functions implies that either $\varkappa(\lambda)\equiv 0$ or the set of all eigenvalues for (3.5) is discrete. The former is impossible for "typical" values of parameters (see, for example, [13], where the spectrum was estimated for a particular case $p_1=p_2=p_3=1$, $p_4=p_5=0$). We do not estimate solutions of the characteristic equation (3.23) here. Such a study requires additional assumptions on the mechanical parameters, based on real measurements, and is not of principal interest for this work.

4. The Galerkin approximation

To derive a Galerkin approximation (see, e.g., [6]), we consider a variational formulation of the boundary value problem as follows: if $(\tilde{\varphi}(t), \tilde{w}(x,t), \tilde{\psi}(x,t))$ $(0 \le x \le l)$ is a solution of (3.2), corresponding to $M(t)$, on an interval $t \in \mathcal{J} \subset \mathbb{R}$ then

$$\ddot{\tilde{\varphi}}(t) - v = 0,$$

$$\begin{aligned}
\tilde{\mu} = \int_0^l \delta\tilde{w}(x,t)\{&(\ddot{\tilde{w}} + xv - g\tilde{\varphi}\sin\varphi_0)\rho \\
&+ (K(\tilde{\psi} - \tilde{w}') + \rho_1 g(\tilde{w}'\sin\varphi_0 + \tilde{\varphi}w_0'\cos\varphi_0))' + \cdots\}dx \\
&+ \int_0^l \delta\tilde{\psi}(x,t)\{I_\rho\ddot{\tilde{\psi}} + K(\tilde{\psi} - \tilde{w}') - (EI\tilde{\psi}')' + I_\rho v\}dx + \delta\tilde{\psi}(l,t)\{J_c\ddot{\tilde{\psi}} + EI\tilde{\psi}' + J_c v\}|_{x=l} \\
&+ \delta\tilde{w}(l,t)\{K(\tilde{w}' - \tilde{\psi}) + m(\ddot{\tilde{w}} + lv - g(\tilde{\varphi} + \tilde{w}')\sin\varphi_0 - g\tilde{\varphi}w_0'\cos\varphi_0) + \cdots\}|_{x=l} = 0, \\
&\qquad \forall t \in \mathcal{J},
\end{aligned} \tag{4.1}$$

for each admissible variation $(\delta\tilde{w}(x,t), \delta\tilde{\psi}(x,t))$ satisfying the boundary conditions $\delta\tilde{w}(0, t) = 0$ and $\delta\tilde{\psi}(0,t) = 0$. (The derivation of $\tilde{\mu}$ from (3.2) uses the standard technique: integration by parts, collecting terms, and so forth. The expression (4.1) may also be obtained by expanding (2.9) in a neighborhood of the equilibrium and neglecting the higher order terms.) Here v is given by the expression (3.3).

Let us fix an integer number $N \ge 1$ and consider nontrivial solutions (λ_j, w_j, ψ_j) of (3.5) for $j = 1,2,\ldots,N$. We assume that all λ_j are different and substitute finite sums

$$\tilde{w}(x,t) = \sum_{j=1}^N q_j(t)w_j(x), \; \tilde{\psi}(x,t) = \sum_{j=1}^N q_j(t)\psi_j(x) \tag{4.2}$$

into (3.3) and (4.1). We also restrict $\delta\tilde{w}$ and $\delta\tilde{\psi}$ to finite-dimensional subspaces:

$$\delta\tilde{w}(\cdot,t) \in \operatorname{span}\{w_1(\cdot),\ldots,w_N(\cdot)\}, \qquad \delta\tilde{\psi}(\cdot,t) \in \operatorname{span}\{\psi_1(\cdot),\ldots,\psi_N(\cdot)\}. \tag{4.3}$$

By assuming $\delta\tilde{w}(x,t) = w_i(x)$ and $\delta\tilde{\psi}(x,t) = \psi_i(x)$ in (4.1) for $i = 1,2,\ldots,N$ and exploiting Lemma 3.1, we obtain the following control system with respect to $\tilde{\varphi}, q_1, q_2, \ldots, q_N$:

$$\begin{aligned}
\dot{z}_1 &= A_{11}z_1 + A_{12}z_2 + B_1u + R_1(z,u), \\
\dot{z}_2 &= A_{21}z_1 + A_{22}z_2 + B_2u + R_2(z,u), \quad z = \left(z_1^T, z_2^T\right)^T,
\end{aligned} \tag{4.4}$$

where z is the state, u is the control,

$$z_1 = (\widetilde{\varphi}, \dot{\widetilde{\varphi}})^T, \qquad z_2 = (q_1, \dot{q}_1, q_2, \dot{q}_2, \ldots, q_N, \dot{q}_N)^T,$$

$$u = \frac{\widetilde{M}}{J_0 + \int_0^l (w_0^2 \rho - {w_0'}^2 \rho_2)\,dx + m w_0^2(l)},$$

$$A_{11} = \begin{pmatrix} 0 & 1 \\ d_0 & 0 \end{pmatrix}, \qquad A_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ d_1 & 0 & d_2 & 0 & \cdots & d_N & 0 \end{pmatrix},$$

$$B_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad B_2 = (0, -b_1, \ldots, 0, -b_N)^T,$$

$$A_{21} = \begin{pmatrix} 0 & 0 \\ a_1 - b_1 d_0 & 0 \\ 0 & 0 \\ a_2 - b_2 d_0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ a_N - b_N d_0 & 0 \end{pmatrix}, \qquad A_{22} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -\lambda_1 - b_1 d_1 & 0 & -b_1 d_2 & 0 & \cdots & -b_1 d_N & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ -b_2 d_1 & 0 & -\lambda_2 - b_2 d_2 & 0 & \cdots & -b_2 d_N & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ -b_N d_1 & 0 & -b_N d_2 & 0 & \cdots & -\lambda_N - b_N d_N & 0 \end{pmatrix},$$

$$a_j = g \frac{(\int_0^l \rho w_j\,dx + m w_j(l)) \sin\varphi_0 + \int_0^l \rho_1 w_0' w_j'\,dx \cos\varphi_0}{\int_0^l (\rho w_j^2 + I_\rho \psi_j^2)\,dx + m w_j^2(l) + J_c \psi_j^2(l)},$$

$$b_j = \frac{\int_0^l (\rho x w_j + I_\rho \psi_j)\,dx + m l w_j(l) + J_c \psi_j(l)}{\int_0^l (\rho w_j^2 + I_\rho \psi_j^2)\,dx + m w_j^2(l) + J_c \psi_j^2(l)},$$

$$d_0 = \frac{\int_0^l [(\rho w_0 - \rho_1 w_0') \cos\varphi_0 - (1/2)(w_0')^2 \rho_1 \sin\varphi_0]\,dx + m w_0(l) \cos\varphi_0}{J_0 + \int_0^l (w_0^2 \rho - {w_0'}^2 \rho_2)\,dx + m w_0^2(l)} g,$$

$$d_j = \frac{g \int_0^l [(w_0' \cos\varphi_0 - \sin\varphi_0) \rho_1 w_j' + \rho w_j \sin\varphi_0]\,dx + m g w_j(l) \sin\varphi_0 + EI \psi_j'(0)}{J_0 + \int_0^l (w_0^2 \rho - {w_0'}^2 \rho_2)\,dx + m w_0^2(l)}, \tag{4.5}$$

the nonlinear term $R(z,u) = (R_1^T, R_2^T)^T$ satisfies the estimate

$$||R(z,u)|| = O(\|z\|^2 + u^2) \tag{4.6}$$

around the equilibrium point $z = 0$, $u = 0$. The control system (4.4) is a finite dimensional approximation of (3.2) corresponding to the flexible coordinates of order up to N.

5. Stabilization in finite dimensions

In this section, an explicit procedure for stabilizing controller design is proposed.

Theorem 5.1. *Assume that all eigenvalues* $(\lambda_1,\dots,\lambda_N)$ *are positive and different, and that* $a_j+\lambda_j b_j \neq 0$ *for each* $j=1,2,\dots,N$. *Then system (4.4) is stabilizable by the following feedback control:*

$$u = Kz, \qquad K=(K_1,K_2), \qquad K_1 = \left(-d_0 - \frac{h_1+\sum_{j=1}^N a_j(b_j+a_j/\lambda_j)}{h_2}, -\frac{h_0}{h_2}\right),$$
$$K_2 = \left(-d_1 + \frac{a_1+\lambda_1 b_1}{h_2}, 0, -d_2 + \frac{a_2+\lambda_2 b_2}{h_2}, 0, \dots, -d_N + \frac{a_N+\lambda_N b_N}{h_2}, 0\right), \tag{5.1}$$

where h_0, h_1, *and* h_2 *are arbitrary positive constants.*

Proof. Consider a Lyapunov function candidate

$$2V(z) = \left(h_1 + \sum_{j=1}^N \frac{a_j^2}{\lambda_j}\right)\tilde{\varphi}^2 + \left(h_2 + \sum_{j=1}^N b_j^2\right)\dot{\tilde{\varphi}}^2 + \sum_{j=1}^N \left(\lambda_j {q_j}^2 + \dot{q}_j^2 - 2a_j\tilde{\varphi}q_j + 2b_j\dot{\tilde{\varphi}}\dot{q}_j\right). \tag{5.2}$$

By applying the Cauchy-Schwartz inequality, we get

$$2V \geq G_1\left(-|\tilde{\varphi}|, \left(\sum_{j=1}^N \lambda_j q_j^2\right)^{1/2}\right) + G_2\left(-|\dot{\tilde{\varphi}}|, \left(\sum_{j=1}^N \dot{q}_j^2\right)^{1/2}\right), \tag{5.3}$$

where

$$G_1(\alpha,\beta) = \left(h_1 + \sum_{j=1}^N \frac{a_j^2}{\lambda_j}\right)\alpha^2 + 2\left(\sum_{j=1}^N \frac{{a_j}^2}{\lambda_j}\right)^{1/2}\alpha\beta + \beta^2,$$
$$G_2(\alpha,\beta) = \left(h_2 + \sum_{j=1}^N b_j^2\right)\alpha^2 + 2\left(\sum_{j=1}^N {b_j}^2\right)^{1/2}\alpha\beta + \beta^2. \tag{5.4}$$

Sylvester's criterion for quadratic forms G_1 and G_2 implies that both G_1 and G_2 are positive definite if $h_1>0$ and $h_2>0$. Then the quadratic form V is positive definite due to estimate (5.3).

The time-derivative of V along the trajectories of the linear part of (4.4) is

$$\dot{V} = h_2\dot{\tilde{\varphi}}v + \left(h_1 + h_2 d_0 + \sum_{j=1}^N a_j\left(\frac{b_j+a_j}{\lambda_j}\right)\right)\tilde{\varphi}\dot{\tilde{\varphi}} + \dot{\tilde{\varphi}}\sum_{j=1}^N q_j(h_2 d_j - a_j - \lambda_j b_j). \tag{5.5}$$

We choose a constant $h_0>0$ arbitrarily and define the feedback control in order to have $\dot{V} = -h_0\dot{\tilde{\varphi}}^2$. This yields expression (5.1).

Now we apply the Barbashin-Krasovskii theorem (or LaSalle's invariance principle, cf. [21]). For this purpose consider the set

$$Z_0 = \{(\tilde{\varphi}, \dot{\tilde{\varphi}}, q_1, \dot{q}_1, \dots, q_N, \dot{q}_N) \in \mathbb{R}^{2N+2} : \dot{V} = 0\}. \tag{5.6}$$

Each positive semitrajectory of the linear approximation of (4.4) with (5.1) on Z_0 satisfies the following relations:

$$\begin{gathered} \ddot{q}_j = -\lambda_j q_j + a_j \tilde{\varphi}, \\ \sum_{j=1}^{N} \left[-a_j b_j \tilde{\varphi} + (a_j + \lambda_j b_j) q_j \right] = \left(h_1 + \sum_{j=1}^{N} \frac{a_j^2}{\lambda_j} \right) \tilde{\varphi} = \text{const}, \quad t \geq 0. \end{gathered} \tag{5.7}$$

The above relations imply

$$\sum_{j=1}^{N} (a_j + \lambda_j b_j) \left(r_{1j} \cos\left(\sqrt{\lambda_j} t\right) + r_{2j} \sin\left(\sqrt{\lambda_j} t\right) \right) = h_1 \tilde{\varphi} \tag{5.8}$$

for some constants r_{1j}, r_{2j}, and $\tilde{\varphi}$. Exploiting the fact that

$$\left\{ 1, \sin\left(\sqrt{\lambda_j} t\right), \cos\left(\sqrt{\lambda_j} t\right) \right\}_{j=1}^{N} \tag{5.9}$$

are linearly-independent functions on $[0, +\infty)$ (cf. [22]), we get that (5.8) is possible only if $\tilde{\varphi} = 0$ and $r_{1j} = r_{2j} = 0$ for all $j = 1, 2, \dots, N$. Thus, the only semitrajectory of the linearized closed-loop system on Z_0 is the trivial one, and the trivial solution of the linear part of (4.4), (5.1) is asymptotically stable by the Barbashin-Krasovskii theorem (LaSalle's invariance principle). Now local asymptotic stability of the nonlinear closed-loop system follows from Lyapunov's theorem on stability using linearization. □

Remark 5.2. As it follows from the representation $\dot{V} = -h_0 \dot{\tilde{\varphi}}^2$, the choice of constant h_0 affects the decay rate of the Lyapunov function along trajectories of the closed-loop system. On the one hand, the more h_0 the faster convergence of solutions to the equilibrium could be achieved (for solutions with $\dot{\tilde{\varphi}} \neq 0$). On the other hand, for large h_0, the gain $-h_0/h_2$, appearing in formula (5.1), may take large values if h_2 is small. This suggests us to choose h_0 as maximal as possible, and to select h_2 in such a way that the term $-(h_0/h_2)\dot{\tilde{\varphi}}$, appearing in $u = Kz$, would not bring the control input u to its saturation bound (for typical disturbances $\dot{\tilde{\varphi}}$). The constant h_1 should be then defined according to a desired geometry of the level surfaces for the quadratic form V. Indeed, constants h_1 and h_2 define a relation between semiaxes for the ellipsoids $V(z) = \text{const}$, and hence a desired ratio between overshoots for $\tilde{\varphi}$ and $\dot{\tilde{\varphi}}$ can be estimated in terms of h_1 and h_2. Certainly, this suggestion is based on the linearized system and does not give a rigorous characterization of the global behavior.

6. Observer design

In order to implement the feedback law (5.1) in practice, one should reconstruct the complete state vector of (4.4) from the outputs which can be measured. The values of $w(x,t)$

and $\psi(x,t)$ cannot be directly estimated in a real flexible manipulator. Instead, there is a set of strain gauges located at a point $x = l_0$, $0 \le l_0 \le l$, which allows measurement of some components of the strain tensor. By using only the principal part of the strain at $x = l_0$, we get the output $\psi'(x,t)|_{x=l_0}$ for each $t \ge 0$. By subtracting from the signals $\varphi(t)$ and $\psi'(x,t)|_{x=l_0}$ their steady-state values and rescaling, we assume that the following outputs are available for the finite dimensional approximation (4.4):

$$y_1(t) = \tilde{\varphi}(t), \qquad y_2(t) = l^2 \tilde{\psi}'(x,t)|_{x=l_0} = \sum_{j=1}^{N} \chi_j q_j(t), \tag{6.1}$$

where $\chi_j = l^2 \psi_j'(l_0)$. We introduce the factor l^2 in order to get the dimension of length for the output y_2.

Let us rewrite the output (6.1) as follows:

$$y_1 = C_1 z_1, \quad y_2 = C_2 z_2, \quad C_1 = (1,0), \quad C_2 = (\chi_1, 0, \chi_2, 0, \dots, \chi_N, 0). \tag{6.2}$$

Lemma 6.1. *The control system (4.4), (6.2) is locally observable at $z = 0$ if*

$$\begin{vmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{N1} & \pi_{N2} & \cdots & \pi_{NN} \end{vmatrix} \ne 0, \tag{6.3}$$

where $\pi_{1,j} = \chi_j$, $\pi_{k,j} = -\lambda_j \pi_{k-1,j} - d_j \sum_{i=1}^{N} \pi_{k-1,i} b_i$, $j = \overline{1,N}$, $k = \overline{2,N}$.

In particular, the condition (6.3) is equivalent to $\chi_1 \ne 1$ if $N = 1$ or

$$\chi_1 \chi_2 (\lambda_1 - \lambda_2 + b_1 d_1 - b_2 d_2) + b_2 \chi_2^2 d_1 - b_1 \chi_1^2 d_2 \ne 0 \quad \text{if } N = 2. \tag{6.4}$$

Proof. The linear part of (4.4), (6.2) can be written in terms of output y_1 as follows:

$$\begin{aligned} & z_1 = (y_1, \dot{y}_1)^T, \\ & \dot{z}_2 = A_{22} z_2 + B_2 u + (0, a_1 - b_1 d_0, 0, a_2 - b_2 d_0, \dots, 0, a_N - b_N d_0)^T y_1, \\ & y_2 = C_2 z_2. \end{aligned} \tag{6.5}$$

Hence, the above system is observable if the pair (A_{22}, C_2) satisfies the Kalman observability condition (cf. [23, Theorem 3.1, page 58]):

$$\operatorname{rank} \begin{pmatrix} C_2 \\ C_2 A_{22} \\ \vdots \\ C_2 A_{22}^{2N-1} \end{pmatrix} = 2N. \tag{6.6}$$

Straightforward computations show that

$$\det\begin{pmatrix} C_2 \\ C_2A_{22} \\ \vdots \\ C_2A_{22}^{2N-1} \end{pmatrix} = \begin{vmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{N1} & \pi_{N2} & \cdots & \pi_{NN} \end{vmatrix}^2. \tag{6.7}$$

Therefore, (6.3) implies the observability rank condition for the linear part of (4.4), (6.2). It also means that (4.4), (6.2) is strongly locally observable at $z = 0$ by the Hermann-Krener theorem [24]. □

The following theorem gives an explicit procedure for the Luenberger-type observer design.

Theorem 6.2. *Suppose that the control system (4.4), (6.2) satisfies the observability condition (6.3), all λ_j are positive and different, $a_j + \lambda_j b_j \neq 0$, and $b_j d_j > 0$ for $j = \overline{1,N}$. Then the origin $z = 0$, $\overline{z} = 0$ of the extended system (4.4), (6.2) and*

$$\begin{aligned} \dot{\overline{z}}_1 &= (A_{11} - F_1C_1)\overline{z}_1 + A_{12}\overline{z}_2 + F_1y_1 + B_1u, \\ \dot{\overline{z}}_2 &= (A_{22} - F_{22}C_2)\overline{z}_2 + F_{21}y_1 + F_{22}y_2 + B_2u \end{aligned} \tag{6.8}$$

with $u = K\overline{z}$ is locally asymptotically stable, where K is given by (5.1),

$$F_1 = (\phi_1, d_0 + \phi_2)^T, \qquad F_{21} = (0, a_1 - b_1d_0, 0, a_2 - b_2d_0, \dots, 0, a_N - b_Nd_0)^T,$$

$$F_{22} = (f_1, 0, f_2, 0, \dots, f_N, 0)^T, \qquad (f_1, f_2, \dots, f_N)^T = \gamma Q^{-1}(\chi_1, \chi_2, \dots, \chi_N)^T,$$

$$Q = \begin{pmatrix} \dfrac{\lambda_1 d_1}{b_1} + d_1^2 & d_1d_2 & \cdots & d_1d_N \\ d_2d_1 & \dfrac{\lambda_2 d_2}{b_2} + d_2^2 & \cdots & d_2d_N \\ \vdots & \vdots & \ddots & \vdots \\ d_Nd_1 & d_Nd_2 & \cdots & \dfrac{\lambda_N d_N}{b_N} + d_N^2 \end{pmatrix}. \tag{6.9}$$

Here ϕ_1, ϕ_2, and γ are any positive constants.

Proof. Consider the observation errors $e_1 = z_1 - \overline{z}_1$, $e_2 = z_2 - \overline{z}_2$. Then subtracting (6.8) from (4.4) yields the following dynamics:

$$\dot{e}_1 = H_1e_1 + A_{12}e_2 + R_1(z,u), \qquad \dot{e}_2 = H_2e_2 + R_2(z,u), \tag{6.10}$$

here $H_1 = A_{11} - F_1C_1$ and $H_2 = A_{22} - F_{22}C_2$. We see that the roots of the polynomial

$$\det(H_1 - \mu I) = \begin{vmatrix} -\phi_1 - \mu & 1 \\ -\phi_2 & -\mu \end{vmatrix} = \mu^2 + \phi_1\mu + \phi_2 \tag{6.11}$$

have negative real parts if and only if $\phi_1 > 0$ and $\phi_2 > 0$. Our goal is to show that the real parts of all eigenvalues of

$$H_2 = \begin{pmatrix} -f_1\chi_1 & 1 & -f_1\chi_2 & 0 & \cdots & -f_1\chi_N & 0 \\ -\lambda_1 - b_1 d_1 & 0 & -b_1 d_2 & 0 & \cdots & -b_1 d_N & 0 \\ -f_2\chi_1 & 0 & -f_2\chi_2 & 1 & \cdots & -f_2\chi_N & 0 \\ -b_2 d_1 & 0 & -\lambda_2 - b_2 d_2 & 0 & \cdots & -b_2 d_N & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -f_N\chi_1 & 0 & -f_N\chi_2 & 0 & \cdots & -f_N\chi_N & 1 \\ -b_N d_1 & 0 & -b_N d_2 & 0 & \cdots & -\lambda_N - b_N d_N & 0 \end{pmatrix} \qquad (6.12)$$

are also negative if the conditions of Theorem 6.2 are satisfied. Let us denote $e_2 = (\xi_1, \eta_1, \dots, \xi_N, \eta_N)^T$ and consider the following quadratic form:

$$2W(e_2) = \sum_{j=1}^{N} \frac{d_j \eta_j^2}{b_j} + (\xi_1, \xi_2, \dots, \xi_N) Q (\xi_1, \xi_2, \dots, \xi_N)^T. \qquad (6.13)$$

This form is positive definite as $\lambda_j > 0$ and $b_j d_j > 0$. Indeed, all principal minors Δ_j of Q are positive:

$$\Delta_j = \frac{(\lambda_1 d_1)(\lambda_2 d_2) \cdots (\lambda_j d_j)}{b_1 b_2 \cdots b_j} \left(1 + \sum_{i=1}^{j} \frac{b_i d_i}{\lambda_i} \right) > 0, \quad j = \overline{1, N}. \qquad (6.14)$$

Then Sylvester's criterion implies that W is positive definite. The inequality $\det(Q) = \Delta_N > 0$ also proves invertibility of Q in (6.9). By computing the time derivative of W along the trajectories of the linear system $\dot{e}_2 = H_2 e_2$, we get

$$\dot{W}(e_2) = -\gamma (C_2 e_2)^2 \le 0, \qquad (6.15)$$

provided that F_{22} is defined by (6.9). As the time derivative of W is negative semi-definite and vanishes on $\ker C_2 = \{e_2 \in \mathbb{R}^{2N} : C_2 e_2 = 0\}$, we check whether the linear system $\dot{e}_2 = H_2 e_2$ admits a nontrivial semitrajectory on $\ker C_2$. Let $C_2 e_2(t) \equiv 0$, $t \ge 0$, then

$$\frac{d^k}{dt^k} C_2 e_2(t) = C_2 (A_{22} - F_{22} C_2)^k e_2(t) = C_2 A_{22}^k e_2(t) = 0, \quad t \ge 0,\ k \ge 0. \qquad (6.16)$$

This implies that, for each $t \ge 0$, $e_2(t)$ is a solution of the following system of linear algebraic equations:

$$C_2 A_{22}^k e_2(t) = 0, \quad k = \overline{0, 2N - 1}. \qquad (6.17)$$

The above system has only the trivial solution $e_2(t) = 0$ because of the observability rank condition (6.3). This proves asymptotic stability of the linear system $\dot{e}_2 = H_2 e_2$ by the Barbashin-Krasovskii theorem (LaSalle's invariance principle).

We have shown that the matrices H_1 and H_2 are Hurwitz. The nonlinear closed-loop system (4.4), (6.2), (6.8) with $u = K\overline{z}$ can be written in variables (z,e) as follows:

$$\begin{pmatrix} \dot{z} \\ \dot{e}_1 \\ \dot{e}_2 \end{pmatrix} = \begin{pmatrix} H_0 & -BK_1 & -BK_2 \\ 0 & H_1 & A_{12} \\ 0 & 0 & H_2 \end{pmatrix} \begin{pmatrix} z \\ e_1 \\ e_2 \end{pmatrix} + \begin{pmatrix} R(z,K(z-e)) \\ R_1(z,K(z-e)) \\ R_2(z,K(z-e)) \end{pmatrix}, \tag{6.18}$$

where

$$H_0 = \begin{pmatrix} A_{11}+B_1K_1 & A_{12}+B_1K_2 \\ A_{21}+B_2K_1 & A_{22}+B_2K_2 \end{pmatrix}, \quad B = (B_1^T, B_2^T)^T. \tag{6.19}$$

As $a_j + \lambda_j b_j \neq 0$ then the conditions of Theorem 5.1 are satisfied and H_0 is Hurwitz. Hence, the trivial solution of (6.18) is asymptotically stable by linear approximation as the spectrum of its matrix is the union of spectra of the Hurwitz matrices H_0, H_1, and H_2. □

7. Conclusions

We have proposed a feedback controller that stabilizes the equilibrium of a Galerkin approximation for a rotating Timoshenko beam, provided that measurements of the raising angle and the strain at a point are available. The feedback law and coefficients of the dynamical observer are computed explicitly for any number of modal coordinates. A potential field of application of these results is the control design for fire-rescue turntable ladders. An advantage of our approach is that the identification procedure can be reduced significantly in comparison with a multibody model. In addition, the higher modes can be calculated explicitly, which is important for the design of an oscillation damping control of a turntable ladder. For a possible implementation of the controller, it is necessary to integrate a system of ordinary differential equations in real time. We do not consider here such issues as spillover analysis, convergence of Galerkin approximations, computational complexity, or limitation of the sampling rate with respect to the calculation time leaving these problems for future work.

Acknowledgments

This work was supported in part by the Alexander von Humboldt Foundation and under Grant of the President of Ukraine for Young Scientists. The authors express gratitude to anonymous reviewers for their valuable comments.

References

[1] Z.-H. Luo, B.-Z. Guo, and O. Morgul, *Stability and Stabilization of Infinite Dimensional Systems with Applications*, Communications and Control Engineering Series, Springer, London, UK, 1999.

[2] H. A. Talebi, R. V. Patel, and K. Khorasani, *Control of Flexible-Link Manipulators Using Neural Networks*, Springer, London, UK, 2001.

[3] O. Sawodny, H. Aschemann, and A. Bulach, “Mechatronical designed control of fire-rescue turntable ladders as flexible link robots,” in *Proceedings of the 15th IFAC World Congress*, Barcelona, Spain, July 2002, CD-ROM file 385.pdf.

[4] K. Grosh and P. M. Pinsky, “Design of Galerkin generalized least squares methods for Timoshenko beams,” *Computer Methods in Applied Mechanics and Engineering*, vol. 132, no. 1-2, pp. 1–16, 1996.

[5] A. Zuyev and O. Sawodny, “Control design for Galerkin approximations of a flexible structure,” in *Proceedings of IFAC Workshop on Generalized Solutions in Control Problems (GSCP ’04)*, pp. 311–318, Fizmatlit, Pereslavl-Zalessky, Russia, September 2004.

[6] J. Donea and A. Huerta, *Finite Element Methods for Flow Problems*, John Wiley & Sons, Chichester, UK, 2003.

[7] J. E. Lagnese and G. Leugering, “Controllability of thin elastic beams and plates,” in *The Control Handbook*, W. S. Levine, Ed., pp. 1139–1156, CRC Press-IEEE Press, Boca Raton, Fla, USA, 1996.

[8] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Continuous and Approximation Theories. II: Abstract Hyperbolic-Like Systems Over a Finite Time Horizon*, vol. 75 of *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, Cambridge, UK, 2000.

[9] C.-Z. Xu and J. Baillieul, “Stabilizability and stabilization of a rotating body-beam system with torque control,” *IEEE Transactions on Automatic Control*, vol. 38, no. 12, pp. 1754–1765, 1993.

[10] A. Zuyev, “Partial asymptotic stabilization of nonlinear distributed parameter systems,” *Automatica*, vol. 41, no. 1, pp. 1–10, 2005.

[11] S. P. Timoshenko, “On the correction for shear of the differential equation for transverse vibrations of prismatic bars,” *Philisophical Magazine*, vol. 41, pp. 744–746, 1921, reprinted in: *The Collected Papers* of S. P. Timoshenko, McGraw-Hill, London, UK, 1953.

[12] J. U. Kim and Y. Renardy, “Boundary control of the Timoshenko beam,” *SIAM Journal on Control and Optimization*, vol. 25, no. 6, pp. 1417–1429, 1987.

[13] W. Krabs and G. M. Sklyar, “On the controllability of a slowly rotating Timoshenko beam,” *Journal for Analysis and Its Applications*, vol. 18, no. 2, pp. 437–448, 1999.

[14] W. Krabs and G. M. Sklyar, “On the stabilizability of a slowly rotating Timoshenko beam,” *Zeitschrift für Analysis und ihre Anwendungen*, vol. 19, no. 1, pp. 131–145, 2000.

[15] Ö. Morgül, “Boundary control of a Timoshenko beam attached to a rigid body: planar motion,” *International Journal of Control*, vol. 54, no. 4, pp. 763–791, 1991.

[16] S. W. Taylor, “A smoothing property of a hyperbolic system and boundary controllability,” *Journal of Computational and Applied Mathematics*, vol. 114, no. 1, pp. 23–40, 2000.

[17] S. W. Taylor and S. C. B. Yau, “Boundary control of a rotating Timoshenko beam,” *Australian & New Zealand Industrial and Applied Mathematics Journal*, vol. 44, part E, pp. 143–184, 2003.

[18] D.-H. Shi, S. H. Hou, and D.-X. Feng, “Feedback stabilization of a Timoshenko beam with an end mass,” *International Journal of Control*, vol. 69, no. 2, pp. 285–300, 1998.

[19] M. Tadi, “Computational algorithm for controlling a Timoshenko beam,” *Computer Methods in Applied Mechanics and Engineering*, vol. 153, no. 3-4, pp. 153–165, 1998.

[20] A. Zuyev and O. Sawodny, “Observer design for a flexible manipulator model with a payload,” in *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 4490–4495, San Diego, Calif, USA, December 2006.

[21] N. Rouche, P. Habets, and M. Laloy, *Stability Theory by Liapunov’s Direct Method*, vol. 22 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1977.

[22] W. Krabs, *On Moment Theory and Controllability of One-Dimensional Vibrating Systems and Heating Processes*, vol. 173 of *Lecture Notes in Control and Information Sciences*, Springer, Berlin, Germany, 1992.

[23] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*, vol. 10 of *Applications of Mathematics (New York)*, Springer, New York, NY, USA, 3rd edition, 1985.

[24] R. Hermann and A. J. Krener, “Nonlinear controllability and observability,” *IEEE Transactions on Automatic Control*, vol. 22, no. 5, pp. 728–740, 1977.

Alexander Zuyev: Institute of Applied Mathematics & Mechanics, National Academy of Sciences of
Ukraine, R. Luxembourg 74, 83114 Donetsk, Ukraine
Email address: al_zv@mail.ru

Oliver Sawodny: Institute for System Dynamics, Universität Stuttgart, Pfaffenwaldring 9,
70569 Stuttgart, Germany
Email address: sawodny@isys.uni-stuttgart.de

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 59372, 23 pages
doi:10.1155/2007/59372

Research Article

Numerical and Analytical Study of Optimal Low-Thrust Limited-Power Transfers between Close Circular Coplanar Orbits

Sandro da Silva Fernandes and Wander Almodovar Golfetto

Received 22 September 2006; Revised 18 April 2007; Accepted 20 May 2007

Recommended by José Manoel Balthazar

A numerical and analytical study of optimal low-thrust limited-power trajectories for simple transfer (no rendezvous) between close circular coplanar orbits in an inverse-square force field is presented. The numerical study is carried out by means of an indirect approach of the optimization problem in which the two-point boundary value problem, obtained from the set of necessary conditions describing the optimal solutions, is solved through a neighboring extremal algorithm based on the solution of the linearized two-point boundary value problem through Riccati transformation. The analytical study is provided by a linear theory which is expressed in terms of nonsingular elements and is determined through the canonical transformation theory. The fuel consumption is taken as the performance criterion and the analysis is carried out considering various radius ratios and transfer durations. The results are compared to the ones provided by a numerical method based on gradient techniques.

1. Introduction

The main purpose of this paper is to present a numerical and analytical study of optimal low-thrust limited-power trajectories for simple transfers (no rendezvous) between close circular coplanar orbits in an inverse-square force field. The study of these transfers is particularly interesting because the orbits found in practice often have a small eccentricity and the problem of slight modifications (corrections) of these orbits is frequently met [1]. Besides, the analysis has been motivated by the renewed interest in the use of low-thrust propulsion systems in space missions verified in the last two decades. Several researchers have obtained numerical, and sometimes analytical, solutions for a number of specific initial orbits and specific thrust profiles [2–10]. Averaging methods are also used in such researches [11–15].

Low-thrust electric propulsion systems are characterized by high specific impulse and low-thrust capability and have a great interest for high-energy planetary missions and certain Earth orbit missions. For trajectory calculations, two idealized propulsion models are of most frequent use [1]: LP and CEV systems. In the power-limited variable ejection velocity systems or, simply, LP systems, the only constraint concerns the power, that is, there exists an upper constant limit for the power. In the constant ejection velocity limited-thrust systems or, simply, CEV systems, the magnitude of the thrust acceleration is bounded. In both cases, it is usually assumed that the thrust direction is unconstrained. The utility of these idealized models is that the results obtained from them provide good insight into more realistic problems. In this paper, only LP systems are considered.

In the study presented in the paper, the fuel consumption is taken as the performance criterion and it is calculated for various radius ratios $\rho = r_f/r_0$, where r_0 is the radius of the initial circular orbit O_0 and r_f is the radius of the final circular orbit O_f, and for various transfer durations $t_f - t_0$. Transfers with small and moderate amplitudes are considered. The optimization problem associated to the space transfer problem is formulated as a Mayer problem of optimal control with Cartesian elements—components of position and velocity vectors—as state variables.

The numerical study is carried out by a neighboring extremal algorithm which is based on the linearization about an extremal solution of the nonlinear two-point boundary value problem defined by the set of necessary conditions for a Bolza problem of optimal control with fixed initial and final times, fixed initial state, and constrained final state [16, 17]. The resulting linear two-point boundary value problem is solved through Riccati transformation. As briefly described in Section 2, a slight modification is introduced in the algorithm to improve the convergence. On the other hand, the analytical study is based on a linear theory expressed in terms of nonsingular orbital elements, similar to the ones presented in [1, 18]. Here, the linear theory is determined through canonical transformation theory using the concept of generalized canonical systems. This approach provides a simple way to compare the numerical solutions and the analytical theory. The numerical and analytical results are compared to the ones obtained through an algorithm based on gradient techniques [19, 20].

2. Neighboring extremal algorithm based on Riccati transformation

For completeness, a brief description of the neighboring extremal algorithm used in the paper is presented in this section. This algorithm has a slight modification when compared to the well-known algorithms in the literature [16, 17, 21]: a constraint on the control variations is introduced. Numerical experiments have shown that this simple device improves the convergence.

Let the system of differential equations be defined by

$$\frac{dx_i}{dt} = f_i(x, u), \quad i = 1, \dots, n, \tag{2.1}$$

where x is an n-vector of state variables and u is an m-vector of control variables. It is assumed that there exist no constraints on the state or control variables. The problem consists in determining the control $u^*(t)$ that transfers the system (2.1) from the initial conditions

$$x(t_0) = x_0, \tag{2.2}$$

to the final conditions at t_f,

$$\psi(x(t_f)) = 0, \tag{2.3}$$

and minimizes the performance index

$$J[u] = g(x(t_f)) + \int_{t_0}^{t_f} F(x,u)dt. \tag{2.4}$$

The functions $f(\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $F(\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g(\cdot) : \mathbb{R}^n \to \mathbb{R}$, and $\psi(\cdot) : \mathbb{R}^n \to \mathbb{R}^q$, $q < n$, are assumed to be twice continuously differentiable with respect to their arguments. Furthermore, it is assumed that the matrix $[\partial\psi/\partial x]$ has a maximum rank.

By applying the Pontryagin maximum principle [21, 22] to the Bolza problem with constrained final state and fixed terminal times defined by (2.1)–(2.4), the following two-point boundary value problem is obtained:

$$\frac{dx}{dt} = H_\lambda^T, \tag{2.5}$$

$$\frac{d\lambda}{dt} = -H_x^T, \tag{2.6}$$

$$H_u = 0, \tag{2.7}$$

with

$$x(t_0) = x_0, \tag{2.8}$$

$$\lambda(t_f) = -(g_x + \mu^T \psi_x)^T, \tag{2.9}$$

$$\psi(x(t_f)) = 0, \tag{2.10}$$

where $H(x,\lambda,u) = -F(x,u) + \lambda^T f(x,u)$ is the Hamiltonian function, λ is an n-vector of adjoint variables and μ is a q-vector of Lagrange multipliers. The quantities $H_x, H_u, g_x, \ldots,$ and so forth, denote the partial derivatives. If x, λ, and u are taken to be column vectors, then H_x, H_λ, and H_u are row vectors. In this way, ψ_x is a $q \times n$-matrix. The superscript T denotes the transpose of a matrix or a (row or column) vector.

Neighboring extremal methods are iterative procedures used for solving the two-point boundary value problem defined through (2.5)–(2.10). These methods are based on the second variation theory and consist in determining iteratively the unknown adjoint variables $\lambda(t_0)$ and Lagrange multipliers μ. Let $\lambda^0(t_0)$ be an arbitrary starting approximation of the unknown adjoint variables at t_0. The trajectory $x^0(t)$ corresponding to these starting values is obtained by integrating (2.5) from t_0 to t_f, with the initial conditions (2.8). The vector of Lagrange multipliers μ is then calculated such that the transversality condition (2.9) is fulfilled. Since ψ_x has a maximum rank, one finds that

$$\mu = -\left(\psi_x \psi_x^T\right)^{-1} \psi_x \left(\lambda(t_f) + g_x^T\right). \tag{2.11}$$

Let $\lambda^1(t_0) = \lambda^0(t_0) + \delta\lambda(t_0)$ and $\mu^1 = \mu^0 + \delta\mu$ be the next approximation. Following [16, 21], the corrections (perturbations) $\delta\lambda(t_0)$ and $\delta\mu$ are obtained in order to satisfy the linear two-point boundary value problem obtained from the linearization of (2.5)–(2.10) about a nominal extremal solution defined by $\lambda^0(t_0)$:

$$\delta\dot{x} = H_{\lambda x}\delta x + H_{\lambda u}\delta u, \tag{2.12}$$

$$\delta\dot{\lambda} = -H_{x\lambda}\delta\lambda - H_{xx}\delta x - H_{xu}\delta u, \tag{2.13}$$

$$H_{ux}\delta x + H_{u\lambda}\delta\lambda + H_{uu}\delta u = 0, \tag{2.14}$$

$$\delta x(t_0) = 0, \tag{2.15}$$

$$\psi_x \delta x(t_f) = -k\psi(x(t_f)), \tag{2.16}$$

$$\delta\lambda(t_f) = -\left(g_{xx} + \mu^T \psi_{xx}\right)\delta x(t_f) - \psi_x^T \delta\mu, \tag{2.17}$$

where the constant k, $0 < k \leq 1$, has been introduced to indicate that the correction is partial. Quantities such as $g_{xx}, H_{xx}, H_{x\lambda}, H_{xu}, \ldots$, and so forth, are matrices of second partial derivatives; for instance, $H_{xu} = [\partial^2 H/\partial x_i \partial u_j]$ is an $n \times m$ matrix. According to our notation $H_{\lambda x} = H_{x\lambda}^T$.

Equations (2.12)–(2.17) form the two-point boundary value problem to the accessory minimum problem associated to the original optimization problem defined by (2.1)–(2.4) [16, 17, 21]. This accessory minimum problem is obtained expanding the augmented performance index, which includes through adjoint variables the constraints represented by the state equations, to second order and all constraints to first order, as described in the appendix.

According to the appendix, (2.14) must be replaced by (A.8) since a constraint on the control variations is imposed [23]. Assuming that W_2 is chosen such that $H_{uu} + W_2$ is nonsingular for $t \in \lfloor t_0, t_f \rfloor$, we may solve (A.8) for $\delta u(t)$, in terms of $\delta x(t)$ and $\delta\lambda(t)$:

$$\delta u(t) = -\left(H_{uu} + W_2\right)^{-1}\left(H_{ux}\delta x(t) + H_{u\lambda}\delta\lambda(t)\right). \tag{2.18}$$

Substituting this equation into (2.12) and (2.13), it follows that

$$\delta\dot{x} = A\delta x + B\delta\lambda, \tag{2.19}$$

$$\delta\dot{\lambda} = C\delta x - A^T\delta\lambda, \tag{2.20}$$

where matrices A, B, and C are given by

$$\begin{aligned} A(t) &= H_{\lambda x} - H_{\lambda u}\left(H_{uu} + W_2\right)^{-1} H_{ux}, \\ B(t) &= -H_{\lambda u}\left(H_{uu} + W_2\right)^{-1} H_{u\lambda}, \\ C(t) &= H_{xu}\left(H_{uu} + W_2\right)^{-1} H_{ux} - H_{xx}. \end{aligned} \tag{2.21}$$

We recall that the matrices A, B and C are evaluated on a nominal extremal solution.

Equations (2.15) through (2.20) represent a linear two-point boundary value problem, whose solution can be obtained through a backward sweep method which uses the Riccati transformation [21]:

$$\begin{aligned} \delta\lambda(t) &= R(t)\delta x(t) + L(t)\delta\mu, \\ k\psi &= L^T(t)\delta x(t) + Q(t)\delta\mu, \end{aligned} \tag{2.22}$$

where R is an $n \times n$ symmetric matrix, L is an $n \times q$ matrix, and Q is a $q \times q$ symmetric matrix. For (2.22) to be consistent with (2.15)–(2.20), the Riccati coefficients must satisfy the differential equations (2.23) with the boundary conditions (2.24) defined below.

The step-by-step computing procedure to be used in the neighboring extremal algorithm is summarized as follows.

(1) Guess the starting approximation for $\lambda(t_0)$, that is, $\lambda^0(t_0)$.

(2) The control $u = u(x,\lambda)$ is obtained from (2.7): $H_u = 0$.

(3) Integrate forward, from t_0 to t_f, the system of differential equations (2.5) and (2.6) with the initial conditions $x(t_0) = x_0$ and $\lambda(t_0) = \lambda^0(t_0)$ in order to obtain $x(t_f)$ and $\lambda(t_f)$.

(4) Compute μ through (2.11).

(5) Integrate backward, from t_f to t_0, the differential equations for the Riccati coefficients

$$\begin{aligned} -\dot{R} &= RA + A^T R + RBR - C, \\ -\dot{L} &= \left(A^T + RB\right)L, \\ -\dot{Q} &= L^T BL, \end{aligned} \tag{2.23}$$

with the boundary conditions

$$\begin{aligned} R(t_f) &= -\left(g_{xx} + \mu^T \psi_{xx}\right), \\ L(t_f) &= -\psi_x^T, \\ Q(t_f) &= 0, \end{aligned} \tag{2.24}$$

and the system of differential equations (2.5) and (2.6) with boundary conditions $x(t_f)$ and $\lambda(t_f)$.

(6) Compute the variation $\delta\mu$ from $\delta\mu = Q(t_0)^{-1}k\psi$.

(7) Compute $\delta\lambda(t_0)$ from $\delta\lambda(t_0) = L(t_0)\delta\mu$.

(8) Test the convergence. If it is not obtained, update the unknown $\lambda(t_0)$, that is, compute the new value $\lambda^1(t_0) = \lambda^0(t_0) + \delta\lambda(t_0)$.

(9) Go back to step 2 and repeat the procedure until convergence is achieved.

3. Optimal low-thrust trajectories

In what follows, the neighboring extremal algorithm presented in previous section is applied to determine optimal low-thrust limited-power transfers between close coplanar circular orbits in an inverse-square force field.

3.1. Problem formulation. A low-thrust limited-power propulsion system, or LP system, is characterized by low-thrust acceleration level and high specific impulse [1]. The ratio between the maximum thrust acceleration and the gravity acceleration on the ground, $\gamma_{\max}/g_0$, is between 10^{-4} and 10^{-2}. For such system, the fuel consumption is described by the variable J defined as

$$J = \frac{1}{2}\int_{t_0}^{t_f} \gamma^2 dt, \tag{3.1}$$

where γ is the magnitude of the thrust acceleration vector $\mathbf{\Gamma}$, used as control variable. The consumption variable J is a monotonic decreasing function of the mass m of the space vehicle:

$$J = P_{\max}\left(\frac{1}{m} - \frac{1}{m_0}\right), \tag{3.2}$$

where $P_{\max}$ is the maximum power and m_0 is the initial mass. The minimization of the final value of the fuel consumption J_f is equivalent to the maximization of m_f.

The optimization problem concerned with simple transfers (no rendezvous) between coplanar orbits will be formulated as a Mayer problem of optimal control by using Cartesian elements as state variables. At time t, the state of a space vehicle M is defined by the radial distance r from the center of attraction, the radial and transverse components of the velocity, u and v, and the fuel consumption J. (Note that the radial component u should not be confused with the control variables defined in Section 2.) The geometry of the transfer problem is illustrated in Figure 3.1.

In the two-dimension optimization problem, the state equations are given by

$$\begin{aligned}
\frac{du}{dt} &= \frac{v^2}{r} - \frac{\mu}{r^2} + R,\\
\frac{dv}{dt} &= -\frac{uv}{r} + S,\\
\frac{dr}{dt} &= u,\\
\frac{dJ}{dt} &= \frac{1}{2}(R^2 + S^2),
\end{aligned} \tag{3.3}$$

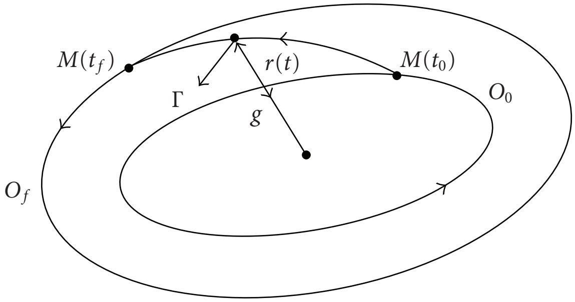


Figure 3.1. Geometry of transfer problem.

where μ is the gravitational parameter (it should not be confused with Lagrange multiplier defined in Section 2), R and S are the components of the thrust acceleration vector in a moving frame of reference, that is, $\mathbf{\Gamma} = R\mathbf{e}_r + S\mathbf{e}_s$, with the unit vector $\mathbf{e}_r$ pointing radially outward and the unit vector $\mathbf{e}_s$ perpendicular to $\mathbf{e}_r$ in the direction of the motion and in the plane of orbit. The optimization problem is stated as follows: it is proposed to transfer a space vehicle M from the initial conditions at t_0,

$$u(0) = 0, \qquad v(0) = 1, \qquad r(0) = 1, \qquad J(0) = 0, \tag{3.4}$$

to the final state at the prescribed final time t_f,

$$u(t_f) = 0, \qquad v(t_f) = \sqrt{\frac{\mu}{r_f}}, \qquad r(t_f) = r_f, \tag{3.5}$$

such that J_f is a minimum.

We note that in the formulation of the boundary conditions above all variables are dimensionless. In this case, $\mu = 1$.

3.2. Two-point boundary value problem. Following the Pontryagin maximum principle [21, 22], the adjoint variables λ_u, λ_v, λ_r, and λ_J are introduced and the Hamiltonian function $H(u,v,r,J,\lambda_u,\lambda_v,\lambda_r,\lambda_J,R,S)$ is formed using the right-hand side of (3.3):

$$H = \lambda_u\left(\frac{v^2}{r} - \frac{\mu}{r^2} + R\right) + \lambda_v\left(-\frac{uv}{r} + S\right) + \lambda_r u + \frac{\lambda_J}{2}\left(R^2 + S^2\right). \tag{3.6}$$

The control variables R and S must be selected from the admissible controls such that the Hamiltonian function reaches its maximum along the optimal trajectory. Thus, we find that

$$\begin{aligned} R^* &= -\frac{\lambda_u}{\lambda_J} \\ S^* &= -\frac{\lambda_v}{\lambda_J}. \end{aligned} \tag{3.7}$$

The variables λ_u, λ_v, λ_r, and λ_J must satisfy the adjoint differential equations and the transversality conditions (2.6) and (2.9).

Therefore, from (3.3)–(3.7), we get the following two-point boundary value problem for the transfer problem defined by (3.3)–(3.5):

$$\begin{aligned}
&\frac{du}{dt} = \frac{v^2}{r} - \frac{\mu}{r^2} - \frac{\lambda_u}{\lambda_J}, \qquad \frac{dv}{dt} = -\frac{uv}{r} - \frac{\lambda_v}{\lambda_J},\\
&\frac{dr}{dt} = u, \qquad \frac{dJ}{dt} = \frac{1}{2\lambda_J^2}\left(\lambda_u^2 + \lambda_v^2\right),\\
&\frac{d\lambda_u}{dt} = \frac{v}{r}\lambda_v - \lambda_r, \qquad \frac{d\lambda_v}{dt} = -2\frac{v}{r}\lambda_u + \frac{u}{r}\lambda_v,\\
&\frac{d\lambda_r}{dt} = \left(\frac{v^2}{r^2} - 2\frac{\mu}{r^3}\right)\lambda_u - \frac{uv}{r^2}\lambda_v, \qquad \frac{d\lambda_J}{dt} = 0,
\end{aligned} \tag{3.8}$$

with the boundary conditions given by (3.4) and (3.5), and the transversality condition

$$\lambda_J(t_f) = -1. \tag{3.9}$$

3.3. Applying the neighboring extremal algorithm. The matrices A, B, and C describing the linearized two-point boundary value problem in the neighboring extremal algorithm can be obtained straightforwardly from (3.8) by calculating the partial derivatives of the right-hand side of the equation with respect to the state and adjoint variables taking into account a diagonal weighting matrix W_2 as described in the next paragraph. These matrices are then given by

$$A = \begin{bmatrix} 0 & \frac{2v}{r} & -a & 0 \\ -\frac{v}{r} & -\frac{u}{r} & \frac{uv}{r^2} & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} -\frac{1}{\lambda_J - \sigma} & 0 & 0 & \frac{\lambda_u}{\lambda_J(\lambda_J - \sigma)} \\ 0 & -\frac{1}{(\lambda_J - \sigma)} & 0 & \frac{\lambda_v}{\lambda_J(\lambda_J - \sigma)} \\ 0 & 0 & 0 & 0 \\ \frac{\lambda_u}{\lambda_J(\lambda_J - \sigma)} & \frac{\lambda_v}{\lambda_J(\lambda_J - \sigma)} & 0 & -\frac{c}{\lambda_J^2(\lambda_J - \sigma)} \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & \frac{\lambda_v}{r} & -\frac{v\lambda_v}{r^2} & 0 \\ \frac{\lambda_v}{r} & -\frac{2\lambda_u}{r} & b & 0 \\ -\frac{v\lambda_v}{r^2} & b & d & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{3.10}$$

where

$$a = \frac{v^2}{r^2} - 2\frac{\mu}{r^3}, \qquad b = \frac{2v}{r^2}\lambda_u - \frac{u}{r^2}\lambda_v,$$
$$c = \lambda_u^2 + \lambda_v^2, \qquad d = \left(-\frac{2v^2}{r^3} + \frac{6\mu}{r^4}\right)\lambda_u + \frac{2uv}{r^3}\lambda_v. \tag{3.11}$$

In Section 5, we present the results obtained through the neighboring extremal algorithm for several ratios $\rho = r_f/r_0$, $\rho = 0.727; 0.800; 0.900; 0.950; 0.975; 1.025; 1.050; 1.100; 1.200; 1.523$, and nondimensional transfer durations of 2, 3, 4, 5. We note that the Earth-Mars transfer corresponds to $\rho = 1.523$ and Earth-Venus to $\rho = 0.727$. The criterion adopted for convergence is a tolerance of 1.0×10^{-8} in the computation of corrections (variations) of the initial value of the adjoint variables. In view of this convergence criterion, the terminal constraints are obtained with an error less than 1.0×10^{-6}, which means that $\|\psi(x(t_f))\| \le 1.0 \times 10^{-6}$. All simulations consider the transfer from low orbit to high orbit, with starting approximation $\lambda^0(t_0) = (0.001 \quad 0.001 \quad 0.001 \quad -1)$, attenuation factor $k = 0.10$ for $\rho = 0.727$ and 1.5236, and $k = 0.15$ for the other values of ρ, and a diagonal matrix W_2 such that $W_{2_{11}} = W_{2_{22}} = -\sigma$, with $\sigma = 5.5$ for all maneuvers, except $\rho = 0.727$ and 0.800, with $t_f - t_0 = 4$ and 5. In these cases, $\sigma = 2.5$.

4. Linear theory

In this section, a first-order analytical solution for the problem of optimal simple transfer defined in Section 3.1 is presented.

The Hamiltonian function H^* governing the extremal (optimal) trajectories can be obtained as follows. Since λ_J is a first integral (3.8) and $\lambda_J(t_f) = -1$, from the transversality condition (3.9), it follows that $\lambda_J(t) = -1$. Thus, from (3.7), we find the optimal thrust acceleration

$$R^* = \lambda_u, \qquad S^* = \lambda_v. \tag{4.1}$$

Introducing these equations into (3.6), it results that

$$H^* = u\lambda_r + \left(\frac{v^2}{r} - \frac{\mu}{r^2}\right)\lambda_u - \frac{uv}{r}\lambda_v + \frac{1}{2}(\lambda_u^2 + \lambda_v^2). \tag{4.2}$$

In what follows, we consider the problem of determining an approximate solution of the system of differential equations governed by the Hamiltonian H^* by means of the theory of canonical transformations. This analytical solution is obtained through the canonical transformation theory using the concept of generalized canonical systems [24, 25].

Now, consider the Hamiltonian function describing a null thrust arc in the two-dimensional formulation of the optimization problem defined in Section 3.1:

$$H = u\lambda_r + \left(\frac{v^2}{r} - \frac{\mu}{r^2}\right)\lambda_u - \frac{uv}{r}\lambda_v + \frac{v}{r}\lambda_\theta. \tag{4.3}$$

Note that H is obtained from (3.6) taking $R = S = 0$ and adding the last term concerning the differential equation of the angular variable θ, which defines the position of the space vehicle with respect to a reference axis in the plane of motion. This variable is important for rendezvous problems and plays no special role for simple transfer problems like the one considered here, but it is necessary to define the canonical transformations described below. In the transformation theory described in the next paragraphs, it is assumed that the Hamiltonian H^* is augmented in order to include this last term, that is,

$$H^* = u\lambda_r + \left(\frac{v^2}{r} - \frac{\mu}{r^2}\right)\lambda_u - \frac{uv}{r}\lambda_v + \frac{v}{r}\lambda_\theta + \frac{1}{2}(\lambda_u^2 + \lambda_v^2). \tag{4.4}$$

Note that H is the undisturbed part of H^* and plays a fundamental role in our theory.

According to the properties of generalized canonical systems, the general solution of the system of differential equations governed by the Hamiltonian H can be expressed in terms of a fast phase and is given by [24, 25]:

$$\begin{aligned}
u &= \sqrt{\frac{\mu}{p}}e\sin f,\\
v &= \sqrt{\frac{\mu}{p}}(1 + e\cos f),\\
r &= \frac{p}{1 + e\cos f},\\
\theta &= \omega + f,\\
\lambda_u &= \sqrt{\frac{p}{\mu}}\sin f\,\lambda_e + \sqrt{\frac{p}{\mu}}\frac{\cos f}{e}(\lambda_f - \lambda_\omega)\\
\lambda_v &= 2\sqrt{\frac{p}{\mu}}r\lambda_p + \sqrt{\frac{p}{\mu}}(2\cos f + e\cos^2 f + e)\frac{r}{p}\lambda_e - \sqrt{\frac{p}{\mu}}\frac{\sin f}{e}\left[1 + \frac{r}{p}\right](\lambda_f - \lambda_\omega),\\
\lambda_r &= 2\frac{p}{r}\lambda_p + \frac{\cos f + e}{r}\lambda_e - \frac{\sin f}{re}(\lambda_f - \lambda_\omega),\\
\lambda_\theta &= \lambda_\omega,
\end{aligned} \tag{4.5}$$

where p is the semi latus rectum, e is the eccentricity, ω is the pericenter argument, f is the true anomaly (fast phase), and $(\lambda_p, \lambda_e, \lambda_f, \lambda_\omega)$ are adjoint variables to (p, e, f, ω).

Equations (4.5) define a Mathieu transformation between the Cartesian elements and the orbital ones,

$$(u, v, r, \theta, \lambda_u, \lambda_v, \lambda_r, \lambda_\theta) \xrightarrow{\text{Mathieu}} (p, e, f, \omega, \lambda_p, \lambda_e, \lambda_f, \lambda_\omega). \tag{4.6}$$

The undisturbed Hamiltonian function H is invariant with respect to this canonical transformation. Thus, the undisturbed Hamiltonian is written in terms of the new variables as

$$H = \frac{\sqrt{\mu p}}{r^2}\lambda_f. \tag{4.7}$$

The general solution of the new differential equations governed by the new undisturbed Hamiltonian function H is closely related to the solution of the time flight equation in the two-body problem for elliptic, parabolic, and hyperbolic motions [25]. For quasicircular motions, this solution is very simple, as described in the next paragraphs.

Equations (4.5) have singularities for circular orbits ($e = 0$). In order to avoid this drawback, a set of nonsingular elements is introduced. The relationships between the singular orbital elements and the nonsingular ones are given by

$$a = \frac{p}{(1-e^2)}, \qquad h = e\cos\omega, \qquad k = e\sin\omega, \qquad L = f + \omega. \tag{4.8}$$

These equations define a Lagrange point transformation and the Jacobian of the inverse of this transformation must be computed in order to get the relationships between the corresponding adjoint variables. Thus, we get

$$\begin{aligned}
\lambda_a &= (1-e^2)\lambda_p,\\
\lambda_h &= \left(\lambda_e - \frac{2ep}{(1-e^2)}\lambda_p\right)\cos\omega + \left(\frac{\lambda_L - \lambda_\omega}{e}\right)\sin\omega,\\
\lambda_k &= \left(\lambda_e - \frac{2ep}{(1-e^2)}\lambda_p\right)\sin\omega - \left(\frac{\lambda_L - \lambda_\omega}{e}\right)\cos\omega,\\
\lambda_L &= \lambda_f.
\end{aligned} \tag{4.9}$$

Equations (4.8) and (4.9) define a new Mathieu transformation between singular and nonsingular elements,

$$(p,e,f,\omega,\lambda_p,\lambda_e,\lambda_f,\lambda_\omega) \xrightarrow{\text{Mathieu}} (a,h,k,L,\lambda_a,\lambda_h,\lambda_k,\lambda_L). \tag{4.10}$$

Substituting (4.8) and (4.9) into (4.5), we get

$$\begin{aligned}
u &= \sqrt{\frac{\mu}{a(1-h^2-k^2)}}(h\sin L - k\cos L),\\
v &= \sqrt{\frac{\mu}{a(1-h^2-k^2)}}(1 + h\cos L + k\sin L),\\
r &= \frac{a(1-h^2-k^2)}{1 + h\cos L + k\sin L},\\
\theta &= L,
\end{aligned}$$

$$\begin{aligned}
\lambda_u &= \sqrt{\frac{a}{\mu}}\left\{2a\lambda_a\frac{(h\sin L-k\cos L)}{\sqrt{1-h^2-k^2}}+\sqrt{1-h^2-k^2}\,(\lambda_h\sin L-\lambda_k\cos L)\right\},\\
\lambda_v &= \sqrt{\frac{a}{\mu}}\left\{2a\lambda_a\sqrt{1-h^2-k^2}\left(\frac{a}{r}\right)\right.\\
&\qquad+\frac{1}{\sqrt{1-h^2-k^2}}\left(\frac{r}{a}\right)\left\{\left[\frac{3}{2}h+2\cos L+\frac{h}{2}\cos 2L+\frac{k}{2}\sin 2L\right]\lambda_h\right.\\
&\qquad\qquad\left.\left.+\left[\frac{3}{2}k+2\sin L-\frac{k}{2}\cos 2L+\frac{h}{2}\sin 2L\right]\lambda_k\right\}\right\},\\
\lambda_r &= 2\left(\frac{a}{r}\right)^2\lambda_a+\frac{1}{r}\left[(h+\cos L)\lambda_h+(k+\sin L)\lambda_k\right],\\
\lambda_\theta &= -k\lambda_h+h\lambda_k+\lambda_L.
\end{aligned}\tag{4.11}$$

These equations are valid for all orbits and define a Mathieu transformation between the Cartesian elements and the nonsingular orbital elements.

For quasicircular orbits, with very small eccentricities, (4.11) can be greatly simplified if higher-order terms in eccentricity are neglected. Thus,

$$\begin{aligned}
u &= na(h\sin\ell-k\cos\ell),\\
v &= na(1+h\cos\ell+k\sin\ell),\\
r &= \frac{a}{1+h\cos\ell+k\sin\ell},\\
\theta &= \ell\\
\lambda_u &= \frac{1}{na}(\lambda_h\sin\ell-\lambda_k\cos\ell),\\
\lambda_v &= \frac{2}{na}[a\lambda_a+(\lambda_h\cos\ell+\lambda_k\sin\ell)],\\
\lambda_r &= 2\lambda_a+\frac{1}{a}(\lambda_h\cos\ell+\lambda_k\sin\ell),\\
\lambda_\theta &= \lambda_\ell,
\end{aligned}\tag{4.12}$$

where $n=\sqrt{\mu/a^3}$ is the mean motion and $\ell=\omega+M$ is the mean latitude. We note that first-order terms in eccentricity are retained in the state variables in order to get a trajectory with better accuracy. For adjoint variables, this is unnecessary since λ_a, λ_h, and λ_k are small quantities for transfers between close circular orbits, that is, for small amplitude transfers.

Introducing (4.12) into the expression for H^*, we finally get

$$H^* = n\lambda_\ell + \frac{1}{2n^2a^2}\left\{4a^2\lambda_a^2 + \frac{5}{2}(\lambda_h^2+\lambda_k^2) + 8a\lambda_a\lambda_k\sin\ell + 8a\lambda_a\lambda_h\cos\ell \right.$$
$$\left. +3\lambda_h\lambda_k\sin 2\ell + \frac{3}{2}(\lambda_h^2-\lambda_k^2)\cos 2\ell\right\}. \tag{4.13}$$

For transfers between close circular coplanar orbits, an approximate solution of the system of differential equations governed by H^* can be obtained through simple integrations if the system is linearized about a reference circular orbit $\overline{O}$ with semimajor axis $\overline{a}$. This solution can be put in the form

$$\Delta x = A\lambda_0, \tag{4.14}$$

where $\Delta x = [\Delta\alpha \quad \Delta h \quad \Delta k]^T$ denotes the imposed changes on nonsingular orbital elements (state variables), $\alpha = a/\overline{a}$, $\lambda_\alpha = \overline{a}\lambda_a$, λ_0 is the 3×1 vector of initial values of the adjoint variables, and A is a 3×3 symmetric matrix. The adjoint variables are constant and the matrix A is given by

$$A = \begin{bmatrix} a_{\alpha\alpha} & a_{\alpha h} & a_{\alpha k} \\ a_{h\alpha} & a_{hh} & a_{hk} \\ a_{k\alpha} & a_{kh} & a_{kk} \end{bmatrix}, \tag{4.15}$$

where

$$a_{\alpha\alpha} = 4\sqrt{\frac{\overline{a}^5}{\mu^3}}\Delta\overline{\ell}, \tag{4.16}$$

$$a_{\alpha h} = a_{h\alpha} = 4\sqrt{\frac{\overline{a}^5}{\mu^3}}(\sin\overline{\ell}_f - \sin\overline{\ell}_0), \tag{4.17}$$

$$a_{\alpha k} = a_{k\alpha} = -4\sqrt{\frac{\overline{a}^5}{\mu^3}}(\cos\overline{\ell}_f - \cos\overline{\ell}_0), \tag{4.18}$$

$$a_{hh} = \sqrt{\frac{\overline{a}^5}{\mu^3}}\left[\frac{5}{2}\Delta\overline{\ell} + \frac{3}{4}(\sin 2\overline{\ell}_f - \sin 2\overline{\ell}_0)\right], \tag{4.19}$$

$$a_{hk} = a_{kh} = -\frac{3}{4}\sqrt{\frac{\overline{a}^5}{\mu^3}}(\cos 2\overline{\ell}_f - \cos 2\overline{\ell}_0), \tag{4.20}$$

$$a_{kk} = \sqrt{\frac{\overline{a}^5}{\mu^3}}\left[\frac{5}{2}\Delta\overline{\ell} - \frac{3}{4}(\sin 2\overline{\ell}_f - \sin 2\overline{\ell}_0)\right]. \tag{4.21}$$

Subscript f stands for the final time, $\overline{\ell} = \overline{\ell}_0 + \overline{n}(t - t_0)$, and t_0 is the initial time. The linear solution, described by (4.14)–(4.21), is in agreement with the one presented in [1, 18], where it is obtained through a different approach.

In view of (4.1) and (4.12), the optimal thrust acceleration $\mathbf{\Gamma}^*$ is expressed by

$$\mathbf{\Gamma}^* = \frac{1}{\overline{n}\overline{a}}\{(\lambda_h \sin\overline{\ell} - \lambda_k \cos\overline{\ell})\mathbf{e}_r + 2(\lambda_\alpha + \lambda_h \cos\overline{\ell} + \lambda_k \sin\overline{\ell})\mathbf{e}_s\}. \tag{4.22}$$

The variation of the consumption variable ΔJ during the maneuver can be obtained straightforwardly from (4.13) and (4.15) by integrating, from t_0 to t_f, the differential equation (see (3.3), (4.1), and (4.13))

$$\frac{dJ}{dt} = H_\gamma^*, \tag{4.23}$$

where H_γ^* denotes the part of the Hamiltonian H^* concerned with the thrust acceleration. Thus,

$$\Delta J = \frac{1}{2}\{a_{\alpha\alpha}\lambda_\alpha^2 + 2a_{\alpha h}\lambda_\alpha\lambda_h + 2a_{\alpha k}\lambda_\alpha\lambda_k + a_{hh}\lambda_h^2 + 2a_{hk}\lambda_h\lambda_k + a_{kk}\lambda_k^2\}, \tag{4.24}$$

where $a_{\alpha\alpha}, a_{\alpha h}, \ldots, a_{kk}$ are given by (4.16)–(4.21); and λ_α, λ_h, and λ_k are obtained from the solution of the linear algebraic system defined by (4.14).

We recall that the extremal (optimal) trajectory is given by (4.12) with the nonsingular elements a, h, and k calculated from (4.14).

For transfers between circular orbits, only $\Delta\alpha$ is imposed. If it is assumed that the initial and final positions of the vehicle in orbit are symmetric with respect to x-axis of the inertial reference frame, that is, $\overline{\ell}_f = -\overline{\ell}_0 = \Delta\ell/2$, the solution of the system (4.14) is given by

$$\begin{aligned}
\lambda_\alpha &= \frac{1}{2}\sqrt{\frac{\mu^3}{\overline{a}^5}}\left\{\frac{\Delta\alpha(5\Delta\overline{\ell} + 3\sin\Delta\overline{\ell})}{10\Delta\overline{\ell}^2 + 6\Delta\overline{\ell}\sin\Delta\overline{\ell} - 64\sin^2(\Delta\overline{\ell}/2)}\right\},\\
\lambda_h &= -\sqrt{\frac{\mu^3}{\overline{a}^5}}\left\{\frac{8\Delta\alpha\sin(\Delta\overline{\ell}/2)}{10\Delta\overline{\ell}^2 + 6\Delta\overline{\ell}\sin\Delta\overline{\ell} - 64\sin^2(\Delta\overline{\ell}/2)}\right\},\\
\lambda_k &= 0.
\end{aligned} \tag{4.25}$$

We note that the linear theory is applicable only to orbits which are not separated by large radial distance, that is, to transfers between close orbits. If the reference orbit is chosen in the conventional way, that is, with the semimajor axis as the radius of the initial orbit, the radial excursion to the final orbit will be maximized [26]. A better reference orbit is defined with a semimajor axis given by an intermediate value between the values of semimajor axes of the terminal orbits. In our analysis, we have chosen $\overline{a} = (a_0 + a_f)/2$ in order to improve the accuracy in the calculations.

In the next section, the results of the linear theory are compared to the ones provided by the neighboring extremal algorithm described in Sections 2 and 3.

Table 5.1. Consumption variable J $(\rho > 1)$.

ρ	$t_f - t_0$	J_{anal}	J_{grad}	J_{neigh}	d_{rel1}	d_{rel2}
1.0250	2.0	3.5856×10^{-4}	3.5855×10^{-4}	3.5854×10^{-4}	0.00	0.00
	3.0	8.4459×10^{-5}	8.4462×10^{-5}	8.4456×10^{-5}	0.00	0.01
	4.0	3.1226×10^{-5}	3.1233×10^{-5}	3.1230×10^{-5}	0.01	0.01
	5.0	1.7138×10^{-5}	1.7147×10^{-5}	1.7143×10^{-5}	0.03	0.02
1.0500	2.0	1.4463×10^{-3}	1.4463×10^{-3}	1.4459×10^{-3}	0.03	0.03
	3.0	3.4169×10^{-4}	3.4166×10^{-4}	3.4164×10^{-4}	0.02	0.01
	4.0	1.2533×10^{-4}	1.2538×10^{-4}	1.2537×10^{-4}	0.03	0.00
	5.0	6.7541×10^{-5}	6.7611×10^{-5}	6.7598×10^{-5}	0.08	0.02
1.1000	2.0	5.8778×10^{-3}	5.8741×10^{-3}	5.8716×10^{-3}	0.11	0.04
	3.0	1.3977×10^{-3}	1.3970×10^{-3}	1.3969×10^{-3}	0.06	0.00
	4.0	5.0619×10^{-4}	5.0666×10^{-4}	5.0664×10^{-4}	0.09	0.00
	5.0	2.6374×10^{-4}	2.6453×10^{-4}	2.6451×10^{-4}	0.29	0.01
1.2000	2.0	2.4187×10^{-2}	2.4097×10^{-2}	2.4097×10^{-2}	0.37	0.00
	3.0	5.8370×10^{-3}	5.8200×10^{-3}	5.8199×10^{-3}	0.29	0.00
	4.0	2.0813×10^{-3}	2.0845×10^{-3}	2.0844×10^{-3}	0.15	0.00
	5.0	1.0260×10^{-3}	1.0346×10^{-3}	1.0345×10^{-3}	0.83	0.00
1.5236	2.0	1.7743×10^{-1}	1.7434×10^{-1}	1.7434×10^{-1}	1.77	0.00
	3.0	4.4947×10^{-2}	4.4067×10^{-2}	4.4066×10^{-2}	1.99	0.00
	4.0	1.6051×10^{-2}	1.5889×10^{-2}	1.5889×10^{-2}	1.02	0.00
	5.0	7.2498×10^{-3}	7.3352×10^{-3}	7.3351×10^{-3}	1.17	0.00

5. Results

The values of the consumption variable J computed through the neighboring extremal algorithm and the ones provided by the linear theory and by a numerical method based on gradient techniques [19] are presented in Tables 5.1 and 5.2, and are plotted in Figures 5.1 and 5.2, as function of the radius ratio ρ of the terminal orbits for various transfer durations $t = t_f - t_0$. The absolute relative difference in percent between the numerical and analytical results is also presented in the tables according to the following definition:

$$d_{\text{rel1}} = \left| \frac{J_{\text{neigh}} - J_{\text{linear}}}{J_{\text{neigh}}} \right| \times 100\%,$$
$$d_{\text{rel2}} = \left| \frac{J_{\text{neigh}} - J_{\text{grad}}}{J_{\text{neigh}}} \right| \times 100\%. \tag{5.1}$$

Tables 5.1 and 5.2 show that $d_{\text{rel1}} < 2\%$ for $\rho > 1$, and $d_{\text{rel1}} < 5.5\%$ for $\rho < 1$. The greater values corresponds to transfers with moderate amplitude $\rho = 0.7270$ and $\rho = 1.5236$. On the other hand, $d_{\text{rel2}} < 0.04\%$ for all cases.

Tables 5.1 and 5.2, and Figures 5.1 and 5.2 show the good agreement between the results. Note that the linear theory provides a good approximation for the solution of the

Table 5.2. Consumption variable J ($\rho < 1$).

ρ	$t_f - t_0$	J_{linear}	J_{grad}	J_{neigh}	d_{rel1}	d_{rel2}
0.7270	2.0	3.7654×10^{-2}	3.7299×10^{-2}	3.7298×10^{-2}	0.95	0.00
	3.0	8.9269×10^{-3}	9.0261×10^{-3}	9.0259×10^{-3}	1.10	0.00
	4.0	4.0482×10^{-3}	4.2133×10^{-3}	4.2131×10^{-3}	3.91	0.00
	5.0	2.8941×10^{-3}	3.0573×10^{-3}	3.0572×10^{-3}	5.33	0.00
0.8000	2.0	2.0951×10^{-2}	2.0842×10^{-2}	2.0842×10^{-2}	0.52	0.00
	3.0	4.9040×10^{-3}	4.9173×10^{-3}	4.9172×10^{-3}	0.27	0.00
	4.0	2.0703×10^{-3}	2.1047×10^{-3}	2.1046×10^{-3}	1.63	0.00
	5.0	1.3838×10^{-3}	1.4198×10^{-3}	1.4197×10^{-3}	2.53	0.00
0.9000	2.0	5.4740×10^{-3}	5.4672×10^{-3}	5.4671×10^{-3}	0.13	0.00
	3.0	1.2771×10^{-3}	1.2772×10^{-3}	1.2771×10^{-3}	0.00	0.01
	4.0	5.0063×10^{-4}	5.0198×10^{-4}	5.0198×10^{-4}	0.27	0.00
	5.0	3.0496×10^{-4}	3.0653×10^{-4}	3.0652×10^{-4}	0.51	0.00
0.9500	2.0	1.3958×10^{-3}	1.3955×10^{-3}	1.3955×10^{-3}	0.02	0.00
	3.0	3.2649×10^{-4}	3.2649×10^{-4}	3.2647×10^{-4}	0.01	0.01
	4.0	1.2451×10^{-4}	1.2459×10^{-4}	1.2458×10^{-4}	0.06	0.00
	5.0	7.2585×10^{-5}	7.2671×10^{-5}	7.2667×10^{-5}	0.11	0.01
0.9750	2.0	3.5225×10^{-4}	3.5231×10^{-4}	3.5223×10^{-4}	0.01	0.02
	3.0	8.2555×10^{-5}	8.2560×10^{-5}	8.2553×10^{-5}	0.00	0.01
	4.0	3.1120×10^{-5}	3.1126×10^{-5}	3.1124×10^{-5}	0.01	0.01
	5.0	1.7765×10^{-5}	1.7772×10^{-5}	1.7771×10^{-5}	0.03	0.00

low-thrust limited-power transfer between close circular coplanar orbits in an inverse-square force field.

Figures 5.1 and 5.2 also show that the fuel consumption can be greatly reduced if the duration of the transfer is increased. The fuel consumption for transfers with duration $t_f - t_0 = 2$ is approximately ten times the fuel consumption for a transfer with duration $t_f - t_0 = 4$.

In order to follow the evolution of the optimal thrust acceleration vector during the transfer, it is convenient to plot the locus of its tip in the moving frame of reference. Figures 5.3 and 5.4 illustrate these plots for small amplitude transfers with $\rho = 0.950$ and 1.050, and for moderate amplitude transfers, Earth-Mars ($\rho = 1.523$), and Earth-Venus ($\rho = 0.727$) transfers, with $t_f - t_0 = 2$. It should be noted that the agreement between the numerical and analytical results is better for small amplitude transfers. For moderate amplitude transfers, this difference increases with the duration of the maneuvers.

Figures 5.5 and 5.6 show the time history of the state variables—u, v, and r—for a small amplitude transfer, $\rho = 1.050$, and a moderate amplitude transfer, $\rho = 1.523$, with $t_f - t_0 = 2$. Again, the agreement between the numerical and analytical results is better for small amplitude transfers.

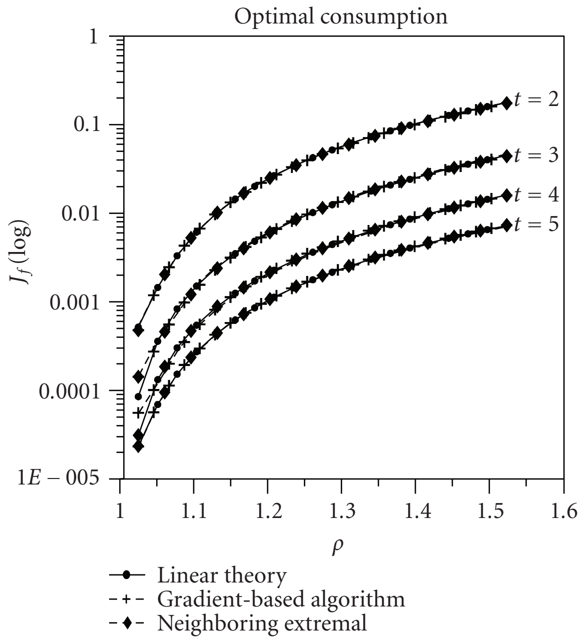


Figure 5.1. Consumption variable J for $\rho > 1$.

The results—consumption variable, thrust acceleration, and trajectory—provided by the gradient-based algorithm are quite similar to the ones provided by the neighboring extremal algorithm.

It should be noted that the numerical algorithms based on the second variation theory—gradient-based algorithm and the neighboring extremal algorithm—provide quite similar results. This fact leads us to suppose that the solutions provided by the both algorithms are really optimal in the sense of a local minimum for the consumption variable J, although the sufficiency conditions are not tested. Besides, we note that the Pontryagin maximum principle is a set of necessary and sufficient conditions for the linearized problem describing the transfers between close circular orbits [26].

6. Conclusion

In this paper, a numerical and analytical study of optimal low-thrust limited-power trajectories for simple transfer (no rendezvous) between close circular coplanar orbits in an inverse-square force field is presented. The numerical study is carried out by means of a neighboring extremal algorithm and the analytical one is based on linear theory obtained through canonical transformation theory, using the concept of generalized canonical systems. The numerical and analytical results have been compared to the ones obtained through a numerical method based on gradient techniques. The great agreement between the results shows that the linear theory provides a good approximation for the

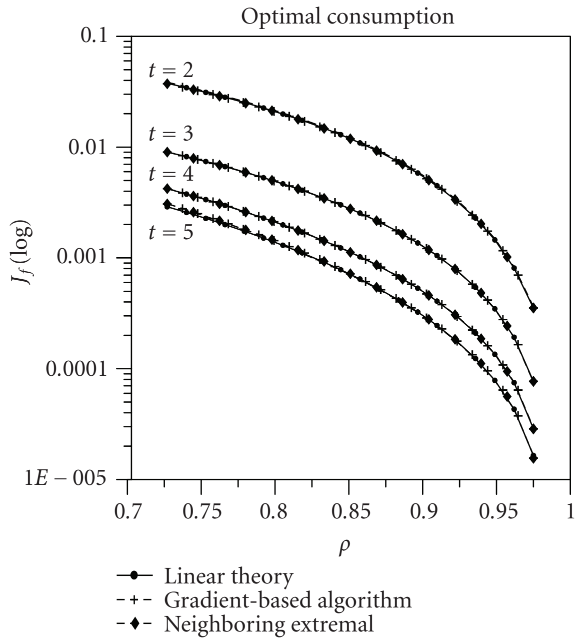


Figure 5.2. Consumption variable J for $\rho < 1$.

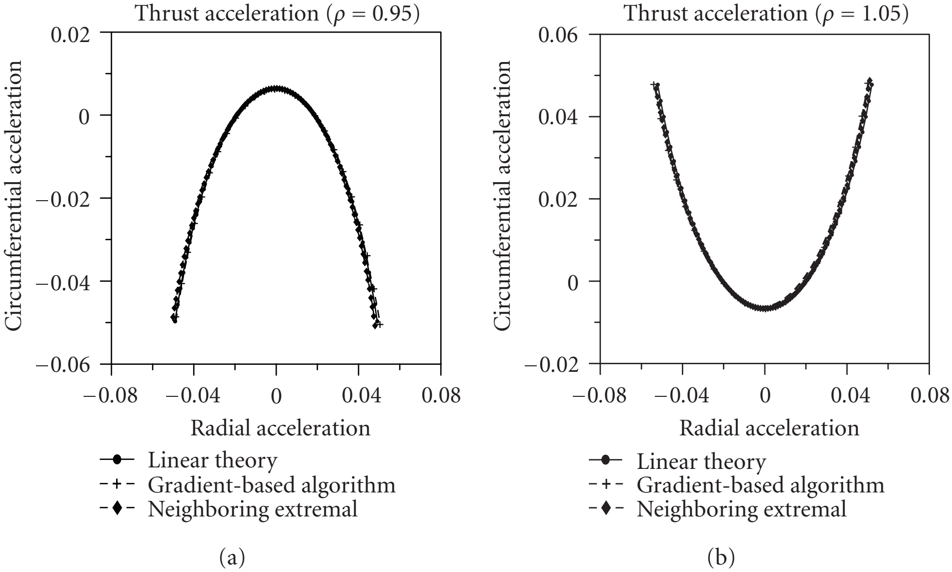


Figure 5.3. Thrust acceleration for $t_f - t_0 = 2$ (transfers with small amplitude).

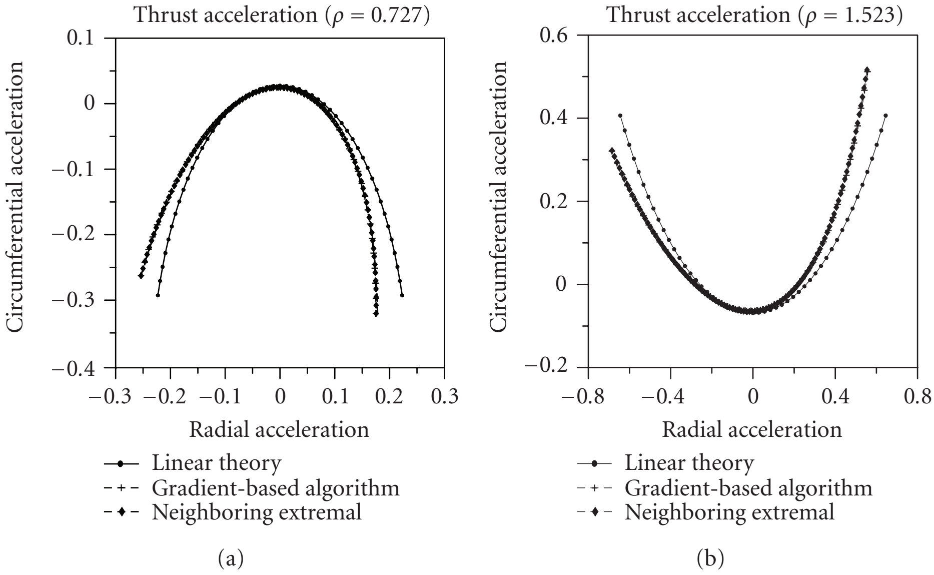


Figure 5.4. Thrust acceleration for $t_f - t_0 = 2$ (transfers with moderate amplitude).

solution of the transfer problem and it can be used in preliminary mission analysis concerning close coplanar circular orbits. On the other hand, the good performance of the algorithms based on the second variation theory shows that they are also good tools in determining optimal low-thrust limited-power trajectories.

Appendix

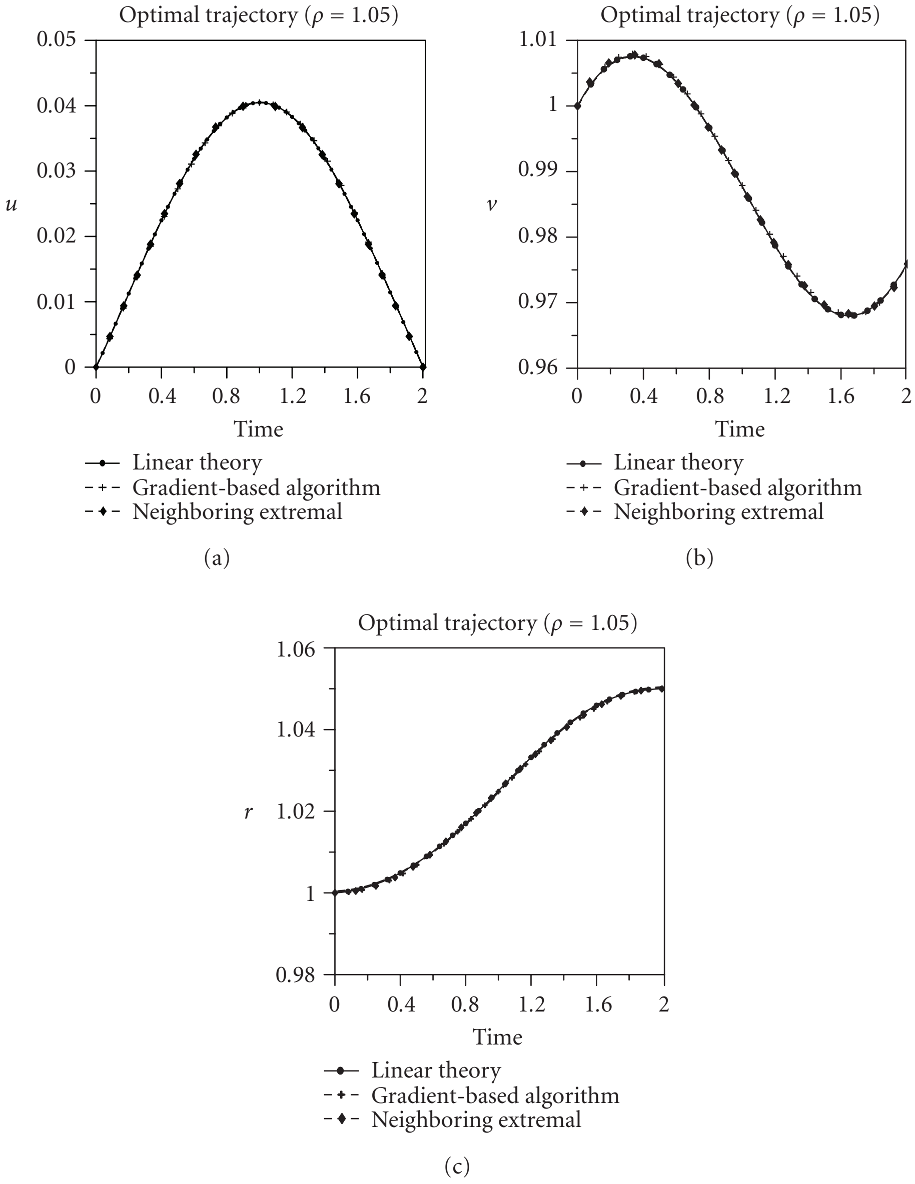
In this appendix, the modified accessory minimum problem is described. Consider the Bolza problem formulated in Section 2. Introducing the adjoint vector $\lambda(t)$ and the vector of Lagrange multipliers μ, the augmented performance index $\overline{J}$ is formed using (2.3) and (2.4),

$$\overline{J} = g(x(t_f)) + \mu^T \psi(x(t_f)) + \int_{t_0}^{t_f} [-H(x,\lambda,u) + \lambda^T \dot{x}]dt, \tag{A.1}$$

where H is the Hamiltonian function previously introduced in Section 2.

Now, consider the expansion of $\overline{J}$ to second-order and the constraints, defined by (2.2)–(2.4), to first order. Taking into account that all first-order terms vanish about a nominal extremal solution (see (2.5)–(2.10)), one finds that

$$\delta^2\overline{J} = \frac{1}{2}\delta x^T(t_f)(g_{xx} + \mu^T \psi_{xx})\delta x(t_f) - \frac{1}{2}\int_{t_0}^{t_f} \{\delta x^T H_{xx}\delta x + 2\delta u^T H_{ux}\delta x + \delta u^T H_{uu}\delta u\} dt, \tag{A.2}$$



(a)

(b)

(c)

Figure 5.5. Time history of state variables for $t_f - t_0 = 2$ and $\rho = 1.050$.

$$\delta\dot{x} = H_{\lambda x}\delta x + H_{\lambda u}\delta u, \tag{A.3}$$

$$\delta x(t_0) = 0, \tag{A.4}$$

$$\psi_x \delta x(t_f) = -k\psi. \tag{A.5}$$

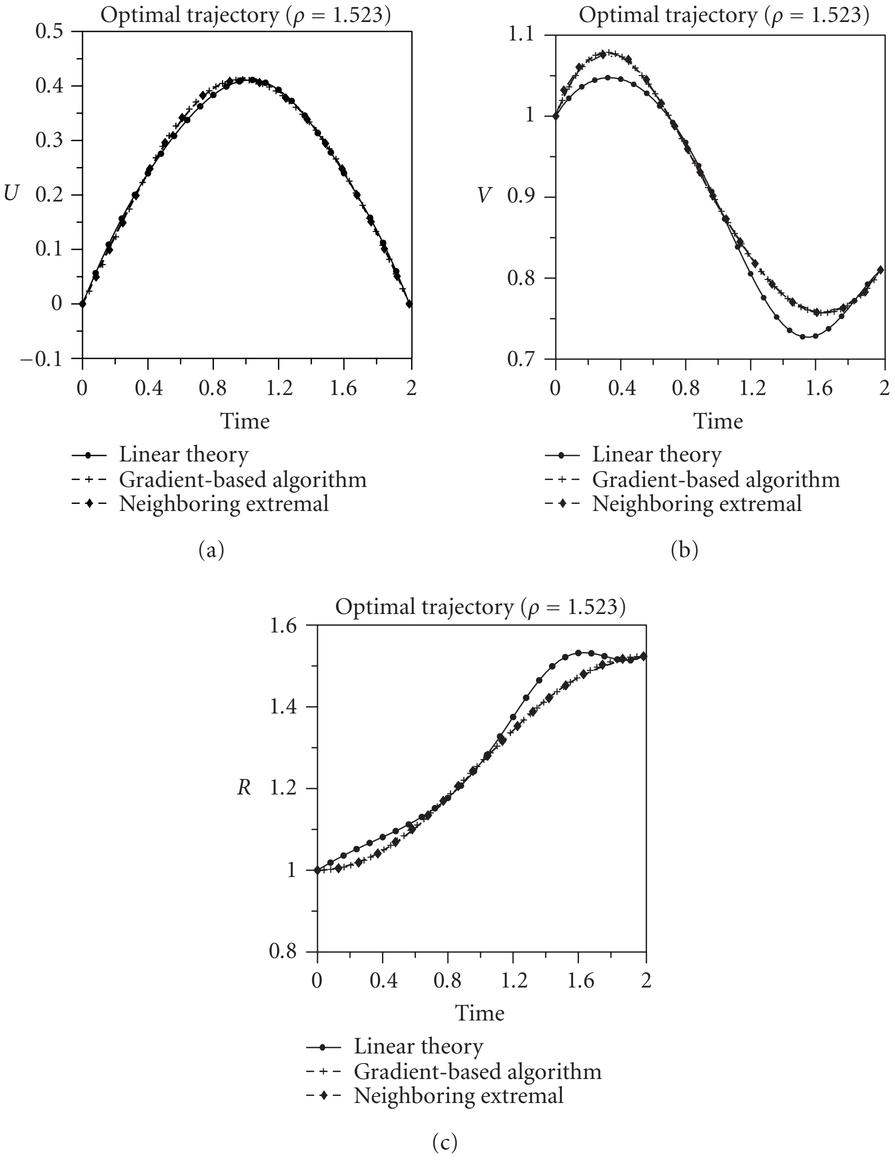


Figure 5.6. Time history of state variables for $t_f - t_0 = 2$ and $\rho = 1.523$.

All quantities, $H_{xx}, H_{\lambda x}, g_{xx}, \psi, \ldots$, in equations above, are evaluated on a nominal extremal solution and k is defined in Section 2.

In order to assure that the expansion above is valid, a constraint is imposed on the control variations:

$$\frac{1}{2}\int_{t_0}^{t_f} \delta u^T W \delta u \, dt = M, \tag{A.6}$$

where $W(t)$ is an arbitrary $m \times m$ positive-definite weighting matrix and $M > 0$ is an arbitrary prescribed value.

Consider the following Bolza problem: determine δu such that $\delta^2 \overline{J}$ is minimized subject to the constraints (A.3) through (A.6). In view of the imposed constraint on the control variations defined by (A.6), this minimization problem is referred to as modified accessory minimum problem. By applying the set of necessary and sufficient conditions to this minimization problem, one finds (2.12) through (2.17), with (2.14) replaced by

$$H_{ux}\delta x + H_{u\lambda}\delta\lambda + (H_{uu} + \alpha W)\delta u = 0, \tag{A.7}$$

where α is a constant Lagrange multiplier associated to the constraint (A.6). Since W and M are arbitrary, α can be included in the choice of matrix W, that is, a new arbitrary $m \times m$ matrix W_2 can be introduced such that $W_2 = \alpha W$. The evaluation of α is unnecessary, as well as the choice of M. Accordingly, (A.7) can be replaced by

$$H_{ux}\delta x + H_{u\lambda}\delta\lambda + (H_{uu} + W_2)\delta u = 0. \tag{A.8}$$

Beside the equations mentioned above, the strengthened Legendre condition must be satisfied, that is,

$$H_{uu} + W_2 < 0. \tag{A.9}$$

Acknowledgment

This research has been partially supported by CNPq under Contract 300450/2003-6.

References

[1] J.-P. Marec, *Optimal Space Trajectories*, Elsevier, New York, NY, USA, 1979.

[2] V. Coverstone-Carroll and S. N. Williams, "Optimal low thrust trajectories using differential inclusion concepts," *Journal of the Astronautical Sciences*, vol. 42, no. 4, pp. 379–393, 1994.

[3] J. A. Kechichian, "Optimal low-thrust rendezvous using equinoctial orbit elements," *Acta Astronautica*, vol. 38, no. 1, pp. 1–14, 1996.

[4] J. A. Kechichian, "Reformulation of Edelbaum's low-thrust transfer problem using optimal control theory," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 5, pp. 988–994, 1997.

[5] J. A. Kechichian, "Orbit raising with low-thrust tangential acceleration in presence of earth shadow," *Journal of Spacecraft and Rockets*, vol. 35, no. 4, pp. 516–525, 1998.

[6] C. A. Kluever and S. R. Oleson, "A direct approach for computing near-optimal low-thrust transfers," in *AAS/AIAA Astrodynamics Specialist Conference*, Sun Valley, Idaho, USA, August 1997, AAS paper 97-717.

[7] A. A. Sukhanov and A. F. B. A. Prado, "Constant tangential low-thrust trajectories near an oblate planet," *Journal of Guidance, Control, and Dynamics*, vol. 24, no. 4, pp. 723–731, 2001.

[8] V. Coverstone-Carroll, J. W. Hartmann, and W. J. Mason, "Optimal multi-objective low-thrust spacecraft trajectories," *Computer Methods in Applied Mechanics and Engineering*, vol. 186, no. 2–4, pp. 387–402, 2000.

[9] M. Vasile, F. B. Zazzera, R. Jehn, and G. Janin, "Optimal interplanetary trajectories using a combination of low-thrust and gravity assist manoeuvres," in *Proceedings of the 51st International Astronautical Congress (IAF '00)*, Rio de Janeiro, Brazil, October 2000, IAF-00-A.5.07.

[10] G. D. Racca, “New challenges to trajectory design by the use of electric propulsion and other new means of wandering in the solar system,” *Celestial Mechanics & Dynamical Astronomy*, vol. 85, no. 1, pp. 1–24, 2003.

[11] T. N. Edelbaum, “Optimum power-limited orbit transfer in strong gravity fields,” *AIAA Journal*, vol. 3, no. 5, pp. 921–925, 1965.

[12] J.-P. Marec and N. X. Vinh, “Optimal low-thrust, limited power transfers between arbitrary elliptical orbits,” *Acta Astronautica*, vol. 4, no. 5-6, pp. 511–540, 1977.

[13] C. M. Hassig, K. D. Mease, and N. X. Vinh, “Minimum-fuel power-limited transfer between coplanar elliptical orbits,” *Acta Astronautica*, vol. 29, no. 1, pp. 1–15, 1993.

[14] S. Geffroy and R. Epenoy, “Optimal low-thrust transfers with constraints—generalization of averaging techniques,” *Acta Astronautica*, vol. 41, no. 3, pp. 133–149, 1997.

[15] B. N. Kiforenko, “Optimal low-thrust orbital transfers in a central gravity field,” *International Applied Mechanics*, vol. 41, no. 11, pp. 1211–1238, 2005.

[16] J. V. Breakwell, J. L. Speyer, and A. E. Bryson Jr., “Optimization and control of nonlinear systems using the second variation,” *SIAM Journal on Control and Optimization*, vol. 1, no. 2, pp. 193–223, 1963.

[17] A. G. Longmuir and E. V. Bohn, “Second-variation methods in dynamic optimization,” *Journal of Optimization Theory and Applications*, vol. 3, no. 3, pp. 164–173, 1969.

[18] J.-P. Marec, *Transferts Optimaux Entre Orbites Elliptiques Proches*, ONERA, Châtillon, France, 1967.

[19] S. da Silva Fernandes and W. A. Golfetto, “Numerical computation of optimal low-thrust limited-power trajectories—transfers between coplanar circular orbits,” *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 27, no. 2, pp. 178–185, 2005.

[20] S. da Silva Fernandes and W. A. Golfetto, “Computation of optimal low-thrust limited-power trajectories through an algorithm based on gradient techniques,” in *Proceedings of the 3rd Brazilian Conference on Dynamics, Control and Their Applications*, pp. 789–796, São Paulo, Brazil, 2004, CD-ROM.

[21] A. E. Bryson Jr. and Y. C. Ho, *Applied Optimal Control*, John Wiley & Sons, New York, NY, USA, 1975.

[22] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, John Wiley & Sons, New York, NY, USA, 1962.

[23] T. Bullock and G. Franklin, “A second-order feedback method for optimal control computations,” *IEEE Transactions on Automatic Control*, vol. 12, no. 6, pp. 666–673, 1967.

[24] S. da Silva Fernandes, “A note on the solution of the coast-arc problem,” *Acta Astronautica*, vol. 45, no. 1, pp. 53–57, 1999.

[25] S. da Silva Fernandes, “Generalized canonical systems applications to optimal trajectory analysis,” *Journal of Guidance, Control, and Dynamics*, vol. 22, no. 6, pp. 918–921, 1999.

[26] F. W. Gobetz, “A linear theory of optimum low-thrust rendez-vous trajectories,” *The Journal of the Astronautical Sciences*, vol. 12, no. 3, pp. 69–76, 1965.

Sandro da Silva Fernandes: Departamento de Matemática, Instituto Tecnológico de Aeronáutica,
12228-900 São José dos Campos, SP, Brazil
Email address: sandro@ita.br

Wander Almodovar Golfetto: Instituto de Aeronáutica e Espaço, Comando Geral de Tecnologia
Aeroespacial, 12228-904 São José dos Campos, SP, Brazil
Email address: wander@iae.cta.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 80321, 20 pages
doi:10.1155/2007/80321

Research Article

Fault Detection and Control of Process Systems

Vu Trieu Minh, Nitin Afzulpurkar, and W. M. Wan Muhamad

Received 8 September 2006; Revised 25 December 2006; Accepted 5 February 2007

Recommended by José Manoel Balthazar

This paper develops a stochastic hybrid model-based control system that can determine online the optimal control actions, detect faults quickly in the control process, and reconfigure the controller accordingly using interacting multiple-model (IMM) estimator and generalized predictive control (GPC) algorithm. A fault detection and control system consists of two main parts: the first is the fault detector and the second is the controller reconfiguration. This work deals with three main challenging issues: design of fault model set, estimation of stochastic hybrid multiple models, and stochastic model predictive control of hybrid multiple models. For the first issue, we propose a simple scheme for designing faults for discrete and continuous random variables. For the second issue, we consider and select a fast and reliable fault detection system applied to the stochastic hybrid system. Finally, we develop a stochastic GPC algorithm for hybrid multiple-models controller reconfiguration with soft switching signals based on weighted probabilities. Simulations for the proposed system are illustrated and analyzed.

1. Introduction

Control systems are becoming more and more powerful and sophisticated. Reliability, availability, and safety are primary goals in the operation of process systems. The aim to develop a fast and reliable control system that could detect undesirable changes in the process (referred to as "faults") and isolate the impact of faults has been attracting much attention of researchers. Various methods for fault detection and control of process systems have been studied and developed over recent years [1–8] but there are relatively few successful developments of controller systems that can deal with faults in stochastic hybrid sense where faults are modeled as multiple-model set with varying variable structure and use of stochastic model predictive control algorithm.

Faults are difficult to foresee and prevent. Traditionally, faults were handled by describing the result behavior of the system and were grouped into a hierarchical structure of fault model [9]. This approach is still widely used in practice: when a failure occurs, the system behavior changes and should be described by a different mode from the one that corresponds to the normal mode. A more appropriate mathematical model for such a system is the so-called stochastic hybrid approach. It differs from the conventional hierarchical structure in that its state may jump as well as it may vary continuously. Apart from the applications to problems involving failures, hybrid systems have found great success in such areas as target tracking and control that involve possible structure changes [10]. Hybrid systems switch among many operating modes, where each mode is governed by its own characteristic dynamic laws. Mode transitions are triggered by variables crossing specific thresholds.

For the fault modeling and verification, design of a model set is the key issue for the application of multiple-model estimator and controller. For simple systems, fault model set can be designed as a fixed structure (FS) or a fixed set of models at all times. The FS has fundamental limitations that only one fixed model set can be represented sufficiently and accurately by all possible failures. Actually, the set of possible system modes is not fixed. It varies and depends on the hybrid state of the system at the previous time. As shown in [11], use of more models in an FS estimation does not improve the performance. In fact, the performance will deteriorate if too many models are used in one fixed model set. Therefore, a variable structure (VS) is considered for modeling faults. The VS overcomes limitations of FS by using a variable set of models determined in real time adaptively. General and representative problems of model-set design, choice, and comparison for multiple-model approach to hybrid estimation are given in [12]. In this paper, a simple scheme for modeling of fault set as discrete and continuous random variables is proposed.

For the fault detection, various methods have been developed in [13–17]. One of the most effective approaches for solving stochastic hybrid systems is based on the use of multiple models (MM): it runs a bank of filters in parallel, each based on a particular model to obtain the model-conditional estimates. MM estimation algorithms appeared in early 1970s when Bar-Shalom and Tse [18] introduced a suboptimal, computationally bounded extension of Kalman filter to cases where measurements were not always available. Then, several multiple-model filtering techniques, which could provide accurate state estimation, have been developed. Major existing approaches for MM estimation are discussed and introduced in [18–23] including the noninteracting multiple model (NIMM), the Gaussian pseudo-Bayesian (GPB1), the second-order Gaussian pseudo-Bayesian (GPB2), and the interacting multiple models (IMM). Among those, we consider and select a fast and reliable algorithm for the fault detection system applied to the above model-set design.

Finally, for the controller reconfiguration (CR), we propose the use of stochastic model predictive control (MPC) algorithm applied to stochastic hybrid multiple models. The problem of determining optimal control laws for hybrid systems has been widely investigated and many results can be found in [24–28]. However, the use of MPC applied to stochastic hybrid systems is unfavorable since the general MPC algorithms follow deterministic perspective approaches. Thus, we propose use of generalized predictive control (GPC), a stochastic MPC technique developed by Clarke el al. [29, 30]. GPC uses the ideas with controlled autoregressive integrated moving average (CARIMA) plant in adaptive context and self-tuning by recursive estimation. Kinnaert [31] developed GPC from CARIMA model into a more general one in MIMO state-space form. We propose the use of a bank of GPC controllers, each based on a particular model. The optimal control action is the summation of probabilistic weighted outputs of each GPC controller. A similar soft switching mechanism based on weighted probabilities has been developed. Simulations for the proposed controller are illustrated and analyzed. Results show its strong ability for real applications to detect faults in dynamic systems.

The outline of this paper is as follows: Section 2 introduces the stochastic hybrid system and fault modeling design; Section 3 considers the selection of a fault detection system; Section 4 develops a controller reconfiguration integrated with fault detection system; examples and simulations are given after each section to illustrate the main ideas of the section; finally conclusions are given in Section 5.

2. Hybrid system and fault modeling design

An important requirement currently exists for improving the safety and reliability of process systems in ways that reduce their vulnerability to failures. When a failure occurs, the system behavior changes and should be described by a different mode from the one that corresponds to the normal mode of operation. An effective way to model faults for dynamic failures, which state may jump as well as vary continuously in a discrete set of modes, is the so-called a hybrid system.

A simplest continuous time hybrid system is described by the following different linear state update equation:

$$\begin{aligned} \dot{x}(t) &= A(t,m(t))x(t)+B(t,m(t))u(t)+T(t,m(t))\xi(t),\\ z(t) &= C(t,m(t))x(t)+\eta(t,m(t)), \end{aligned} \tag{2.1}$$

and a discrete-time hybrid system is the following:

$$\begin{aligned} x(k+1) &= A(k,m(k+1))x(k)+B(k,m(k+1))u(k)+T(k,m(k+1))\xi(k),\\ z(k) &= C(k,m(k))x(k)+\eta(k,m(k)), \end{aligned} \tag{2.2}$$

where A, B, T, and C are the system matrices, $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ is the control input, $z \in \mathbb{R}^p$ is the measured output, $\xi \in \mathbb{R}^{n_\xi}$ and $\eta \in \mathbb{R}^{p_\eta}$ are independent noises with means $\overline{\xi}(t)$ and $\overline{\eta}(t)$, and covariances $\Theta(k)$ and $\Xi(k)$. In this equation, $m(t) \in M$ is the system mode, which may jump or stay unchanged, x is the state variable, which varies continuously. The system mode sequence is assumed to be a first-order Markov chain with transition probabilities:

$$\Pi\{m_j(k+1) \mid m_i(k)\} = \pi_{ij}(k), \quad \forall m_i, m_j \in M_k, \tag{2.3}$$

where $\Pi\{\cdot\}$ denotes probability, $m(k)$ is the discrete-valued modal state, that is, the index of the normal or fault modes at time k, $M_k = \{m_1,\dots,m_N\}$ is the set of all possible system modes at time instant k, $\pi_{ij}(k)$ is the transition probability from mode m_i to mode m_j at time instant k. Obviously, the following relation must be held for any $m_i \in M_k$:

$$\sum_{j=1}^{M_k} \pi_{ij}(k) = \sum_{j=1}^{M_k} \Pi\{m_j(k+1) \mid m_i(k)\} = 1, \quad m_{i,\dots,N} \in M_k \subset M. \tag{2.4}$$

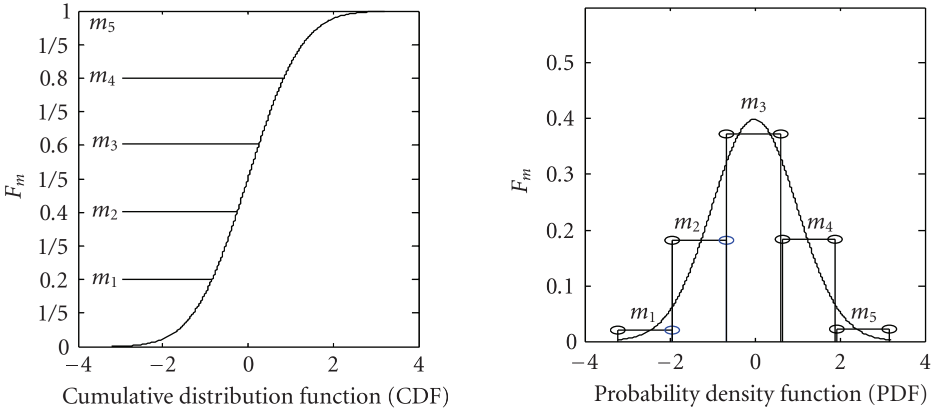
Faults can be modeled by changing the appropriate matrices A, B, C, or T representing the effectiveness of failures in the systems. They can also be modeled by increasing the process noise covariance Θ or measurement noise covariance Ξ in ξ and η. M_k denotes the set of models used at time instant k and M denotes the total set of models used, that is, M is the union of all $M_k's$:

$$m_i \in M_k = \begin{cases} x(k+1) = A_i(k)x(k) + B_i(k)u(k) + T_i(k)\xi_i(k), \\ z(k) = C_i(k)x(k) + \eta_i(k), \end{cases} \tag{2.5}$$

where the subscript i denotes the fault modeling in model set, $m_i \in M_k = \{m_1, m_2, \dots, m_N\}$, each m_i corresponds to a node (a fault) occurring in the process at time instant k. In fixed structure, the model set M_k used is fixed over time, that is, $M_k \overline{\Delta} = M$, for all k, to be determined offline based on the initial information about the system faults. Otherwise, we have a variable structure or the model set M_k varies at any time in the total model set M or $M_k \subset M$. Variable structure model overcomes fundamental limitations of fixed structure mode set because the fixed model set used does not always exactly match the true system mode set at any time, or the set of possible modes at any time varies and depends on the previous state of the system.

For faults varying as continuous variables, we can handle them via probabilistic modeling techniques. In these cases, faults can be modeled as discrete modes based on their cumulative distribution function (CDF) and probability density function (PDF). Data of the past operation fault records (fault rate and percentage of fault type) provide the required probability distribution of the mode. More methods on design of model set as continuous random variables can be read in [12]. In this paper, we just propose the simplest method of equal probability to model a continuous random variable into discrete modes. We assume that the CDF $F_s(x)$ of the true continuous variable s is known and we want to reconstruct it into the CDF $F_m(x)$ of discrete modes. In the equal probability method, we propose to group the CDF $F_s(x)$ into $|M|$ discrete modes of equal probabilities, $m_i = 1/|M|$ (preferably an odd number $3, 5, 7, \dots$, for symmetric distributions). The design of a continuous variable into discrete modes is shown in Figure 2.1 with $|M| = 5$ and PDF is a normal distribution $f(x;\mu,\sigma) = (1/\sigma\sqrt{2\pi})\exp(-(x-\mu)^2/2\sigma^2)$ with mean $\mu = 0$ and variance $\sigma^2 = 1$.

In Figure 2.1, we group a continuous random variable with a normal distribution into five equal probabilities (discrete modes) with a model set (mode set) of $M = \{m_1, m_2, \dots, m_{|M|}\}$.



(a)

(b)

Figure 2.1. Group a normal CDF into equal probabilities.

Example 2.1 (fault model-set design). Consider a continuous process system with the state space model in (2.1):

$$\begin{aligned} \dot{x}(t) &= A_i x(t) + B_i u(t) + T_i \xi_i(t), \\ z(t) &= C_i x(t) + \eta_i(t), \end{aligned} \tag{2.6}$$

where A_i, B_i, T_i, and C_i are system matrices, ξ_i and η_i are independent noises with zero-mean $\overline{\xi} = \overline{\eta} = 0$ and constant covariance $\Theta = \Xi = 0.02^2 I$, $T_i = I$. We assume that at the normal operation mode N, two generic types of faults might take place: one static fault mode S_0 and one varying fault mode V_0. This example is modeled from a chemical process model with four state variables, two inputs and two outputs. For simplicity, we verify only one input.

We have the normal operation mode:

$$A_N = \begin{bmatrix} 1 & 0 & 0.1 & 0 \\ 0 & 1 & 0 & 0.1 \\ -0.08 & 0.06 & 0.7 & 0 \\ 0.1 & -0.1 & 0 & 0 \end{bmatrix}, \qquad B_N = \begin{bmatrix} -0.2 \\ 0.03 \\ 2 \\ 1 \end{bmatrix}, \qquad C_N = \begin{bmatrix} 1 & -0.5 & 1 & 1 \\ -1 & 0.6 & 0 & 1 \end{bmatrix}. \tag{2.7}$$

A static failure mode S_0 happens when an interrupted actuator failure, −50%,

$$B_{S_0} = 0.5 B_N = \begin{bmatrix} -0.1 \\ 0.015 \\ 1 \\ 0.5 \end{bmatrix}. \tag{2.8}$$

A varying failure mode V_0 happens when a continuous varying variable appears in A_N,

$$A_{V_0} = \begin{bmatrix} 1 & 0 & 0.1 & 0 \\ 0 & 1 & 0 & 0.1 \\ -0.08 & 0.06 & 0.7 & \sin(\omega) \\ 0.1 & -0.1 & 0 & 0 \end{bmatrix}, \tag{2.9}$$

where ω is a continuous varying variable (deg/s).

We assume that at the static mode S_0, two other generic types of static faults might take place: mode S_1 with sensor 1 failure -50% or

$$C_{S_1} = \begin{bmatrix} 0.5z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0.5 & -0.25 & 0.5 & 0.5 \\ -1 & 0.6 & 0 & 1 \end{bmatrix}, \tag{2.10}$$

and mode S_2 with sensor 1 failure +50% or

$$C_{S_2} = \begin{bmatrix} 1.5z_1 \\ z_2 \end{bmatrix} \begin{bmatrix} 1.5 & -0.75 & 1.5 & 1.5 \\ -1 & 0.6 & 0 & 1 \end{bmatrix}. \tag{2.11}$$

We continue to assume that the PDF of the continuous varying variable ω in matrix A_{V_0} is the mixture of three normal distributions:

$$f(\omega) = \frac{1}{3\sqrt{2\pi}} \exp\left(-\frac{(\omega-3)^2}{2}\right) + \frac{1}{3\sqrt{2\pi}} \exp\left(-\frac{(\omega)^2}{2}\right) + \frac{1}{3\sqrt{2\pi}} \exp\left(-\frac{(\omega+3)^2}{2}\right). \tag{2.12}$$

Since the PDF of ω is the combination of three normal curves with three mean values $\overline{\omega}_1 = -3^0/\text{s}$, $\overline{\omega}_0 = 0^0/\text{s}$, and $\overline{\omega}_2 = 3^0/\text{s}$, we can group this continuous varying variable into three discrete models (modes) with A_{V_1}, A_{V_0}, and A_{V_2} corresponding to the above three mean values with equal probabilities of 1/3, 1/3, and 1/3. The model set design via CDF and its reconstruction PDF are shown in Figure 2.2.

Hence, in this example, we have total 7 models (modes) grouped into three varying model sets in Figure 2.3:

$$\begin{aligned} &\text{model set 1: } M_1 = \{m_1 = N_{(A_N,B_N,C_N)}, m_2 = S_{0(A_S,B_N,C_N)}, m_3 = V_{0(A_{V_0},B_N,C_N)}\}, \\ &\text{model set 2: } M_2 = \{m_2 = S_{0(A_S,B_N,C_N)}, m_4 = S_{1(A_S,B_N,C_{S_1})}, m_5 = S_{2(A_S,B_N,C_{S_2})}\}, \\ &\text{model set 3: } M_3 = \{m_3 = V_{0(A_{V_0},B_N,C_N)}, m_6 = V_{1(A_{V_1},B_N,C_N)}, m_7 = V_{2(A_{V_2},B_N,C_N)}\}. \end{aligned} \tag{2.13}$$

We assume that the following Markov transition probability matrix in (2.3) is used for all simulations in the total model set $M = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7\}$:

$$\Pi = \begin{bmatrix} 0.94 & 0.03 & 0.03 & 0 & 0 & 0 & 0 \\ 0.03 & 0.93 & 0 & 0.02 & 0.02 & 0 & 0 \\ 0.03 & 0 & 0.93 & 0 & 0 & 0.02 & 0.02 \\ 0 & 0.05 & 0 & 0.95 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0.05 & 0 & 0 & 0 & 0.95 \end{bmatrix}. \tag{2.14}$$

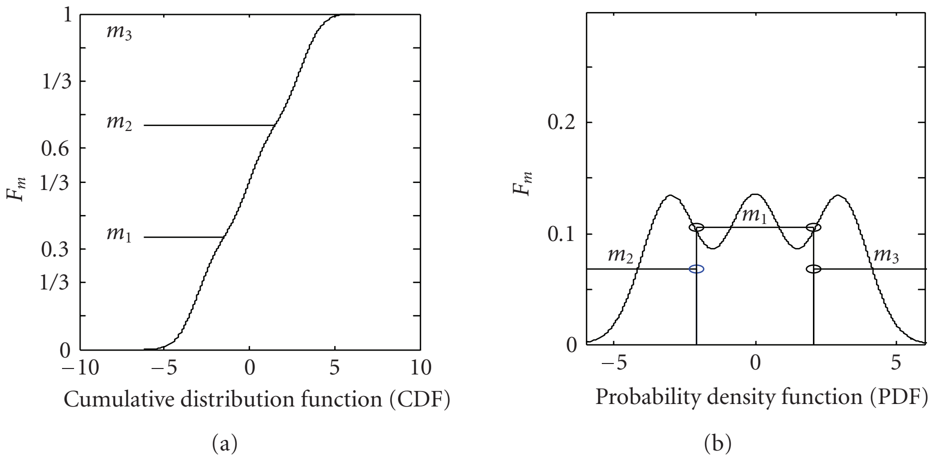


Figure 2.2. Model set design of varying variable ω.

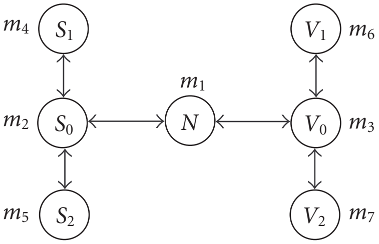


Figure 2.3. Total model set design.

The design of model set now is completed. In the next section, we will consider the selection of a reliable fault detection system applied to this model set.

3. Fault detection system

Fault detection for stochastic hybrid systems has received a great attention in recent years. A variety of different fault detection methods has been developed. For hybrid systems with fixed structure (FS) or variable structure (VS) modeled in mixed logic dynamical (MLD) form or piecewise affine (PWA) systems, the state estimation can be solved by moving horizon estimation (MHE) strategy. MHE has strong ability to incorporate constraints on states and disturbances. Moreover, on the computational side, because MHE algorithms lead to optimization problem of fixed dimension, they are suitable for practical implementation. MHE is applied successfully to constrained linear systems where it can guarantee stability of the estimate when other classical techniques, like Kalman filtering, fail [32]. A number of MHE techniques for fault detection system can be found in [32–35]. However for stochastic hybrid systems where their state can jump as well

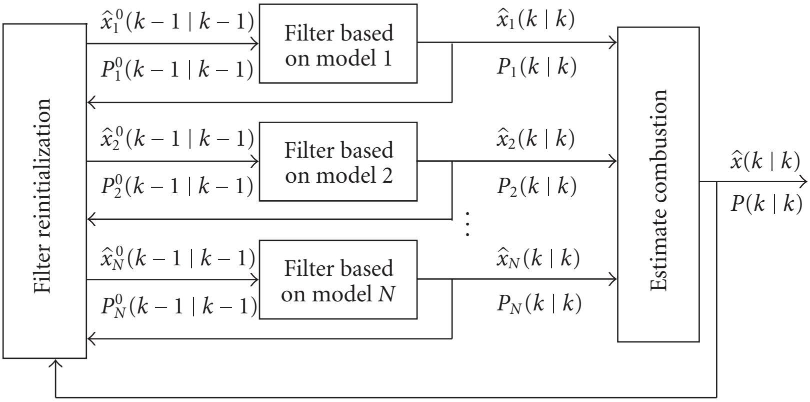


Figure 3.1. Structure of an MM estimator.

as vary continuously and randomly in a model set with the system mode sequence assumed to be a first-order Markov chain in (2.3), a more effective and natural estimation approach is the use of algorithms of multiple-model (MM) estimator. Major existing approaches for MM estimation are discussed and introduced in [18–26]. In this part, we consider and select a reliable fault detection system among the noninteracting multiple models (NIMM), the Gaussian pseudo-Bayesian (GPB1), the second-order Gaussian pseudo-Bayesian (GPB2), and the interaction multiple models (IMM).

From the design of model set (in Section 2), a bank of filters runs in parallel at every time, each based on a particular model, to obtain the model-conditional estimates. The overall state estimate is a probabilistically weighted sum of these model-conditional estimates. The jumps in system modes can be modeled as switching among the assumed models in the set.

Figure 3.1 shows the operation of a recursive multiple-model estimator, where $\hat{x}_i(k \mid k)$ is the estimate of the state $x(k)$ obtained from the filter based on model m_i at time k given the measurement sequence through time k; $\hat{x}_i^0(k-1 \mid k-1)$ is the equivalent reinitialized estimate at time $(k-1)$ as the input to the filter based on model m_i at time k; $\hat{x}(k \mid k)$ is the overall state estimate; $P_i(k \mid k)$, $P_i^0(k-1 \mid k-1)$, and $P(k \mid k)$ are the corresponding covariances.

A simple and straightforward way of filter reinitialization is that each single model-based recursive filter uses its own previous state estimation and state covariance as the input at the current cycle:

$$\begin{aligned} \hat{x}_i^0(k-1 \mid k-1) &= \hat{x}_i(k-1 \mid k-1), \\ P_i^0(k-1 \mid k-1) &= P_i(k-1 \mid k-1). \end{aligned} \tag{3.1}$$

This leads to the so-called noninteracting multiple-model (NIMM) estimator because the filters operate in parallel without interactions with one another, which is reasonable only under the assumption that the system mode does not change.

Another way of reinitialization is to use the previous overstate estimate and covariance for each filter as the required input:

$$\begin{aligned}\hat{x}_i^0(k-1 \mid k-1) &= \hat{x}(k-1 \mid k-1),\\ P_i^0(k-1 \mid k-1) &= P(k-1 \mid k-1).\end{aligned} \tag{3.2}$$

This leads to the first-order generalized pseudo-Bayensian (GPB1) estimator. It belongs to the class of interacting multiple-model estimators since it uses the previous overall state estimate, which carries information from all filters. Clearly, if the transition probability matrix is an identity matrix, this method of reinitialization reduces to the first one.

The GPB1 and GPB2 algorithms were the result of early work by Ackerson and Fu [21] and good overviews are provided in [22], where suboptimal hypothesis pruning techniques are compared. The GPB2 differed from the GPB1 by including knowledge of the previous time step's possible mode transitions, as modeled by a Markov chain. Thus, GPB2 produced slightly smaller tracking errors than GPB1 during nonmaneuvering motion. However in the size of this part, we do not include GPB2 into our simulation test and comparison.

A significantly better way of reinitialization is to use IMM. The IMM was introduced by Zhang and Li in [23]:

$$\begin{aligned}\hat{x}_j^0(k \mid k) &= E\left[x(k) \mid z^k, m_j(k+1)\right] = \sum_{i=1}^{N} \hat{x}_i(k:k)P\{m_i(k) \mid z^k, m_j(k+1)\},\\ P_j^0(k \mid k) &= \operatorname{cov}\left[\hat{x}_j^0(k:k)\right] = \sum_{i=1}^{N} P\{m_i(k) \mid z^k, m_j(k+1)\}\\ &\qquad \times \left\{P_i(k \mid k) + \tilde{x}_{ij}^0(k \mid k)\tilde{x}_{ij}^0(k \mid k)'\right\},\end{aligned} \tag{3.3}$$

where $\operatorname{cov}[\cdot]$ stands for covariance and $\tilde{x}_{ij}^0(k \mid k) = \hat{x}_i^0(k \mid k) - \hat{x}_j(k \mid k)$. In this paper, we will use this approach for setting up a fault detection system.

For each model in $M_k \in M = \{m_1, \ldots, m_N\}$, we can operate a Kalman filter. The probability of each model matching to the system mode provides the required information for mode's chosen decision. The mode decision can be achieved by comparing it with a fixed threshold probability μ_T. If the mode probabilities $\max_i(\mu_i(k)) \geq \mu_T$, mode at $\mu_i(k)$ has occurred and has taken place at the next cycle. Otherwise, there is no new mode detection. The system maintains the current mode for the next cycle calculation.

Example 3.1 (test and selection of fault detection system). From the model-set design in Example 2.1, model-set modes in (2.6) are discretized with 0.1 second, the threshold value for mode probabilities is chosen as $\mu_T = 0.9$. Now we begin to compare the three estimators of NIMM, GPB1, and IMM to test their ability to detect faults. The seven models are run for a time interval $t = 20$ seconds and for the following sequence: $\{m_1, m_2, m_4, m_2, m_5, m_2, m_1, m_3, m_6, m_3, m_7\}$. Results of simulation are shown in Figure 3.2.

In Figure 3.2, we can see that the GPB1 estimator performs as good as IMM estimator while NIMM estimator fails to detect faults in the model set. Next we continue to test the

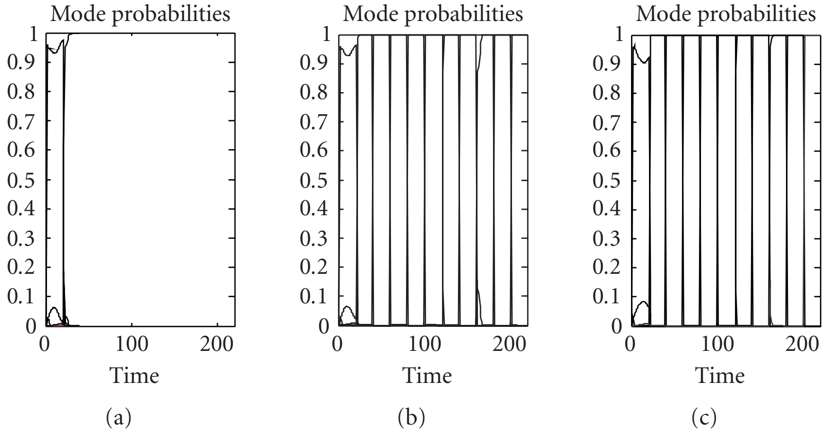


Figure 3.2. Probabilities of estimators (a) NIMM, (b) GPB1, and (c) IMM.

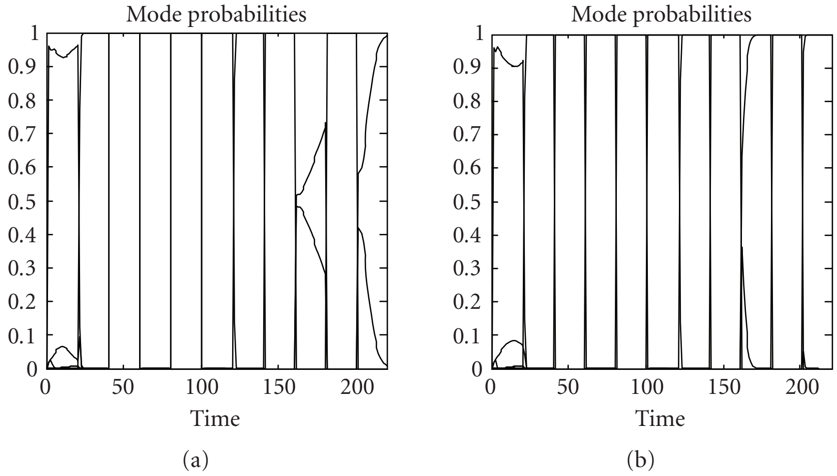


Figure 3.3. Probabilities of estimators (a) GPB1 and (b) IMM.

ability of GPB1 and IMM estimators by narrowing the distances between modes as close as possible until one of methods cannot detect the failures. Now we assume to design new two varying modes of $\{m_6^*, m_7^*\}$ corresponding to a new $A_{V_1}^*$ with $\overline{\omega}_1^* = 0.3^0/\text{s}$ and a new $A_{V_2}^*$ with $\overline{\omega}_2^* = -0.3^0/\text{s}$. With these new parameters, GPB1 fails to detect failures since the distance between modes $\{m_6^*, m_3, m_7^*\}$ is very close, while IMM still proves it is much superior in Figure 3.3.

As a result, we select the IMM for our fault detection system. Now we move to the main part of this paper to set up a controller reconfiguration for the fault detection and control system.

4. Controller reconfiguration

In this section, we develop a new CR which can determine online the optimal control actions and reconfigure the controller accordingly. The problem of determining the optimal control laws for hybrid systems has been widely studied in recent years and many methods have been developed in [24–28]. Optimal quadratic control of piecewise linear and hybrid systems is found in [25, 26]. For complex constrained multivariable control problems, model predictive control (MPC) has become the accepted standard in the process industries [36, 37]. MPC can be applied to multiple models using linear matrix inequalities (LMIs) in [38]. The general MPC algorithms follow deterministic perspective approaches, hence, for stochastic hybrid systems described in (2.1), (2.2), (2.3), and (2.4), there are few MPC ideas applied to control stochastic hybrid systems. Thus, we propose a new controller reconfiguration (CR) using generalized predictive control (GPC) algorithm. We will show how an IMM-based GMC controller can be used as a good fault detection and control system.

Generalized predictive control (GPC) is one of model predictive control (MPC) techniques developed by Clarke et al. [29, 30]. GPC was intended to offer a new adaptive control alternative. GPC uses the ideas with controlled autoregressive integrated moving average (CARIMA) plant in adaptive context and self-tuning by recursive estimation. Kinnaert [31] developed GPC from CARIMA model into a more general form when the models are described in space.

The optimal control problem for the general cost function for GPC controller in (2.1) is

$$\begin{gathered}
\min_{U \triangleq \{u_1, u_2, \ldots, u_{t+N_u-1}\}} \left\{ J(U, x(t)) = x'_{t+N_{y|t}} P x_{t+N_{y|t}} + \sum_{k=0}^{N_y-1} \left[x'_{t+k|t} Q x_{t+k|t} + u'_{t+k|t} R u_{t+k|t} \right] \right\}, \\
\text{subject to } x_{t+k+1|t} = A x_{t+k|t} + B u_{t+k} + T \xi_{t+k|t}, \\
u_{t+k} = -K x_{t+k|t}, \quad k \geq N_u, \\
x_{t+k} \in \mathbb{X}, \quad u_{t+k} \in \mathbb{U},
\end{gathered} \tag{4.1}$$

where $Q = Q' > 0$, $R = R' \geq 0$ are the weighting matrices for predicted state and input, respectively. Linear feedback gain K and the Lyapunov matrix $P > 0$ are the solution of Riccati equation. For simplicity, we assume that the predictive horizon is set equal to the control horizon, that is, $N_u = N_y = N_P$.

By substituting $x_{t+N_p|t} = A^{N_P} x(t) + \sum_{j=0}^{N_P-1} A^j B u_{t+N_p-1-j} + A^{N_P-1} T \xi(t)$, (4.1) can be rewritten as

$$\min_U \left\{ \frac{1}{2} U' H U + x'(t) F U + \xi'(t) Y U \right\}, \quad \text{subject to } GU \leq W + Ex(t), \tag{4.2}$$

where the column vector $U \triangleq [u'_t, \ldots, u'_{t+N_p-1}]' \in \mathbb{R}^U$ is the predictive optimization vector, $H = H' > 0$, and H, F, Y, G, W, E are obtained from (4.1) as only the optimizer

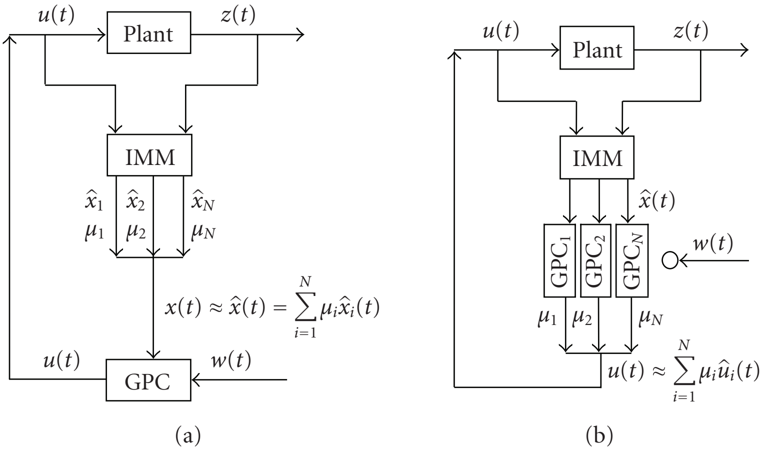


Figure 4.1. Two diagrams of IMM-based GPC controllers.

U is needed. Then, the optimization problem (4.2) is a quadratic program and depends on the current state $x(t)$ and noise $\xi(t)$. The implementation of GPC requires the online solution of a quadratic program at each time step.

For the controller reconfiguration (CR), we can use the output of IMM, the overall state estimate $x(k) \approx \hat{x}(k) = \sum_{i=1}^{N} \mu_i \hat{x}_i(k)$ in (3.3), where N is the number of models in the current model set, as the input for (a) a GPC controller or for (b) a bank of GPC controllers shown in Figure 4.1.

In Figure 4.1(a), IMM provides the overall state estimate $\hat{x}(t)$ and indicates one "most reliable" mode m_i in the mode set. Thus, we can build up a GPC controller corresponding to this "most reliable" mode. The stability of the system is assured if we can find a positive Lyapunov matrix in (4.1).

In Figure 4.1(b), assuming the model probabilities $\mu_i(t)$ are constant during the predictive control horizon, we can derive a new GPC control law that the "true" model $\overline{m}$ is the union of all model modes m_i in the current model set M_k in (2.5) or $\overline{m} = \sum_{i=1}^{N} \mu_i m_i$. Thus, we can achieve a new control law by using a bank of GPC controllers for each model in the model set and have the overall control input $u(t) \approx \sum_{i=1}^{N} \mu_i \hat{u}_i(t)$. The stability of the system is guaranteed if we can find out a common Lyapunov matrix P for all models in the model set.

Lemma 4.1. *The optimal control problem for the general cost function for GPC controller in (4.1) applied to control stochastic hybrid system in (2.1) can guarantee the global and asymptotical stability if there exist positive definite matrices P and θ_i such that $A_iP + PA_i' = -\theta_i$, for all i.*

Proof. For simplicity, we assume that the control input $u(t+N_P) = 0$ after $k \geq N_P$ predictive control horizon so that a common Lyapunov matrix for each model in (4.1) is the solution of Riccati equations $A_iP + PA_i' = -\theta_i$ since the state update equation then becomes $\dot{x}(t) = \sum_{i=1}^{N} \mu_i A_i x(t)$. For a positive Lyapunov function $V(x) = x'(t)Px(t)$, we have always a negative definite time derivative $\dot{V}(x) < 0$, and the system is stable:

$$\dot{V}(x) = \left(\sum_{i=1}^{N} \mu_i A_i x\right)' Px + x'P\left(\sum_{i=1}^{N} \mu_i A_i x\right) = \sum_{i=1}^{N} \mu_i x'\left(A_i P + PA_i'\right)x = \sum_{i=1}^{N} \mu_i x'\left(-\theta_i\right)x < 0. \tag{4.3}$$

Otherwise, the closed-loop feedback in (4.1) $u_{t+k} = -Kx_{t+k|t}$ for $k \geq N_P$, and we have $\dot{x}(t) = Ax(t) + Bu(t)$ or $\dot{x}(t) = (\sum_{i=1}^{N} \mu_i(t)(A - BK_i))x(t) = (\sum_{i=1}^{N} \mu_i(t)A_{Li})x(t)$ can also satisfy Lemma 4.1 in (4.3). A similar result was found in [38] when we can apply a common Lyapunov matrix to find a robust stabilizing state feedback for uncertain hybrid systems using LMIs.

For the controller reconfiguration (CR), we can apply hard switching or soft switching. For hard switching, we use only one controller implemented at any time—similar scheme in Figure 4.1(a). As indicated in [6], even if each controller globally stabilizes, there can exist a switching sequence that destabilizes the closed-loop dynamics. Now we consider some possible soft switching signals where the outputs of each controller are weighted by a continuous, time-varying, probability vector $\nu_i(t)$ which can guarantee the closed-loop stability, $u(t) = \sum_{i=1}^{N} \nu_i \hat{u}_i(t)$, in which $\sum_{i=1}^{N} \nu_i(t) = 1$, $\nu_i(t) \in [0,1]$ for all i, t.

It is difficult to find out a common Lyapunov matrix for all models in the model set (4.3). Recently, a new type of parameter-dependent Lyapunov function has been introduced in the form that $P_L = \sum_{i=1}^{N} \nu_i P_i$ is a parameter-dependent Lyapunov function for any $A_L = \sum_{i=1}^{N} \nu_i A_{Li}$. That is true since we always have a negative derivative $\dot{V}(x) < 0$ in (4.3) as $\dot{V}(x) = \sum_{i=1}^{N} \nu_i x'(A_i P_i + P_i A_i')x = \sum_{i=1}^{N} \nu_i x'(-\theta)x < 0$. However, parameter-dependent Lyapunov matrices do not insure the stability in switching sequence as indicated in [6].

The existence of a direct common Lyapunov matrix $A_i P + PA_i' = -\theta_i$ can be searched using software for solving LMIs. However we propose another method which can find a common Lyapunov matrix with LMIs from their discrete equations. □

Lemma 4.2. *The optimal control problem for the general cost function for GPC controller in (4.1) applied to control stochastic hybrid system in (2.2) can guarantee the global and asymptotical stability if there exist positive definite matrices P and scalar γ such that*

$$\begin{bmatrix} P & PA_i' & \gamma \\ A_i P & P & 0 \\ \gamma & 0 & \gamma I \end{bmatrix} > 0, \quad \forall i. \tag{4.4}$$

Proof. Suppose there exists a Lyapunov function in (4.1) and the system will be stable if the Lyapunov function is decreasing, that is, $J(x(t+N_P+1)) < J(x(t+N_P))$, or $x(t+N_P+1)'Px(t+N_P+1) - x(t+N_P)'Px(t+N_P) < 0$, or $P - A_i'PA_i > 0$, for all i. By adding a scalar $\gamma > 0$, we have $P - A_i'PA_i - \gamma I > 0$, or $P - (A_i'P)P^{-1}(PA_i) - (\gamma)I\gamma^{-1}(\gamma) > 0$. And using Schur complement, this equation is equivalent to the LMI in Lemma 4.2.

Hence, the indirect common Lyapunov matrix in Lemma 4.2 is the solution to the following LMI:

$$\min_{P>0,\,\gamma>0} \gamma, \quad \text{subject to} \begin{bmatrix} P & PA_i' & \gamma \\ A_iP & P & 0 \\ \gamma & 0 & \gamma I \end{bmatrix} > 0, \ \forall i. \tag{4.5}$$

The above is CR design proposal for nonoutput tracking GPC controllers. However in reality, the primary control objective is to force the plant outputs to track their set points. What is about the CR design for tracking GPC controllers? In tracking GPC, the state space of the stochastic model in (2.2) now can be changed into a new innovation form [31]:

$$\begin{aligned} \hat{x}(t+1 \mid t) &= \widetilde{A}\hat{x}(t \mid t-1) + \widetilde{B}\Delta u(t) + \widetilde{T}\xi(t), \\ z(t) &= \widetilde{C}\hat{x}(t \mid t-1) + \xi(t), \end{aligned} \tag{4.6}$$

where $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$, and $\widetilde{T}$ are fixed matrices from A, B, C, and T in (2.2), $\eta = \xi$, $z(t) \in \mathbb{R}^p$, $\Delta u(t) = u(t) - u(t-1) \in \mathbb{R}^m$, and $\hat{x}(t \mid t-1)$ is an estimate of state $x(t) \in \mathbb{R}^n$ obtained from a Kalman filter. For a moving horizon control, the prediction of $x(t+j \mid t)$ in (4.6) given the information $\{z(t), z(t-1), \dots, u(t-1), u(t-2), \dots\}$ is

$$\hat{x}(t+j \mid t) = A^j\hat{x}(t \mid t-1) + \sum_{i=0}^{j-1} A^{j-1-i}B\Delta u(t+i) + A^{j-1}T\xi(t), \tag{4.7}$$

and the prediction of the filtered output is

$$\hat{z}(t+j \mid t) = CA^j\hat{x}(t \mid t-1) + \sum_{i=0}^{j-1} CA^{j-1-i}B\Delta u(t+i) + CA^{j-1}T\xi(t). \tag{4.8}$$

If we form $\widetilde{u}(t) = [\Delta u'(t), \dots, \Delta u'(t+N_P-1)]$ and $\widetilde{z}(t) = [\hat{z}'(t \mid t, \dots, \hat{z}'(t+N_P-1 \mid t)]$, we can write the global prediction model for the filtered-out from 1 to N_P prediction horizon as

$$\hat{z}(t) = \begin{bmatrix} CB & \cdots & 0 \\ CAB & \cdots & 0 \\ \vdots & \vdots & \vdots \\ CA^{N_P-1}B & CA^{N_P-2}B\vdots & CB \end{bmatrix} \widetilde{u}(t) + \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^{N_P} \end{bmatrix} \hat{x}(t \mid t-1) + \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N_P-1} \end{bmatrix} T\xi(t). \tag{4.9}$$

For simplicity, we can rewrite (4.9) as

$$\hat{z}(t) = U\widetilde{u}(t) + V\hat{x}(t \mid t-1) + WT\xi(t). \tag{4.10}$$

Consider the new tracking cost function of GPC [29]:

$$\min_{\widetilde{u}(t)=[\Delta u'(t),\ldots,\Delta u'(t+N_P-1)]}\left\{J(\widetilde{u}(t),x(t))=\sum_{j=1}^{N_P}\left[||z(t+j)-w(t+j)||+||\Delta u(t+j-1)||_{\Gamma}\right]\right\},$$
$$\text{subject to } x_{t+k}\in\mathbb{X},\quad z_{t+k}\in\mathbb{Z},\quad u_{t+k}\in\mathbb{U},\quad \Delta u_{t+k}\in\Delta\mathbb{U} \tag{4.11}$$

where N_P is the prediction horizon, $w(t+j)$ is the output reference, and Γ is the control weighting matrix, the control law that minimizes this tracking cost function is

$$\widetilde{u}(k)=-(U'U+\Gamma)^{-1}\left(V\hat{x}(t\mid t-1)+WT\xi(t)-w(t)\right) \tag{4.12}$$

then the first input $\Delta u(t)$ in $\widetilde{u}(t)$ will be implemented into the system. □

LEMMA 4.3. *Let (x_c,u_c) be an equilibrium pair and the corresponding equilibrium variable $z(t)=z_c$ at $w(t)=w_c$ assuming that the initial state $x(0)$ is such that a feasible solution of (4.11) exists at time $t=0$. Then the GPC law (4.12) stabilizes the system in* $\lim_{t\to\infty}x(t)=x_c$, $\lim_{t\to\infty}z(t)=w_c$, *and* $\lim_{t\to\infty}\Delta u(t)=0$ *while fulfilling constraints in (4.11).*

Proof. This stability problem follows easily from standard Lyapunov theory. Let $\widetilde{u}(0)$ denote the optimal control sequence $\widetilde{u}(0)=[\Delta u'(0),\ldots,\Delta u'(N_P-1)]$, let $V(t)\overset{\Delta}{=}J(\widetilde{u}(0),x(t))$ denote the corresponding value attained by the cost function, and let $\widetilde{u}(1)$ be the sequence $\widetilde{u}(1)=[\Delta u'(1),\ldots,\Delta u'(N_P-2)]$. Then, $\widetilde{u}(1)$ is feasible at time $t+1$, along with the vectors $\Delta u(k\mid t+1)=\Delta u(k+1\mid t)$, $z(k\mid t+1)=z(k+1\mid t)$, $k=0,\ldots,N_P-2$, $u(N_P-1\mid t+1)=u_c$, $z(N_P-1\mid t+1)=z_c$, because $x(N_P-1\mid t+1)=x(N_P\mid t)=x_c$. Hence,

$$V(t+1)\leq J(\widetilde{u}(1),x(t))=V(t)-||z(0)-w_c||-||\Delta u(0)||_{\Gamma} \tag{4.13}$$

and $V(t)$ is reducing. Since $V(t)$ is lower bounded by 0, there exists $V_\infty=\lim_{t\to\infty}V(t)$, which implies that $V(t+1)-V(t)\to 0$. Therefore, each term of the sum

$$||z(t)-w_c||+||\Delta u(t)||_{\Gamma}\leq V(t)-V(t+1) \tag{4.14}$$

converges to zero as well, and the system is stable.

The tracking cost function of GPC in (4.11) and (4.12) does not require to find out a Lyapunov matrix as in general cost function (4.1) and (4.2) so that the tracking GPC controller can guarantee the system stability for systems which do not have solution for the direct Lyapunov method, and can handle input and output constraints in the optimal control problem.

For tracking GPC controllers, we also propose two CR schemes for hard switcher and soft switcher as in Figure 4.1. For hard switcher, we run a tracking GPC controller corresponding to the "most reliable" mode detected by IMM as in Figure 4.1(a). However for a continuous varying variable system, a better control law is to mix all mode probabilities into a "true" model. We then build a bank of tracking GPC controllers for each model in the model set as in Figure 4.1(b). Assuming the mode probabilities are constant during the control horizon, we can easily derive a new GPC control law in (4.10) by forming

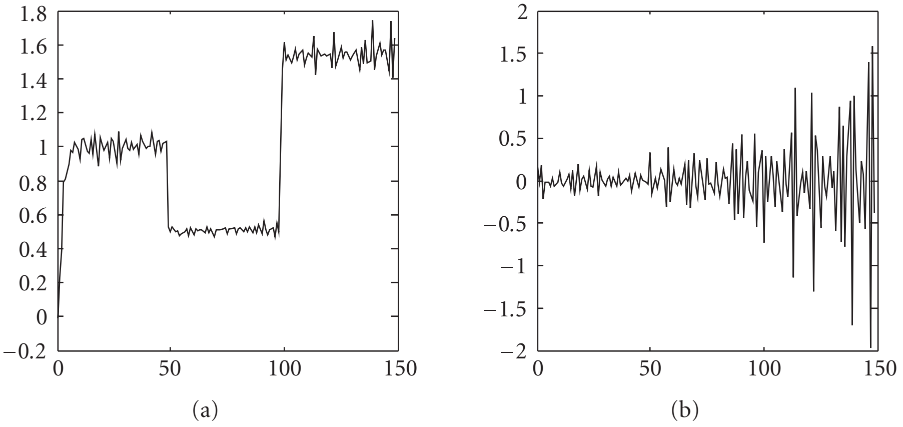


Figure 4.2. Normal GPC controller with sensor errors: (a) output and (b) input.

$U = (\sum_{i=1}^{N}\mu_i U_i)$, $V = (\sum_{i=1}^{N}\mu_i V_i)$, and $W = (\sum_{i=1}^{N}\mu_i W_i)$ matrices that correspond to the "true" model $\overline{m} = (\sum_{i=1}^{N}\mu_i m_i)$, and find out the optimal control action in (4.12). Then the first input $\Delta u(t)$ in $\widetilde{u}(t)$ will be implemented into the system. Next, we will run some simulations to test the above proposed fault detection and control system. □

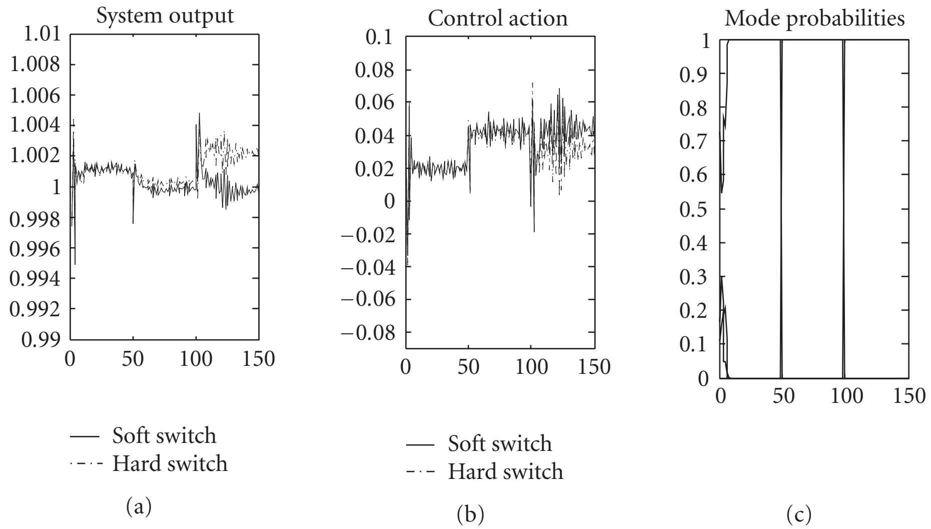
Example 4.4 (controller reconfiguration). The existence of a common Lyapunov matrix in (4.3) can be found by using LMI of Lemma 4.2. For simplicity, we assume that the control input $u(t+N_P) = 0$ after $k \geq N_P$ predictive control horizon so that the solution to the LMI

$$\min_{P>0,\gamma>0} \gamma, \quad \text{subject to} \begin{bmatrix} P & PA_i' & \gamma \\ A_i P & P & 0 \\ \gamma & 0 & \gamma I \end{bmatrix} > 0, \ \forall i, \tag{4.15}$$

can be applied directly to matrices $A_i = \{A_N, A_{V_0}, A_{V_1}, A_{V_2}\}$. We found that a common Lyapunov matrix for all A_i is

$$P = \begin{bmatrix} 6.43 & 1.69 & -1.62 & 0.24 \\ 1.69 & 4.34 & 0.15 & -0.30 \\ -1.62 & 0.15 & 4.14 & -0.10 \\ 0.24 & -0.30 & -0.10 & 3.25 \end{bmatrix}. \tag{4.16}$$

For tracking GPC controller, firstly we run a normal GPC controller with the predictive horizon $N_y = N_u = N_P = 4$, the weighting matrix $\Gamma = 0.1$, and with a reference set point $w = 1$. We assume that the current mode is mode S_0 from time $k = 1 - 50$, mode S_1 with sensor 1 failure -50% from time $k = 51 - 100$, and mode S_2 with sensor 1 failure $+50\%$ from time $k = 101 - 150$. Of course, the normal GPC controller provides wrong outputs (Figure 4.2).



(a)

(b)

(c)

Figure 4.3. IMM-based GPC controllers: (a) output, (b) input, and (c) probabilities (IMM).

Next we run GPC controller simulations using CR system with hard switcher and soft switcher (Figure 4.3). Our new FDMP system still keeps the output at the desired set point since the IMM estimator easily finds out accurate fault modes and activates the CR system online. The soft switcher provides a smoother and smaller offset error in tracking process due to the interaction of mode probabilities that are always mixed into the "true" mode.

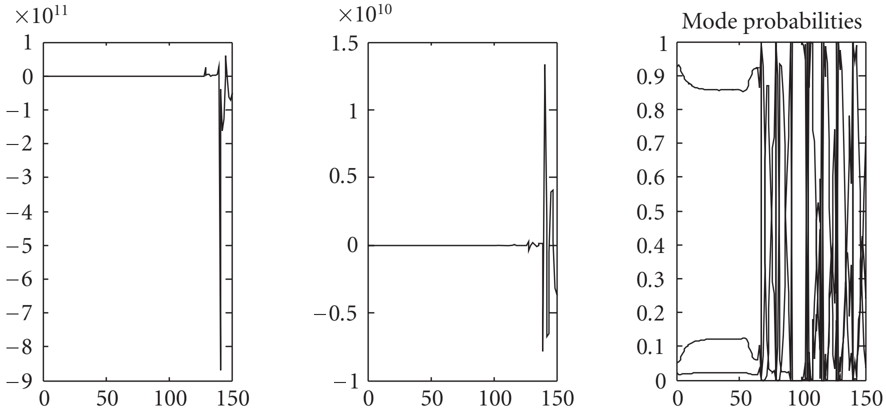
We then test the ability of the system to detect and control continuous varying variable in model set $M_3 = \{m_3, m_6, m_7\}$. Similar results are shown in Figures 3.3 and 4.3 that the IMM-based GPC controller can detect faults online and control well the varying variables with even small mode distances.

Finally, when we continue to narrow the distance between modes as we run the simulation with modes $\{m_6^*, m_3, m_7^*\}$ corresponding to $A_{V_1}^*$ with $\overline{\omega}_1^* = 0.1^0/\mathrm{s}$ and $A_{V_2}^*$ with $\overline{\omega}_2^* = -0.1^0/\mathrm{s}$, the IMM estimator fails to detect faults since the distance between modes becomes too close as shown in Figure 3.3(a), GPB1.

Low magnitude of input signals can also lead to failure of IMM-based GPC controller. If we reduce the reference set point to a very low value at $w = 0.01$, the system becomes uncontrollable (Figure 4.4): when the magnitude of the input signals is very small, the residuals of Kalman filters will be very small, and therefore, the likelihood functions for the modes will approximately be equal. This will lead to unchanging (or very slow changing) mode probabilities which in turn make the IMM estimator incapable to detect failures.

5. Conclusions

Systems subject to dynamic failures can be modeled as a set of variable structures using a variable set of models. The new structure can handle with faults varying continuously



(a) Output

(b) Input

(c) IMM

Figure 4.4. IMM-based GPC controller with low magnitude of input signals.

as random variables. In that case, faults can be modeled as discrete modes based on their cumulative distribution function.

One of the best methods for a fault detection of stochastic hybrid systems is using IMM algorithm. In our simulations, IMM system proves its higher ability to detect multiple failures of a dynamic process compared with that of GPB1 since the GPB1 estimator runs each elemental filter only once in each cycle while the input to each elemental filter in IMM is a weighted sum of the most recent estimates from all elemental filters.

Our proposed IMM-based GPC controller can provide real-time optimal control performance subject to input and output constrains and detection of failures. Simulations in this study show that the system can maintain the output set points amid failures. One of the main advantages of the GPC algorithm is that the controller can provide soft switching signals based on weighted probabilities of the outputs of different models. The tracking GPC controller does not require finding a common Lyapunov matrix as in the general cost function so that the tracking GPC controller can guarantee the stability of systems which are unstable for the direct Lyapunov method.

The main difficulty of this approach is the choice of modes on the model set as well as the transition probability matrix that assigns probabilities jumping from one mode to another since IMM algorithms are sensitive to the transition probability matrix and distance between modes. Another limitation related to IMM-based GPC controller is the magnitude of the noises and the input. When we change the output set points to small values, the input signals might become very small and this leads to unchanging mode probabilities, or IMM-based GPC controller cannot detect failures. Lastly, this approach does not consider issues of uncertainty in the control system.

References

[1] R. J. Patton, “Fault-tolerant control: the 1997 situation,” in *Proceedings of the 3rd IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes (SAFEPROCESS ’97)*, pp. 1033–1055, Hull, UK, August 1997.

[2] M. Blanke, Z. R. Izadi, S. A. Bogh, and C. P. Lunau, “Fault-tolerant control systems—a holistic view,” *Control Engineering Practice*, vol. 5, no. 10, pp. 693–720, 1997.

[3] M. Bodson and J. E. Groszkiewicz, “Multivariable adaptive algorithms for reconfigurable flight control,” *IEEE Transactions on Control Systems Technology*, vol. 5, no. 2, pp. 217–229, 1997.

[4] G.-H. Yang, S.-Y. Zhang, J. Lam, and J. Wang, “Reliable control using redundant controllers,” *IEEE Transactions on Automatic Control*, vol. 43, no. 11, pp. 1588–1593, 1998.

[5] Y. M. Zhang and J. Jiang, “Integrated design of reconfigurable fault-tolerant control systems,” *Journal of Guidance, Control, and Dynamics*, vol. 24, no. 1, pp. 133–136, 2001.

[6] M. S. Branicky, “Multiple Lyapunov functions and other analysis tools for switched and hybrid systems,” *IEEE Transactions on Automatic Control*, vol. 43, no. 4, pp. 475–482, 1998.

[7] A. Bemporad, W. P. M. H. Heemels, and B. De Schutter, “On hybrid systems and closed-loop MPC systems,” *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 863–869, 2002.

[8] H. Nael, G. Adiwinata, and D. Panagiotis, “Fault-tolerant control of process systems using communication networks,” *AIChE Journal*, vol. 51, no. 6, pp. 1665–1682, 2005.

[9] F. Christian, “Understanding fault-tolerant distributed system,” *Communications of the ACM*, vol. 34, no. 2, pp. 56–78, 1991.

[10] X.-R. Li, “Hybrid estimation techniques,” in *Control and Dynamic Systems*, C. T. Leondes, Ed., vol. 76, pp. 213–287, Academic Press, New York, NY, USA, 1996.

[11] X.-R. Li and Y. Bar-Shalom, “Multiple-model estimation with variable structure,” *IEEE Transactions on Automatic Control*, vol. 41, no. 4, pp. 478–493, 1996.

[12] X.-R. Li, Z.-L. Zhao, P. Zhang, and C. He, “Model-set design, choice, and comparison for multiple-model approach to hybrid estimation,” in *Proceedings of the Workshop on Signal Processing, Communications, Chaos and Systems*, pp. 59–92, Newport, RI, USA, June 2002.

[13] M. Basseville, “Detecting changes in signals and systems—a survey,” *Automatica*, vol. 24, no. 3, pp. 309–326, 1988.

[14] P. M. Frank, “Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results,” *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[15] J. Gertler, “Survey of model-based failure detection and isolation in complex plants,” *IEEE Control Systems Magazine*, vol. 8, no. 6, pp. 3–11, 1988.

[16] R. Isermann, “Process fault detection based on modeling and estimation method—a survey,” *Automatica*, vol. 20, no. 4, pp. 387–404, 1984.

[17] R. Patton, P. Frank, and R. Clark, *Fault Diagnosis in Dynamic Systems, Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[18] Y. Bar-Shalom and E. Tse, “Tracking in a cluttered environment with probabilistic data association,” *Automatica*, vol. 11, no. 5, pp. 451–460, 1975.

[19] A. G. Jaffer and S. C. Gupta, “On estimation of discrete processes under multiplicative and additive noise conditions,” *Information Science*, vol. 3, no. 3, pp. 267–276, 1971.

[20] C. B. Chang and M. Athans, “State estimation for discrete systems with switching parameters,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 14, no. 3, pp. 418–425, 1978.

[21] G. Ackerson and K. Fu, “On state estimation in switching environments,” *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 10–17, 1970.

[22] J. Tugnait, “Detection and estimation of abruptly changing systems,” *Automatica*, vol. 18, no. 5, pp. 607–615, 1982.

[23] Y. Zhang and X.-R. Li, “Detection and diagnosis of sensor and actuator failures using IMM estimator,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1293–1313, 1998.

[24] M. S. Branicky, V. S. Borkar, and S. K. Mitter, “A unified framework for hybrid control: model and optimal control theory,” *IEEE Transactions on Automatic Control*, vol. 43, no. 1, pp. 31–45, 1998.

[25] S. Hedlund and A. Rantzer, “Optimal control of hybrid systems,” in *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 4, pp. 3972–3976, Phoenix, Ariz, USA, December 1999.

[26] A. Rantzer and M. Johansson, “Piecewise linear quadratic optimal control,” *IEEE Transactions on Automatic Control*, vol. 45, no. 4, pp. 629–637, 2000.

[27] P. Riedinger, F. Kratz, C. Iung, and C. Zanne, “Optimal control of hybrid systems,” in *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 3, pp. 3059–3064, Phoenix, Ariz, USA, December 1999.

[28] X. Xu and P. Antsaklis, “An approach to switched systems optimal control based on parameterization of the switching instants,” in *Proceedings of the 15th IFAC World Congress*, Barcelona, Spain, July 2002.

[29] D. Clarke, C. Mohtadi, and P. Tuffs, “Generalized predictive control—part I. The basic algorithm,” *Automatica*, vol. 23, no. 2, pp. 137–148, 1987.

[30] D. Clarke, C. Mohtadi, and P. Tuffs, “Generalized predictive control—part II. Extension and Interpretations,” *Automatica*, vol. 23, no. 2, pp. 149–160, 1987.

[31] M. Kinnaert, “Adaptive generalized predictive controller for MIMO systems,” *International Journal of Control*, vol. 50, no. 1, pp. 161–172, 1989.

[32] G. Ferrari-Trecate, D. Mignone, and M. Morari, “Moving horizon estimation for hybrid systems,” in *Proceedings of the American Control Conference (ACC ’00)*, pp. 1684–1688, Chicago, Ill, USA, June 2000.

[33] A. Bemporad, D. Mignone, and M. Morari, “Fault detection and state estimation for hybrid systems,” in *Proceedings of the American Control Conference (ACC ’99)*, San Diego, Calif, USA, June 1999.

[34] H. Michalska and D. Q. Mayne, “Moving horizon observers,” in *Proceedings of the 2nd IFAC Symposium on Nonlinear Control Systems Design*, M. Fliess, Ed., Bordeaux, France, June 1992.

[35] K. Muske and J. B. Rawlings, “Nonlinear moving horizon state estimation,” in *Methods of Model Based Process Control*, pp. 349–365, Kluwer, Boston, Mass, USA, 1995.

[36] 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.

[37] S. Qin and T. Badgwell, “An overview of industrial model predictive control technology,” in *Proceedings of the 5th International Conference on Chemical Process Control*, vol. 93 of *AIChE Symposium Series, no. 316*, pp. 232–256, Tahoe City, Calif, USA, 1997.

[38] V. T. Minh and A. Nitin, “Robust model predictive control for input saturated and soften state constraints,” *Asian Journal of Control*, vol. 7, no. 3, pp. 323–329, 2005.

Vu Trieu Minh: The Sirindhorn International Thai-German Graduate School of Engineering
(TGGS), King Mongkut’s Institute of Technology North Bangkok, 1518 Pibulsongkram, Bangsue,
Bangkok 10800, Thailand
Email address: vutrieuminh@kmitnb.ac.th

Nitin Afzulpurkar: Mechatronics, Industrial System Engineering Group (ISE),
School of Engineering and Technology, Asian Institute of Technology, Klong Luang,
Pathumthani 12120, Thailand
Email address: nitin@ait.ac.th

W. M. Wan Muhamad: Institute of Product Design and Manufacturing, Universiti Kuala Lumpur,
Kuala Lumpur 50250, Malaysia
Email address: drwmansor@iprom.unikl.edu.my

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 12361, 19 pages
doi:10.1155/2007/12361

Research Article

Joint Dynamics Modeling and Parameter Identification for Space Robot Applications

Adenilson R. da Silva, Luiz C. Gadelha de Souza, and Bernd Schäfer

Received 31 August 2006; Accepted 5 March 2007

Recommended by José Manoel Balthazar

Long-term mission identification and model validation for in-flight manipulator control system in almost zero gravity with hostile space environment are extremely important for robotic applications. In this paper, a robot joint mathematical model is developed where several nonlinearities have been taken into account. In order to identify all the required system parameters, an integrated identification strategy is derived. This strategy makes use of a robust version of least-squares procedure (LS) for getting the initial conditions and a general nonlinear optimization method (MCS—multilevel coordinate search—algorithm) to estimate the nonlinear parameters. The approach is applied to the intelligent robot joint (IRJ) experiment that was developed at DLR for utilization opportunity on the International Space Station (ISS). The results using real and simulated measurements have shown that the developed algorithm and strategy have remarkable features in identifying all the parameters with good accuracy.

1. Introduction

Modeling and simulation of the dynamic behavior in a microgravity environment is mandatory for mission success. Not only the reduced gravity is severely changing the dynamic behavior, but also often much more strongly, it is the outer space environment that impacts on physical parameters like joint, structural damping, stiffness in gears and limb structures. That hostile environment is mainly due to big temperature oscillations, solar irradiation, eclipse phases, and hard space radiation. These influences affect predominantly the material, the lubrication properties of space manipulators [1], and other servomechanisms, especially in long-term mission applications. As a result, it is desired a proper knowledge of the time-dependent variances of material behavior in terms of their relevant physical parameters which in turn affects the governing differential equations of motion. Space technology demonstration experiments are required to validate the proposed strategies and algorithms for physical parameter identification. The effect of reduced gravity, temperature [2], and the expected physical parameter change due to material degradation act severely on the proper joint nonlinear dynamic modeling process. These variation especially effects the backlash, friction, stiffness, and control of space manipulators systems. In-flight systems parameters identification, both online and offline versions as well as dynamic model validation are, therefore, a very important pre-requisite to increased confidence in the modeling process.

Slow motion of any mechanical machine has been found to exhibit a highly non-linear friction [3] behavior like: stribeck effect, stick-slip [4], periodic cycle alternating motion, arrest, and so forth. It is well known that the friction and stiffness effect can strongly affect the performance of the robot arm control system, thus, the entire mission success may directly depend on the accuracy of the modeling. In order to obtain a good description of the system, especially in low velocity operation, the nonlinear friction models should be taken into account. However, the identification of nonlinear parameters is extremely difficult to deal with due to the problems of local minima, initial condition, computation time, and so forth. Previous works [3] have reported algorithms that have run time of several days. Such algorithms are almost impracticable if the identification procedure must be performed more than one time, as is the case for space applications, where one is interested in monitoring the parameters time-varying behavior. In this work, a balance between complexity and accuracy is made in order to have a model that accurately describes the friction and stiffness behavior, but also allowing the identification process to be practicable. A friction model that takes into account both, low and high velocity effects, has been derived. The identification strategy uses two versions of LS to identify the parameters, which are linearly dependent upon the measurements. For the nonlinear parameters, a nonlinear global optimization algorithm based on multilevel coordinate search (MCS) [5] has shown a good compromise between accuracy and computation time.

2. Experiment description

The IRJ experiment (Figure 2.1) developed at DLR—Institute of Robotics and Mechatronics, has served as an experimental setup. The design and construction of IRJ incorporate new features like no bulk wiring on the joint and also a number of sensors that monitor the joint performance. The joints are based on special light-weight harmonic drive (HD) gears, while measuring with high precision all relevant state variables: (a) on the input side, motor angular position and speed via an analogous hall sensor, and commanded electric current, (b) on the output side, off-drive position by using optoelectronics, and a torque measurement device based on strain gauge systems. The sensors used give a high degree of intelligence to the joint. In some tests, two accelerometers have been also attached on the top of the link to measure the acceleration in radial and tangential directions. The motors used are inland brushless DC type, which were redesigned by DLR to provide hollow axes where all cabling are fed through. DLR has also developed a lightweight small robot system with a total weight of less than 20 kg and a length of

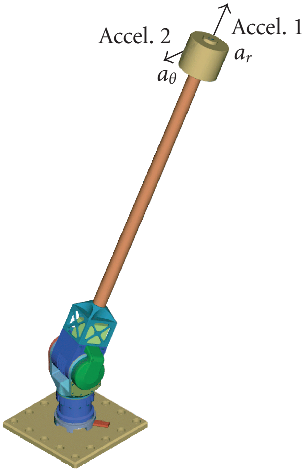


Figure 2.1. IRJ experimental setup of two joint configurations for identification purposes.

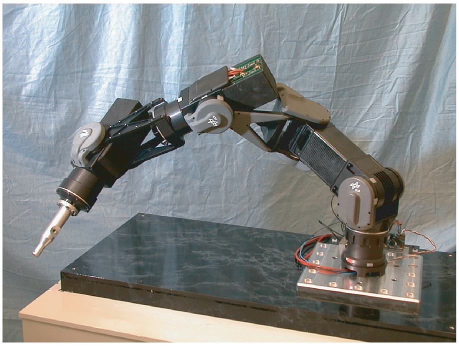


Figure 2.2. Prototype of DLR lightweight robot.

1.50 m (Figure 2.2). This design allows a very favorable payload to total weight ratio of about 1 : 3 to almost 1 : 2, compared to conventional industrial robots of more than 1 : 20.

This new design makes the robot very attractive for space-based demonstration missions as on ISS. Currently, there are some studies underway to contemplate about the space experimental use and possible accommodation opportunities at the ISS. However, if not the entire robot system is likely to be operational in the ISS early opportunity utilization phase, the IRJ experiment more probably is expected to get ready for experimental usage. The IRJ experiment will consist of a combination of two of such intelligent rotary joints. The two axes are kinematically combined in order to build up a roll-pitch configuration (Figure 2.2).

3. Joint dynamics modeling

The main emphasis of the intended space-based identification experiments is directed towards obtaining modeling confidence by proper knowledge of the time-varying joint dynamics parameters, mainly viscous damping, friction/stiction effects, and elasticity within the gears, all of those expecting to be of strongly nonlinear nature. Therefore, the following investigations have been restricted to the modeling and understanding of the nonlinear dynamics of one single intelligent joint. More complex models have already been elaborated for a two-joint configuration and also multibody models have been developed for the seven joint configurations, that is, the entire robotic system, using multibody [1] software code for model generation and simulation. Appropriate identification algorithms have been studied [6, 7] and others are still being developed and are underway for these multidegree of freedom systems.

The mathematical model to be used in the identification process is based on Newton's laws that are used to determine the dynamic force interactions and to derive the equations of motions of the joint. Here, only the main steps of the derivation are focused, a detailed description of the modeling is found in [6].

In the joints in the IRJ experiment, the wave generator (wg) is driven by a motor mounted to the circular spline (cs) and the flexspline is attached to the ground. The output is driven by the circular spline. Damping torques, both at the input and output side, have been considered. According to Figures 3.1 and 3.2, while making use only of the pitch (θ) rotary joint, the equations of motion for the IRJ can be described by

$$\begin{aligned} J_{\text{in}}\ddot{\theta}_{\text{in}} &= T_m - T_{d\_\text{in}} - T_{\text{wg}}, \\ J_{\text{out}}\ddot{\theta}_{\text{out}} &= T_{\text{cs}} - T_{d\_\text{out}} - T_{d\_\text{fscs}} + T_{\text{load}}, \end{aligned} \tag{3.1}$$

where $T_m = K_m I_a$ is the applied motor torque with motor constant K_m and electric current I_a. J_{in} and J_{out} are the input and output inertia, θ_{in} and θ_{out} the respective angular positions. The elasticity within the HD gear is given by the stiffness torque T_{stiff} with $T_{\text{wg}} = T_{\text{stiff}} + T_{d\_\text{wg}}$ on the input side of the gear and $T_{\text{cs}} = (N+1)T_{\text{wg}}$. N is the gear reduction.

The various damping torques are denoted by T_d, attributed with appropriate indices. The applied load on the link (Figure 2.1) side is due to gravity and is given by $T_{\text{load}} = \hat{T}_g \sin\theta_{\text{out}}$ with the load amplitude $\hat{T}_g$.

Based upon the HD manufacturer's catalog values, this gear type typically exhibits the well-known nonlinear behavior. Usually, the dependency between applied torque and the relative angular position $\Delta\theta$ ($\theta_{\text{in}} - \theta_{\text{out}}$) is given by a combination of piecewise linear functions, $T_{\text{stiff}} = f(\Delta\theta)$, depending upon the operational range of the gear. For the identification algorithm to be developed further, it is necessary to replace this piecewise linear behavior by a continuous curve. As a first approach, it has been proved sufficient to apply a third-order polynomial to represent the stiffness torque given by

$$T_{\text{stiff}} = k_1\Delta\theta + k_2(\Delta\theta)^3 \tag{3.2}$$

with coefficients k_1 and k_2 to be adjusted.

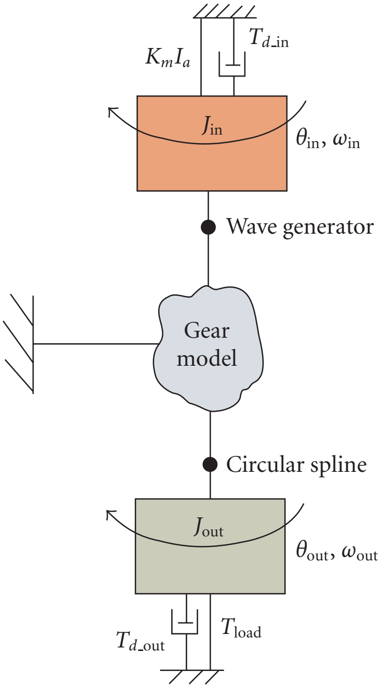


Figure 3.1. Dynamic representation of the intelligent robotic joint (IRJ).

Moreover, regarding the mechanical nature of the torque measurement system with strain gauges attached to spokes and rings, it may be worthwhile to account also for some compliance in that system. This is necessary to model it as a further spring, being serially connected to the HD gear spring. In total, this would result in a combined softer spring, and can be considered within new stiffness constants k_1 and k_2 that now would enter as unknown parameters within the identification algorithm.

According to experimental results of many authors [3, 6], the damping torques T_d that appear on the input side, the output side, and inside the HD gear are assumed to capture two facets of damping behavior, namely T_{visc} and T_{fric}. These are a viscous and a dry friction or Coulomb-type part. Thus, total damping torques is written as

$$T_d = T_{\text{visc}} + T_{\text{fric}}, \tag{3.3}$$

the viscous part can be strongly nonlinear with a cubic relationship in the angular velocity,

$$T_{\text{visc}} = b_1\dot{\theta} + b_2\dot{\theta}^3 \tag{3.4}$$

with the linearly depending coefficients b_1 and b_2. For the dry friction, a modified classical Coulomb friction model is required. This is necessary to account for the well-known

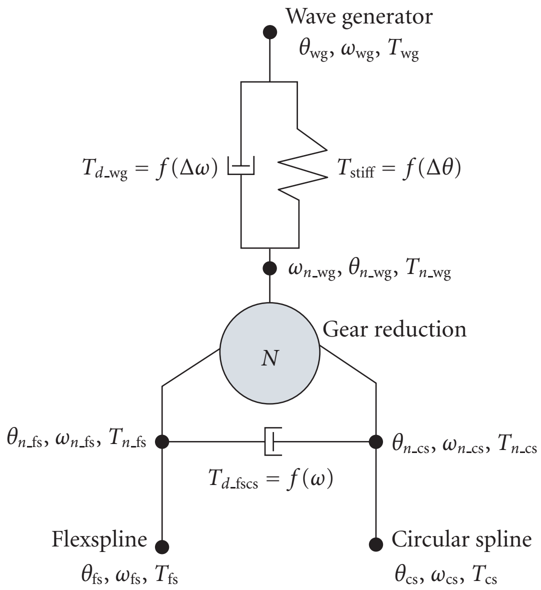

Wave generator
$\theta_{\text{wg}}, \omega_{\text{wg}}, T_{\text{wg}}$
$T_{d\_\text{wg}} = f(\Delta\omega)$
$T_{\text{stiff}} = f(\Delta\theta)$
$\omega_{n\_\text{wg}}, \theta_{n\_\text{wg}}, T_{n\_\text{wg}}$
Gear reduction
N
$\theta_{n\_\text{fs}}, \omega_{n\_\text{fs}}, T_{n\_\text{fs}}$
$\theta_{n\_\text{cs}}, \omega_{n\_\text{cs}}, T_{n\_\text{cs}}$
$T_{d\_\text{fscs}} = f(\omega)$
Flexspline
$\theta_{\text{fs}}, \omega_{\text{fs}}, T_{\text{fs}}$
Circular spline
$\theta_{\text{cs}}, \omega_{\text{cs}}, T_{\text{cs}}$

Figure 3.2. Harmonic drive gear model (wave generator wg, circular cs, and flexible fs spline).

Stribeck effects. This observes the fact that for low velocities, the friction torque is normally decreasing continuously with increasing velocity, not in a discontinuous manner. Another problem that arise in using the classical Coulomb friction model is the discontinuity at zero velocity. In order to account both problems, Stribeck effects and zero discontinuity, an empirical mathematical model has been adopted,

$$T_{\text{fric}} = \left| T_N \right| \cdot \left(\mu \cdot \tanh\left(\frac{\dot{\theta}_i}{\omega_1} \right) + \frac{\dot{\theta}_i}{\omega_2} \cdot e^{-|\omega/\omega_S|^{\delta_S}} \right), \tag{3.5}$$

where T_N is the normal torque, μ is the friction coefficient, ω_S is the Stribeck velocity, $i = \text{in}, \text{out}$, δ_S is the exponential parameter that is commonly taken either as 0.5, 1 or even 2. Another possibility is to let δ_S to be identified by the nonlinear part of the algorithm together with ω_1 and ω_2.

According to [8], periodic variations in the frictional torque might appear in the HD operation. Thus, we have introduced periodic variations in the frictional torque on the HD output,

$$T_{\text{cyclic}} = A_{\text{cyclic}} \sin\left(\theta_{\text{out}} + \gamma_{\text{cyclic}}\right). \tag{3.6}$$

As this relationship indicates, frictional torque fluctuations of amplitude A_{cyclic} complete one cycle every time the flexspline makes one complete rotation relative to the circular spline. To match this model to experimental observations, a phase shift of γ_{cyclic} is also included. In order to obtain a linear dependency of the two parameters, this relationship can be easily transformed to

$$T_{\text{cyclic}} = A_1 \sin \theta_{\text{out}} + A_2 \cos \theta_{\text{out}} \tag{3.7}$$

with the new linearly depending parameters $A_1 = A_{\text{cyclic}} \cos \gamma_{\text{cyclic}}$ and $A_2 = A_{\text{cyclic}} \sin \gamma_{\text{cyclic}}$, from where A_{cyclic} and γ_{cyclic} can be recovered.

It is self-evident that not all of the envisaged damping torques given in (3.1) are expected to capture both types, that is, viscous damping and dry friction parts. Where appropriate, only linear viscous damping is considered in order to keep the amount of parameters to be identified at a minimum, as well as the complexity of the joint dynamic model. It has to be kept in mind that the final manipulator configuration consists of seven kinematic degrees of freedom, which otherwise would drive the amount of parameters intensively high. Recalling the given kinematic constraints, the various torques in (3.1) can now be formulated in terms of the input and output positions, θ_{in} and θ_{out}, and their respective velocities

$$\begin{aligned} J_{\text{in}} \ddot{\theta}_{\text{in}} &= K_m I_a - T_{d\_\text{in}}(\dot{\theta}_{\text{in}}, \theta_{\text{in}}) - T_{d\_\text{wg}}(\dot{\theta}_{\text{in}} - (N+1)\dot{\theta}_{\text{out}}) - T_{\text{stiff}}(\theta_{\text{in}} - (N+1)\theta_{\text{out}}), \\ J_{\text{out}} \ddot{\theta}_{\text{out}} &= (N+1) \cdot \left[T_{\text{stiff}}(\theta_{\text{in}} - (N+1)\theta_{\text{out}}) + T_{d\_\text{wg}}(\dot{\theta}_{\text{in}} - (N+1)\dot{\theta}_{\text{out}}) \right] - T_{d\_\text{out}}(\dot{\theta}_{\text{out}}) \\ &\quad - T_{d\_\text{fscs}}(\theta_{\text{out}}, \dot{\theta}_{\text{out}}) + \widehat{T}_g \sin \theta_{\text{out}}. \end{aligned} \tag{3.8}$$

The damping dependent on the position that appears in T_{d_in} is related to Dahl effect [4]. It is necessary to include a position-dependent term also on the input side in order to get good agreement between dynamic model and measured data.

4. Identification model and strategy

In order to identify the dynamic parameters of the robotic joint, (3.8) have been taken as the dynamic model representation for the identification process. The problem of identifying, especially rigid body dynamics parameters of a robot, has been extensively studied and a vast amount of literature can be found [9–11]. However, these methods have a common idea: the robot is moved along a selected trajectory while the joint motion and torques are measured. Then, the parameters are offline estimated using a standard offline LS-based technique. In addition, most of these works have used an industrial robot as a test bed. The strategy and algorithm proposed in this paper should guarantee to cope with several requirements, like online procedure, ability to track time-variant parameters, possibility to identify parameters with nonlinear dependency with respect to the measurements in fast way at low-computational cost.

For the algorithm development, (3.8) are rewritten in order to set up a linear combination of the unknown parameters, given by the vector Θ, and the known information, given by the measurement vector ϕ. The parameters that appear in vector Θ are identified by an RLS with variable forgetting [7] factor and the parameters ω_1 and ω_2 which appear inside the matrix ϕ are identified by the MCS algorithm. The measured signals are θ_{in}

and I_a on the input side, θ_{out}, $\ddot{\theta}_{\text{out}}$, and T_{out} on the output side. The respective velocities $\dot{\theta}_{\text{in}}$ and θ_{out} and the acceleration signal $\ddot{\theta}_{\text{in}}$ are calculated numerically while regarding filtering techniques to remedy bad measurement signals.

The following specific torque functional relationships have been considered for the dynamic model:

$$T_{d\_\text{in}} = b_{\text{in}}\dot{\theta}_{\text{in}} + |\,T_N\,| \cdot \mu \tanh \cdot \left(\frac{\dot{\theta}_{\text{in}}}{\omega_1}\right) + |\,T_N\,| \cdot \frac{\dot{\theta}_{\text{in}}}{\omega_2} \cdot e^{-|\dot{\theta}_{\text{in}}/\omega_S|^{\delta_S}} + b_{\text{in}_D}\theta_{\text{in}}, \tag{4.1}$$

$$T_{\text{stiff}} = k_1(\theta_{\text{in}} - (N+1)\theta_{\text{out}}) + k_2(\theta_{\text{in}} - (N+1)\theta_{\text{out}})^3 = k_1\Delta\theta + k_2(\Delta\theta)^3, \tag{4.2}$$

$$T_{d\_\text{fscs}} = b_{\text{fscs}\,1}\operatorname{sign}(\dot{\theta}_{\text{out}}) + A_1 \sin\theta_{\text{out}} + A_2 \cos\theta_{\text{out}}, \tag{4.3}$$

$$T_{d\_\text{out}} = b_{\text{out}1}\dot{\theta}_{\text{out}}, \tag{4.4}$$

where $T_{\text{wg}} = T_{n\_\text{wg}} = T_{\text{cs}}/(N+1) = T_{\text{out}}/(N+1)$. Then, the identification model in the linear regression format can be described by

$$Y_k = \phi \cdot \Theta^T, \tag{4.5}$$

where

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} J_{\text{in}}\ddot{\theta}_{\text{in}} - K_m I_a \\ J_{\text{out}}\ddot{\theta}_{\text{out}} - \widehat{T}_g \sin\theta_{\text{out}} \end{bmatrix} = \begin{bmatrix} J_{\text{in}}\ddot{\theta}_{\text{in}} - K_m I_a \\ T_{\text{out}} \end{bmatrix},$$

$$\phi = \begin{bmatrix} -\Delta\theta & -\Delta\theta^3 & -\tanh\left(\frac{\dot{\theta}_{\text{in}}}{\omega_1}\right) & -\dot{\theta}_{\text{in}} \cdot e^{-|\dot{\theta}_{\text{in}}/\omega_S|} & -\dot{\theta}_{\text{in}} & -\theta_{\text{in}} & 0 & 0 & 0 & 0 \\ \Delta\theta & \Delta\theta^3 & 0 & 0 & 0 & 0 & -\dot{\theta}_{\text{out}} & -\operatorname{sign}(\dot{\theta}_{\text{out}}) & -\sin(\theta_{\text{out}}) & -\cos(\theta_{\text{out}}) \end{bmatrix},$$

$$\Theta = \begin{bmatrix} k_1 & k_2 & C_1 & C_2 & b_{\text{in}} & b_{\text{in}D} & b_{\text{out}} & b_{\text{out}C} & A_1 & A_2 \end{bmatrix}^T. \tag{4.6}$$

$C_1 = |T_N| \cdot \mu$ and $C_2 = |T_N|/\omega_2$. The exponential coefficient δ_S has been set to 1.

Using the model given by (4.5), a prediction of Y is given by

$$\widehat{Y}_k = \phi \cdot \widehat{\Theta}^T. \tag{4.7}$$

For a given discrete measurement time t_k, the predicted error to be minimized in LS sense is

$$\varepsilon_k = Y_k - \widehat{Y}_k. \tag{4.8}$$

Using the singular value decomposition (SVD) approach, the excitation level and linear combination in the information matrix is verified:

$$\phi = U\Sigma V^T, \quad \Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m) \tag{4.9}$$

with U and V being the isometric matrices. If some states are not well excited or there exist some linear combination in the ϕ matrix, the related singular value σ_i will have small magnitude, close to machine precision. After testing the matrix ϕ, the initial condition for the recursive algorithm is obtained by standard batch least squares estimation:

$$\widehat{\Theta}_{\text{initial}} = \left(\phi^T \phi\right)^{-1} \phi^T Y_k \tag{4.10}$$

and by applying (4.9), one obtains

$$\widehat{\Theta}_{\text{initial}} = V\Sigma^{-T} U^T Y_k. \tag{4.11}$$

Once the initial conditions are obtained, the recursive identification is carried out by using the algorithm described in [7].

5. Nonlinear optimization: multilevel coordinate search (MCS) algorithm

Two parameters in (4.1) have nonlinear dependency with respect to measurement data; therefore, they cannot be identified by the RLS approach. There exist several methods that can be used; local minimizer or global one. The local minimizer requires a good starting point and sometimes they deliver only a mathematical solution for the problem. In these cases, the parameters have no longer physical meaning. Most of the global algorithms have very hard computational load, making the identification process almost impracticable. In this paper, the MCS algorithm has been used in combination with RLS approach.

The MCS algorithm has a very interesting combination of local and global search of the minimum. Here, we will point out only the basic ideas of the algorithm, the interested reader is directed to the work of Huyer and Neumaier [5].

Consider the bound-constrained optimization problem

$$\min f(x), \quad x \in [u, v], \tag{5.1}$$

with finite or infinite bounds

$$[u, v] := \{x \in R^n \mid u_i \leq x_i \leq v_i,\ i = 1, \ldots, n\}. \tag{5.2}$$

With u and v being n-dimensional vectors with components in $\overline{R} := R \cup \{-\infty, \infty\}$ and $u_i < v_i$ for $i = 1, \ldots, n$, that is, only points with finite components are regarded of a box $[u, v]$ whereas its bounds can be infinite. If all the bounds are set to infinity, an unconstrained optimization problem is obtained.

The MCS algorithm tries to find the minimizer by splitting the search space into smaller boxes. These boxes contain a distinguished point, the so-called *base point*, whose function value is known. In splitting the boxes, a nonuniform procedure is used. Parts where low values of the function are expected are carefully examined. In order to speed up the computation procedure, the MCS algorithm combines global search (splitting the boxes with large parts) and local search (splitting the boxes with good function values). This gives a good balance between convergence to the global minimum and computation time.

The nonlinear algorithm requires an index of performance (IP) to be minimized. There exists several ways to define IP criteria. In this work, two criteria have been tested:

a quadratic function of the error,

$$\mathrm{IP} = \frac{1}{2}[y - \hat{y}][y - \hat{y}]^T \tag{5.3}$$

and the absolute value of the error,

$$\mathrm{IP} = \|y - \hat{y}\|, \tag{5.4}$$

where y is the plant output (friction torque) and $\hat{y}$ the estimation of y by considering the optimal linear parameters (LS estimation), and $\|\cdot\|$ means the Euclidian norm of ε.

6. Integrated algorithm LS: MCS

For solving the identification problem characterized by (4.5), an integrated algorithm using LS and MCS approach is derived. The proposed strategy is divided in two different operational modes: a starting procedure and a normal mode. In the starting procedure, the measurements are collected and stored in a batch with a preselected length. The batch of measurements is continuously updated, this work is like a moving window of measurements. Given an initial guess for the nonlinear parameters (in our case, ω_1 and ω_S), the parameters with linear dependency with respect to measurements (thereafter called just as linear parameters) are estimated by the LS part. Then the linear parameters are passed to MCS algorithm in order to estimate the nonlinear ones. This process is repeated until the convergence criteria are completely fulfilled, namely the norms of the errors are smaller than selected threshold (δ and $\delta_{\Theta_{\min}}$). When convergence criteria are fulfilled, the online identification algorithm for the linear parameters is started, and the non-linear parameters are assumed constant for the period where the norm of the errors is smaller than δ. If the error increases, the nonlinear parameters are updated by using the MCS algorithm. Using this procedure, for our example, the space of search in the identification problem is reduced from 10 to 2 for the nonlinear algorithm. This drastically reduces the computation time and the efficiency of the MCS algorithm in finding the global minimum. Thus, the integrated algorithm has an online update for the linear parameters and a random update for the nonlinear parameters.

The integrated algorithm can be summarized in the following steps.

(i) *Initial procedure:*

(1) input: u, v (boundaries for ω_1 and ω_2), $\omega_{1_{\text{initial}}}$ and $\omega_{2_{\text{initial}}}$; While $N_{\text{err}} = \|y - \hat{y}\| > \delta$ and $\Delta\Theta = (\Theta_k - \Theta_{k-1}) > \delta_{\Theta_{\min}}$;

(2) collect the measurements;

(3) compute ϕ;

(4) check rank of ϕ (SVD);

(5) estimate the Θ_L (LS part);

(6) estimate the $\breve{\Theta}_{\text{NL}}$ ($\breve{\Theta}_{\text{NL}}$ means global optimizer in the box described by $[u, v]$) coefficient (MCS algorithm);

(7) evaluate N_{err} and $\Delta\Theta$;

(8) if $N_{\text{err}} < \delta$ and $\Delta\Theta < \delta_{\Theta_{\min}}$, stop and keep Θ.

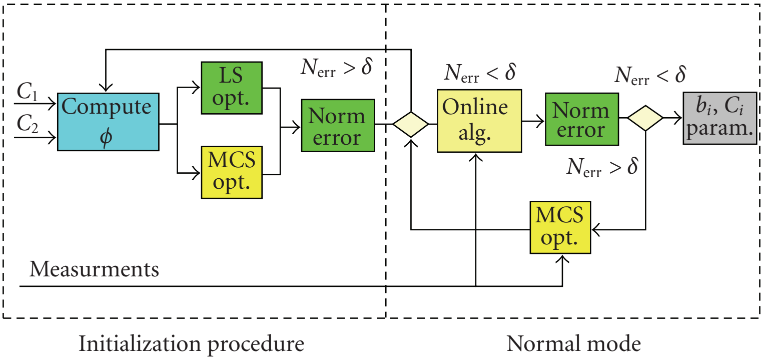


Figure 6.1. Schematic representation of the integrated identification algorithm.

(ii) *Normal mode:*

(9) using Θ, start the online algorithm;

(10) check $N_{\rm err}$;

(11) if $N_{\rm err} > \delta$, call MCS algorithm and using the latest measurement window, evaluate the new $\breve{\Theta}_{\rm NL}$ coefficient;
else $\breve{\Theta}_{\rm NL}$ is still the minimum (no change in the non-linear parameters);

(12) takes the next measurement.

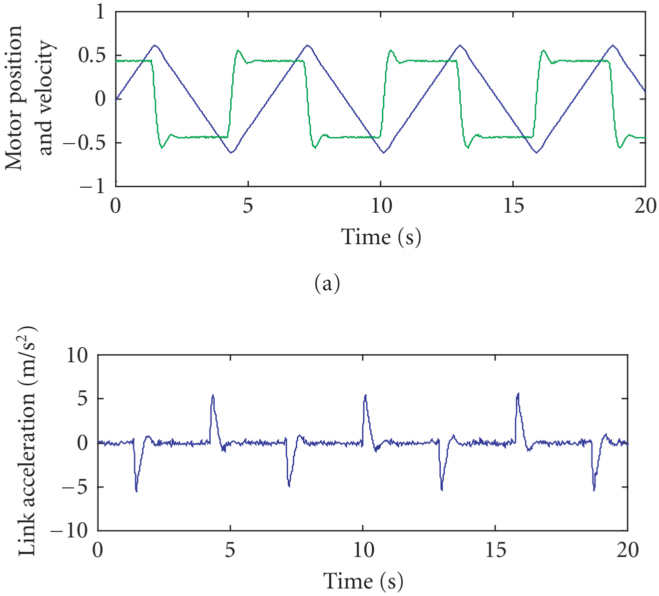
Working in this way, the proposed algorithm can track in real time variations in the linear parameters and update the nonlinear parameters only when some corrections are required. The schematic diagram of the integrated algorithm is shown in Figure 6.1.

7. Results

The proposed strategy and algorithms have been tested in two different situations: first, using only the measured information from IRJ; second, a jump in $\Theta_{\rm NL}$ has been simulated in order to check the ability of the algorithm in tracking time variations in $\Theta_{\rm NL}$. Besides, in order to have a normalized system, the data and parameters values of the motor side have been translated to link side, meaning that the gear reduction is 1. ($N = 1$).

7.1. Case 1: using measured data from IRJ. In this test, the measurements are taken from IRJ with time length of one minute. Figure 7.1, in the upper part shows the motor position and velocity by using a triangular trajectory and on the bottom, the link acceleration. In order to have better resolution, only 20 seconds of measurements are shown.

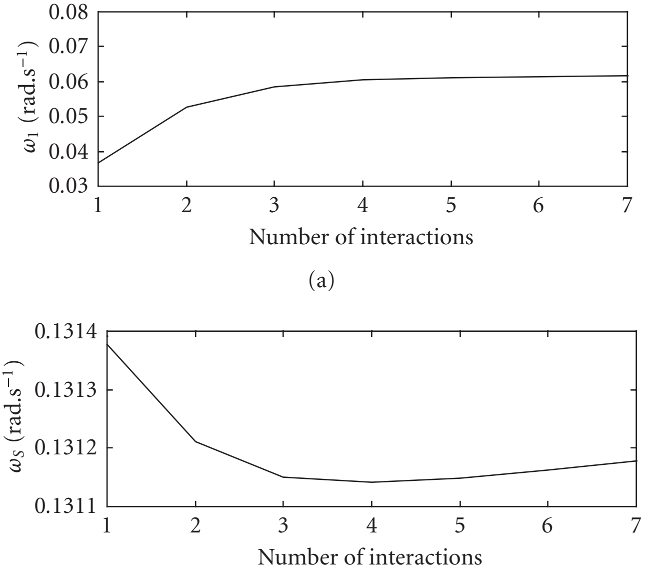
Using (4.5) as a model and the integrated algorithm, all parameters which appear in Θ have been identified. After 13 seconds collecting data, the matrix ϕ gets full rank and the starting procedure is completed. The nonlinear parameters are identified by using MCS algorithm and the linear ones are identified by a batch LS. Figure 7.2 shows the convergence process of the non-linear parameters. It can be noted that after few interactions, the convergence criterion has been fulfilled and the nonlinear optimization has been stopped.



(a)

(b)

Figure 7.1. Measurements from IRJ.



(a)

(b)

Figure 7.2. Nonlinear parameters identified by MCS algorithm.

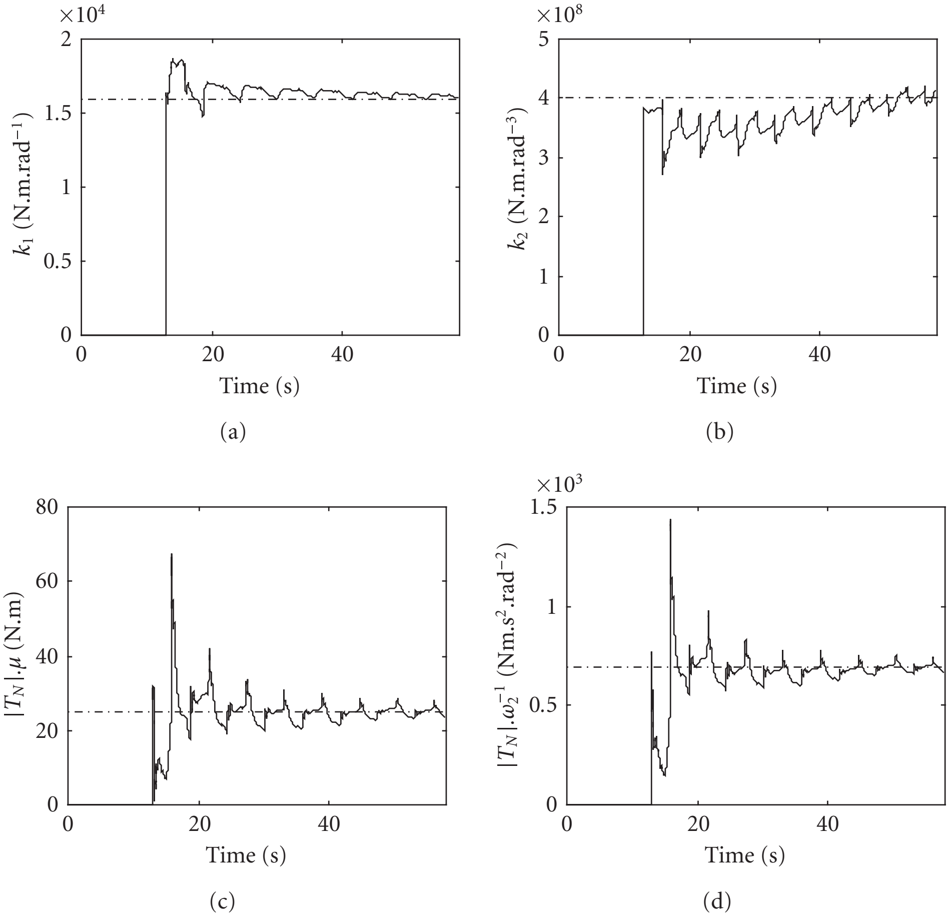


Figure 7.3. Linear parameters identified by RLS part—stiffness and nonlinear damping.

In the plots, there is a period where all parameters have zero value, this period corresponds to the initialization procedure where there is no online identification. The measurements are collected and an analysis in matrix ϕ is performed sequentially with the nonlinear and batch estimation. The stiffness coefficients and the nonlinear damping are shown in Figure 7.3. The dashed lines are constant values obtained by an offline procedure using all the data available. It can be noted that all the parameters converge to the offline estimation showing the good convergence and robustness of the RLS algorithm. As expected, the parameter related to the cubic stiffness has low rate of convergence. This fact is early observed in the singular values of the information matrix. The related singular value has the smallest magnitude meaning that this parameter is very difficult to identify. Despite of its small excitation, the cubic stiffness parameter converges to the expected mean value (offline estimation).

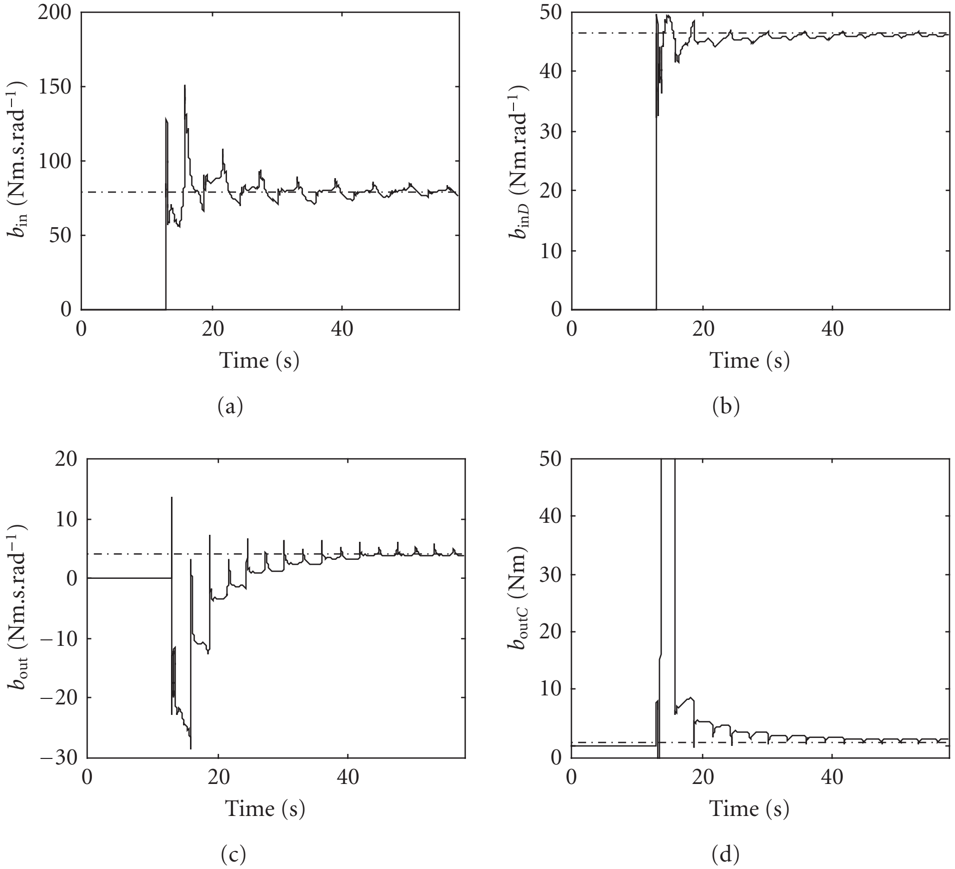
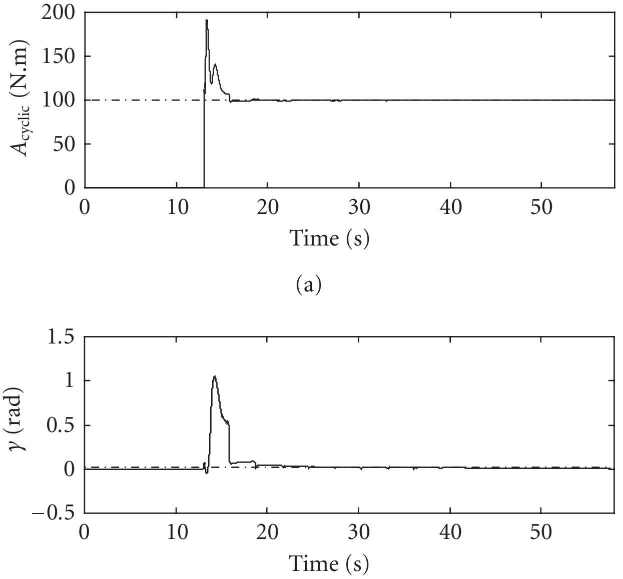


Figure 7.4. Linear parameters identified by RLS part—viscous damping.

Figures 7.4 and 7.5 show the rest of the parameters, which appear in vector Θ. It can be noted that all parameters have stable behavior converging to their expected mean values obtained by full batch identification.

Finally, Figure 7.6 shows the statistical performance of the identification process. It can be observed that the algorithm has good ability in tracking the reference signal. Most of the errors lies below 5%, the peak of the errors (20%) occurs due to the nature of the trajectory (Figure 7.1) used. In the point where the velocity changes the sign, there exists a peak in the torque and the algorithm cannot predict this high torque immediately.

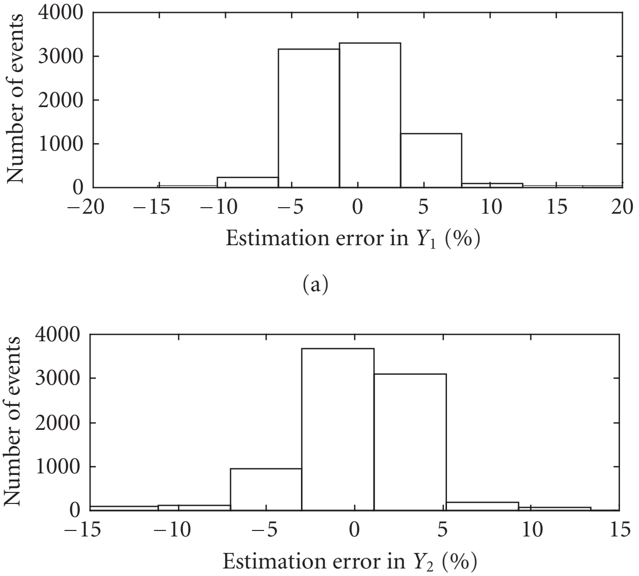
7.2. Case 2: simulation of time-variant parameters. In order to test the integrated linear and nonlinear identification procedure in case of time-variant parameters, a mixed data set has been used: the angular velocity has been taken from the experiment setup and the friction torque is calculated by setting the values of the parameters in (4.1). The parameter δ_s has been set to 1 and the other values used are shown in Table 7.1.



(a)

(b)

Figure 7.5. Linear parameters identified by RLS part—periodic damping and phase.



(a)

(b)

Figure 7.6. Estimation error.

Table 7.1. Offline estimation for the parameters.

Parameters	Value	Parameters	Value
b_{in}	76 Nm.s.rad^{-1}	$\lvert T_N \rvert \cdot \omega_2^{-1}$	590 Nm.s^2.rad^{-2}
$b_{\text{in}D}$	40 Nm.rad^{-1}	ω_1	0.0616 rad.s^{-1}
$\lvert T_N \rvert \cdot \mu$	25 Nm	ω_S	0.1312 rad.s^{-1}

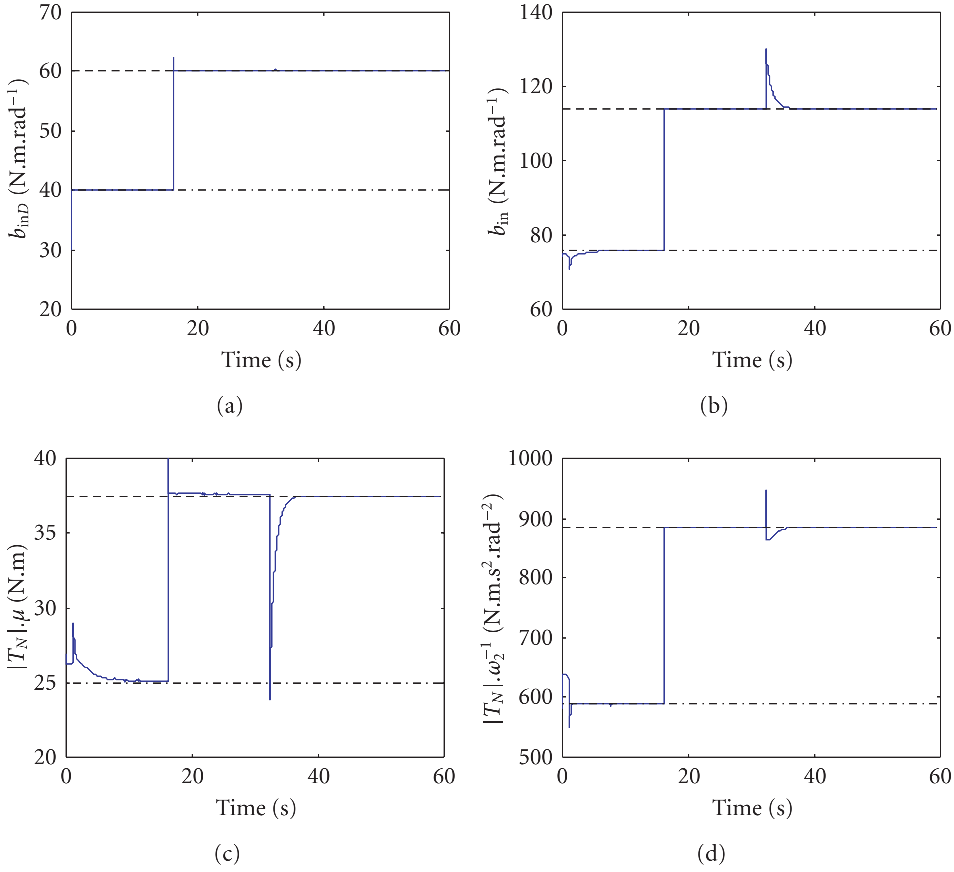


Figure 7.7. Simulation of time-variant systems—linear parameters.

By using these values, a simulated friction torque (T_{d_in}) has been obtained to be used in the test of the algorithm. In order to simulate time variant system, the parameters have experimented variations at instant 16 seconds and 32 seconds in the linear and nonlinear ones, respectively. At time 16 seconds, the plant output T_{d_in} has been recalculated and a jump of 50% in the linear parameters has been set. Immediately, the RLS algorithm is able

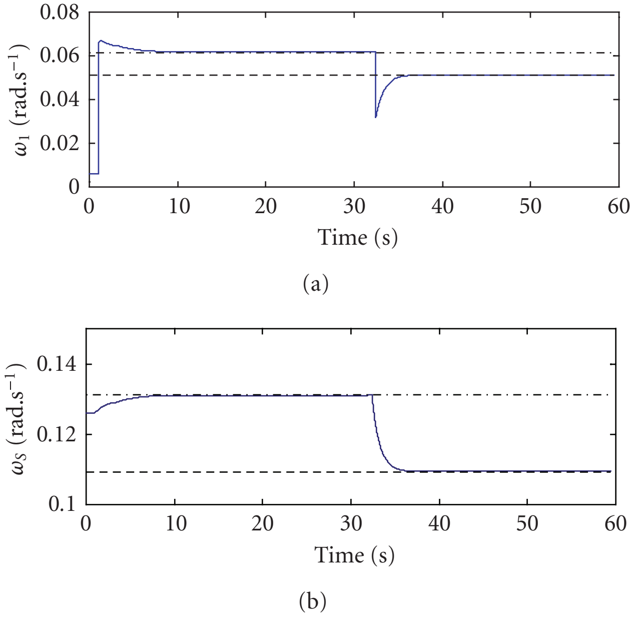


Figure 7.8. Simulation of time-variant systems—nonlinear parameters.

to notice the changes in the parameters, and from that it can estimete the new parameters values, as shown in Figure 7.7. At this time, the correction in the linear parameters is sufficient to keep the error smaller than the threshold δ. Therefore, the MCS algorithm has been not activated. The dashed-dot lines represent the parameters values before the jump and the dot lines the values after the jump.

Figure 7.7 shows that after the initialization procedure, the parameters identified by the RLS part have fast convergence to the nominal values. The jump in the linear parameters is compensated avoiding the nonlinear optimization. When the nonlinear parameters are changed, the linear ones are affected (peaks in Figure 7.7), but according the convergence in the nonlinear one is reached, the linear parameters approach to the correct values.

At instant $t = 32$ seconds, the nonlinear parameters have been changed by 20% of their initial values. Then, the norm of error increases and the corrections in the linear parameters are not sufficient to decrease it. Thus, the MCS is activated and the nonlinear parameters are recalculated. When the norm of the error decreases, the nonlinear optimization is stopped and only the fast (RLS) part of the algorithm is running. Figure 7.8 shows the behavior of the nonlinear parameters, it can be noted that the algorithm has fast convergence in both situation: in the initialization and when are recalculated. Due to fast corrections in the linear parameters, change in these parameters does not affect the nonlinear ones. On the other hand, the linear parameters are affected when corrections in the nonlinear ones are required. This occurs because the corrections in the non-linear parameters are not so fast. Due to this feature, the procedure presented here has very low-computational load allowing one to track time-variant systems, which contain nonlinear parameters.

8. Conclusions

In this work, the complete model of the robotic joint has been derived. The obtained model takes into account several nonlinearities; as for the stiffness as well as in the friction model. The typical nonlinear behavior of the friction at low velocity has been taken into account. An integrated (independent linear and nonlinear parts) identification algorithm has been derived and tested by using data from IRJ experiment and also a mixed data to simulate time-variant systems.

The results have shown that strategy presented gives excellent precision at very low-computational cost; the integrated algorithm is more than 20 times faster than the completely nonlinear counterpart (if all the parameters is to be identified by MCS algorithm alone). This allows an online identification for almost all of the measurement period, except for a short period, when an update in the nonlinear parameters is necessary; the online identification was not possible. The ability in tracking time-variant parameters has been also tested by using simulated data and the results have shown a fast and accurate response to the variations in both set of parameters: linear and nonlinear ones. Another very important feature of the proposed approach is that there is no necessity of initial guess for all the parameters; they are automatically adjusted by the initialization procedure. It is only required to set the boundary for the nonlinear parameters, even though this is not a requirement but save computation time.

References

[1] R. Krenn and B. Schäfer, “Limitations of hardware-in-the-loop simulations of space robotics dynamics using industrial robots,” in *Proceedings of the 5th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS ’99)*, Noordwijk, The Netherlands, June 1999.

[2] M. Grotjahn, M. Daemi, and B. Heimann, “Friction and rigid body identification of robot dynamics,” in *Proceedings of the 6th Pan American Congress of Applied Mechanics (PACAM ’99)*, pp. 1557–1560, Rio do Janeiro, Brazil, January 1999.

[3] B. Armstrong, *Control of Machines with Friction*, Kluwer Academic Publishers, Boston, Mass, USA, 1991.

[4] O. Henrik, *Control systems with friction*, Ph.D. thesis, Lund Institute of Technology, Lund, Sweden, 1996.

[5] W. Huyer and A. Neumaier, “Global optimization by multilevel coordinate search,” *Journal of Global Optimization*, vol. 14, no. 4, pp. 331–355, 1999.

[6] B. Schäfer and A. R. da Silva, “Space robotics experiments for increasing dynamic modeling fidelity,” in *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS ’00)*, Barcelona, Spain, September 2000.

[7] A. R. da Silva, B. Schäfer, L. C. G. de Souza, and R. A. Fonseca, “Space robotics joints nonlinear modeling and on-line parameters identification,” in *Proceedings of the 31st International Symposium on Robotics (ISR ’00)*, pp. 461–467, Montreal, Canada, May 2000.

[8] T. D. Tuttle, “Understanding and modeling the behavior of harmonic drive gear transmission,” Master dissertation, Massachusetts Institute of Technology, Cambridge, Mass, USA, 1992.

[9] M. Daemi and M. Grotjahn, “Practical experiences with L. S. methods for the identification of robot dynamics,” in *Proceedings of the 2nd ECPD International Conference on Advanced Robotics*, pp. 535–540, Vienna, Austria, September 1996.

[10] M. Gautier and W. Khalil, “On the identification of the inertial parameters of robots,” in *Proceedings of the 27th IEEE Conference on Decision and Control*, pp. 2264–2269, Austin, Tex, USA, December 1988.

[11] H. B. Olsen and G. A. Bekey, “Identification of parameters in models of robots with rotary joints,” in *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 1045–1049, Saint Louis, Mo, USA, March 1985.

Adenilson R. da Silva: Space System Division, National Institute for Space Research (INPE),
Avenida dos Astronautas 1758, 12201-970 São José dos Campos, SP, Brazil
Email address: adenilson.silva@dss.inpe.br

Luiz C. Gadelha de Souza: Space Mechanics and Control Division, National Institute for Space
Research (INPE), Avenida dos Astronautas 1758, 12201-970 São José dos Campos, SP, Brazil
Email address: gadelha@dem.inpe.br

Bernd Schäfer: Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen,
82234 Wessling, Germany
Email address: bernd.schaefer@dlr.de

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 28262, 9 pages
doi:10.1155/2007/28262

Research Article

Quadratic Stabilization of LPV System by an LTI Controller Based on ILMI Algorithm

Wei Xie

Received 8 August 2006; Revised 23 February 2007; Accepted 25 May 2007

Recommended by José Manoel Balthazar

A linear time-invariant (LTI) output feedback controller is designed for a linear parameter-varying (LPV) control system to achieve quadratic stability. The LPV system includes immeasurable dependent parameters that are assumed to vary in a polytopic space. To solve this control problem, a heuristic algorithm is proposed in the form of an iterative linear matrix inequality (ILMI) formulation. Furthermore, an effective method of setting an initial value of the ILMI algorithm is also proposed to increase the probability of getting an admissible solution for the controller design problem.

1. Introduction

A linear parameter-varying (LPV) system is formalized as a certain type of nonlinear system, and is successfully applied in developing a control strategy which is based on classical gain-scheduled methodology [1]. Several tutorial papers and special publications concerning the gain-scheduled method of LPV control system are [2–7]. These gain-scheduled LPV controller design approaches are applicable under the assumption that the dependent parameters can be measured online. In practical application, it is often difficult to satisfy this requirement. Therefore, it is crucial to design an effective LTI controller to get robust stability for an LPV plant with immeasurable dependent parameters. Here, these dependent parameters are assumed to vary in a polytopic space. In robust control framework of LPV system, a necessary and sufficient condition of quadratic stability for polytopic LPV system is formulated in terms of a finite LMIs optimization problem [8]. The underlying quadratic Lyapunov functions are also used to derive bounds on robust performance measures. Several heuristic procedures [9–13] have also been proposed to solve some control problems with nonconvex constraints such as a controller with fixed or reduced order of the decentralized structure. In [10], a method is presented to solve some controller design problems when structure constraints are imposed. The procedure is based on a two-stage optimization process, each stage requires the solution of a convex optimization problem based on a kind of LMI expression, in which either the controller gain matrix or the Lyapunov function is considered as the optimization variable.

This paper proposes a way of designing a quadratically stabilizing LTI output feedback controller for LPV system where dependent parameters vary in a polytopic space. Different from gain-scheduled LPV controller design, besides rank constraints, another constraint condition in which the controller matrix should be the same one for each vertex plant of LPV system is added. This problem still remains a complex issue and not numerically tractable. Here, a heuristic ILMI approach is presented to solve an admissible solution for this control problem. And a method of setting an initial value for the Lyapunov matrix is also proposed to increase the possibility of obtaining a feasible solution to the ILMI approach. The proposed method is better than random assignment of the initial value. Even though this approach is not guaranteed to converge globally, it may provide a useful alternative design tool in practice.

2. Notation and definitions

Consider an LPV plant $P(\theta(t))$ described by state space equations as

$$\begin{aligned} \dot{x}(t) &= A(\theta)x(t) + B_u u(t), \\ y(t) &= C_y u(t). \end{aligned} \tag{2.1}$$

Here, state-space matrices have compatible dimensions of time-varying dependent parameters $\theta(t) = [\theta_1(t)\theta_2(t)\cdots\theta_r(t)]^T \in \mathbb{R}^r$. Moreover, we have the following assumptions.

(1) The system state matrix $A(\theta)$ is a continuous and bounded function and depends affinely on $\theta(t)$.

(2) The immeasurable real parameters $\theta(t)$ exist in the LPV plant and vary in a polytope Θ as

$$\begin{aligned} \theta(t) \in \Theta &:= Co\{\omega_1, \omega_2, \ldots, \omega_N\} \\ &= \left\{ \sum_{i=1}^{N} \alpha_i(t)\omega_i : \alpha_i(t) \geq 0,\ \sum_{i=1}^{N} \alpha_i(t) = 1,\ N = 2^r \right\}. \end{aligned} \tag{2.2}$$

(3) The LPV plant is quadratically detectable and quadratically stabilizable.

With the above assumptions, the system state matrix $A(\theta)$ can be expressed as

$$A(\theta) = \sum_{i=1}^{N} \alpha_i(t)A_i \quad \text{with } \alpha_i \geq 0, \qquad \sum_{i=1}^{N} \alpha_i = 1. \tag{2.3}$$

Remark 2.1. It is assumed that the matrices B_u, C_y of the LPV plant are time invariant. When they are time varying, a simple way is to satisfy the requirement by filtering the control input and output through lowpass filters. These filters should have sufficiently large bandwidth. Then, the dependent parameters are shifted into the state matrix $A(\theta)$ in [3].

Definition 2.2 (quadratic stability [14]). Considering a LPV system, $\dot{x}(t) = A(\theta)x(t)$ is said to be quadratically stable if and only if there exists $P > 0$ such that

$$A^T(\theta)P + PA(\theta) < 0. \tag{2.4}$$

Remark 2.3. For polytopic LPV system, we have the equivalent conditions for (2.4) as

$$A_i^T P + PA_i < 0, \quad i = 1,\dots,N. \tag{2.5}$$

It should be noted that if LPV system is quadratically stable one, it is also exponentially stable.

3. Main results

In this section, a LTI output feedback controller is designed to achieve quadratic stability for LPV system where dependent parameters vary in a polytopic space.

We seek to design a controller ($A_K \in \mathbb{R}^{n_k \times n_k}$) of fixed order n_k as

$$\begin{aligned} \dot{x}_k &= A_k x_k + B_k y, \\ u &= C_k x_k + D_k y, \end{aligned} \tag{3.1}$$

where $x_K \in \mathbb{R}^{n_k}$ is the controller state. Substituting (3.1) into (2.1), the closed-loop state matrix A_{cl} has the following expression:

$$A_{\text{cl}}(\theta) = \begin{bmatrix} A(\theta) + B_u D_k C_y & B_u C_K \\ B_K C_y & A_K \end{bmatrix}. \tag{3.2}$$

First, the following definitions are made as

$$J = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix}, \qquad \overline{A}(\theta) = \begin{bmatrix} A(\theta) & 0 \\ 0 & 0 \end{bmatrix}, \qquad \overline{B}_u = \begin{bmatrix} 0 & B_u \\ I & 0 \end{bmatrix}, \qquad \overline{C}_y = \begin{bmatrix} 0 & I \\ C_y & 0 \end{bmatrix}, \tag{3.3}$$

which are totally dependent on the state-space matrices of the controller and the LPV plant. Then, the closed-loop relation is parameterized in terms of the controller realization as

$$A_{\text{cl}}(\theta) = \overline{A}(\theta) + \overline{B}_u J \overline{C}_y. \tag{3.4}$$

Theorem 3.1. *Suppose LPV system is given in (3.4), and then the following are equivalent conditions.*

(1) *The closed-loop state matrix $A_{cl}(\theta)$ is quadratically stable.*

(2) *There exist a symmetric positive definite matrix P and matrix J such that*

$$\overline{A}(\theta)P + P\overline{A}^T(\theta) + \overline{B}_u J \overline{C}_y P + P\overline{C}_y^T J^T \overline{B}_u^T < \delta I \tag{3.5}$$

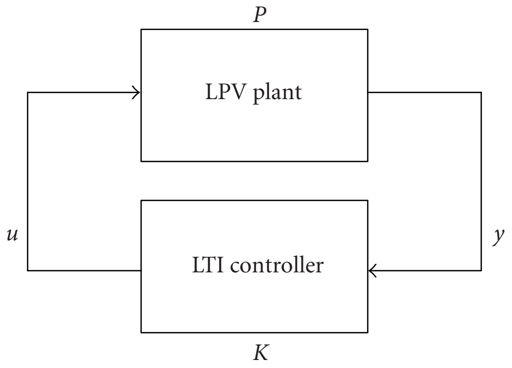


FIGURE 3.1. Relevant LPV control scheme.

or

$$\overline{A}_i P + P\overline{A}_i^T + \overline{B}_u J \overline{C}_y P + P\overline{C}_y^T J^T \overline{B}_u^T < \delta I, \tag{3.6}$$

$i = 1,\ldots,N$, *for δ being a negative scalar value.*

Proof. According to Definition 2.2, the claims (3.5) or (3.6) can be established easily. □

From (3.6), system matrix J of the controller (3.1) should be the same one for each vertex plant of LPV system (3.2): it is also a nonconvex constraint and difficult to be solved. In the following section, necessary conditions for the existence of a constant matrix J for (3.6) are presented, then a heuristic ILMI algorithm is presented to supply a solution of J for (3.6). The choosing of an appropriate initial value to ILMI is very important to converge quickly to a feasible solution. Here, a method of setting an initial value to ILMI algorithm is also proposed.

THEOREM 3.2. *Given an LPV plant (2.1), if there exists a fixed order LTI controller of order n_k that makes the closed-loop LPV system as Figure 3.1 quadratically stable, then there exist $n \times n$ symmetric positive definite matrices X, Y such that*

$$N_o^T\left(A^T(\theta)X + XA(\theta)\right)N_o < 0, \qquad N_c^T\left(YA(\theta) + A^T(\theta)Y\right)N_c < 0. \tag{3.7}$$

Using the polytopic characteristic of the LPV plant, (3.7) can be equivalent to

$$N_o^T\left(A_i^T X + XA_i\right)N_o < 0, \qquad N_c^T\left(YA_i + A_i^T Y\right)N_c < 0, \tag{3.8}$$

$$\begin{gathered}\begin{bmatrix} X & I \\ I & Y \end{bmatrix} \geq 0, \\ \mathit{rank}\begin{bmatrix} X & I \\ I & Y \end{bmatrix} \leq n + n_k, \end{gathered} \tag{3.9}$$

where N_o and N_c are full column rank matrices such that

$$\operatorname{Im} N_o = \ker C_y, \qquad \operatorname{Im} N_c = \ker B_u^T. \tag{3.10}$$

The proof of the theorem can be easily taken from earlier results [3, 15].

Theorem 3.2 tells us necessary conditions of the existence of a stabilizing output feedback LTI controller for the LPV plant (2.1). Meanwhile, it also provides an efficient method for setting an initial value of the common Lyapunov matrix P, which is used to construct a stabilizing output feedback LTI controller.

Remark 3.3. Now, let us overview some results of LPV controller design for LPV plant. Consider the LPV plant (2.1), since this plant is assumed to be quadratically stabilizable and quadratically detectable, (3.8)-(3.9) are sufficient and necessary conditions for the existence of such a full-order gain-scheduled LPV controller that quadratically stabilizes LPV plant (2.1). In contrast to gain-scheduled LPV controller design [3], here only an LTI controller is designed to quadratically stabilize the LPV plant and conditions (3.8)-(3.9) become not sufficient but necessary just as Theorem 3.2.

Note that the matrix inequality (3.6) is a bilinear matrix problem with the constraint that controller gain matrix should be constant, and it is a nonconvex optimization problem. Here, a heuristic approach of alternately solving convex optimization problems is proposed based on LMI formulation. We minimize δ, over P and J, subject to (3.6). This problem is a convex optimization problem in J and δ for fixed P, and is convex in P and δ for fixed J. It also should be noted that this approach is guaranteed to converge, but not necessarily to the global optimum of the problem. The assignment of a proper initial value to P is the key to enhance probability of converging to the global optimum. Here, conditions (3.8)-(3.9) supply necessary conditions for the existence of such an LTI controller of order n_k. Therefore, conditions (3.8)-(3.9) of Theorem 3.2 also give us an effective method of setting an initial value to P.

Therefore, the ILMI algorithm proceeds as shown in Algorithm 3.1.

If, after the procedure is alternated several times, solution J is still infeasible, there are two cases: one is that a feasible J may still exit, for this procedure does not necessarily guarantee to the solution J; the other is that the LPV plant may not be quadratically stabilizable by only an LTI controller.

4. Numerical example

In this section, two numerical examples are considered to illustrate the proposed method. All LMI-related computations are performed with LMI toolbox of Matlab [4].

Example 4.1. We consider the problem of controlling the yaw angles of a satellite system that appears in [4]. The satellite system consisting of two rigid bodies joined by a flexible link has the state-space representation as

$$\begin{bmatrix}\dot{\theta}_1\\ \dot{\theta}_2\\ \ddot{\theta}_1\\ \ddot{\theta}_2\end{bmatrix}=\begin{bmatrix}0&0&1&0\\0&0&0&1\\-k&k&-f&f\\k&-k&f&-f\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \dot{\theta}_1\\ \dot{\theta}_2\end{bmatrix}+\begin{bmatrix}0\\0\\0\\1\end{bmatrix}u, \qquad y=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \dot{\theta}_1\\ \dot{\theta}_2\end{bmatrix}, \tag{4.1}$$

```
Step 1.
    Set initial value $i = 0$, obtain    $P_i = \begin{bmatrix} X & X_2 \\ X_2^T & I \end{bmatrix}$
    subject to (3.8)-(3.9), where $X - Y^{-1} = X_2 X_2^T$.
    Let $\delta_i$ be an arbitrary large positive real number.
    $\delta_{\text{old}} = \delta_i$.
Step 2.
    Repeat {
    OP1: Solve eigenvalue problem, "minimize $\delta_{i1}$, over $J_i$ and $\delta_{i1}$, subject to (3.6);"
            $\delta_i = \delta_{i1\text{opt}}$, $J_i = J_{\text{opt}}$.
            If $\delta_i < 0$, exit. $J_i$ is an admissible solution.
    OP2: Solve eigenvalue problem, "minimize $\delta_{i2}$, over $P_i$ and $\delta_{i2}$, subject to (3.6)
     and $P_i > 0$";
            $P_{i+1} = P_{i2\text{opt}}$. $\delta_i = \delta_{i2\text{opt}}$.
            If $\delta_i < 0$, exit. $J_i$ is an admissible solution.
    If $\|\delta_i - \delta_{\text{old}}\| < \gamma$, a predetermined tolerance, exit.
    Else $\delta_{\text{old}} = \delta_i$.
    $i = i + 1$.
  }
```

Algorithm 3.1

where k and f are torque constant and viscous damping, which vary in the following uncertainty ranges: $k \in [0.09 \quad 0.4]$ and $f \in [0.0038 \quad 0.04]$. A state-feedback controller $u = Kx$ is designed to achieve quadratic stability for all possible parameter trajectories in the polytopic space. The pre-determined tolerance γ is set to $1.0e - 4$. The following two cases are considered.

(1) Setting an arbitrary matrix to the initial P such as identity matrix. After 12 iterations, δ_{12} converges to -0.0857, therefore solution K is found as

$$K = [1061463.3 \quad -1061463.3 \quad -258208.45 \quad -7338.2]. \tag{4.2}$$

(2) Setting an initial matrix to P proposed in this paper. In this case, a state feedback is considered to construct, then an initial matrix of P satisfying (3.8) is chosen as

$$P_0 = \begin{bmatrix} 961.4 & 518.14 & -118.4 & 278.06 \\ 518.1 & 930.3 & -247.3 & -167.8 \\ -118.4 & -247.3 & 95.46 & -55.25 \\ 278.06 & -167.8 & -55.25 & 972.54 \end{bmatrix}. \tag{4.3}$$

Using the initial matrix P_0, after only 1 iteration, δ_1 converges to -9395817.73. An admissible K is found as

$$K = [10541311.8 \quad -24814284.7 \quad -60435459.05 \quad -15945712.7]. \tag{4.4}$$

Therefore, the proposed method has a quicker convergence to a feasible solution than the method of setting an arbitrary matrix as the initial matrix P.

Example 4.2. A classical example of parameter-varying unstable plant that can be viewed as a mass-spring-damper system with time-varying spring stiffness is considered [16]. The state-space equation of this unstable LPV plant is as follows:

$$A(\theta) = \begin{bmatrix} 0 & 1 \\ -0.5-0.5\theta & -0.2 \end{bmatrix}, \qquad B_u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \\ C_y = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad D_{uy} = 0. \tag{4.5}$$

Here, the scope of time-varying parameter $\theta(t)$ is assumed in the polytope space $\Theta := Co\{-1, 1\}$. An LTI output feedback controller is designed to achieve quadratic stability for all possible parameter trajectories in the polytopic space. The predetermined tolerance γ is set to $1.0e-4$.

Just like Example 4.1, the following two cases are considered.

(1) Setting an arbitrary matrix to the initial P, such as identity matrix. After 5 iterations, δ_5 converges to 0.163, which is larger than zero. Therefore solution J is found infeasible.

(2) Setting an initial matrix to P proposed in this paper. In this case, a full-order output feedback controller is considered to construct, then an initial matrix of P satisfying (3.8)-(3.9) is as follows:

$$P_0 = \begin{bmatrix} 17.62 & -11.99 & 4.01 & -1.22 \\ -11.99 & 35.23 & -1.22 & 5.80 \\ 4.01 & -1.22 & 1 & 0 \\ -1.22 & 5.80 & 0 & 1 \end{bmatrix}. \tag{4.6}$$

Using the initial matrix P_0, after only 1 iteration, δ_1 converges to -3.998. An admissible J is solved as

$$J = 1.0e8 * \begin{bmatrix} -4.00 & -0.53 & -7.8e-7 \\ -0.53 & -5.56 & 2.94e-7 \\ 3.76e-8 & 6.0e-10 & 2.0e-7 \end{bmatrix}. \tag{4.7}$$

Therefore, an LTI output feedback controller to satisfy quadratic stability of closed-loop LPV system is constructed as

$$A_k = 1.0e8 * \begin{bmatrix} -4.00 & -0.53 \\ -0.53 & -5.56 \end{bmatrix}, \qquad B_K = \begin{bmatrix} -78 \\ 29.4 \end{bmatrix}, \\ C_K = \begin{bmatrix} 3.76 & 0.06 \end{bmatrix}, \qquad D_k = 20. \tag{4.8}$$

When the trajectory of dependent parameter is assumed as $\theta(t) = 0.63 + 0.1 \cdot e^{-t}$, the trajectory of the output of this plant can be drawn for the initial values $x(0) = [-0.25 \quad 0.15]^T$ as shown in Figure 4.1.

Comparing these two cases above, numerical examples demonstrate that the proposed method of setting the initial value to ILMI algorithm is more efficient than the method of setting an arbitrary matrix as the initial value.

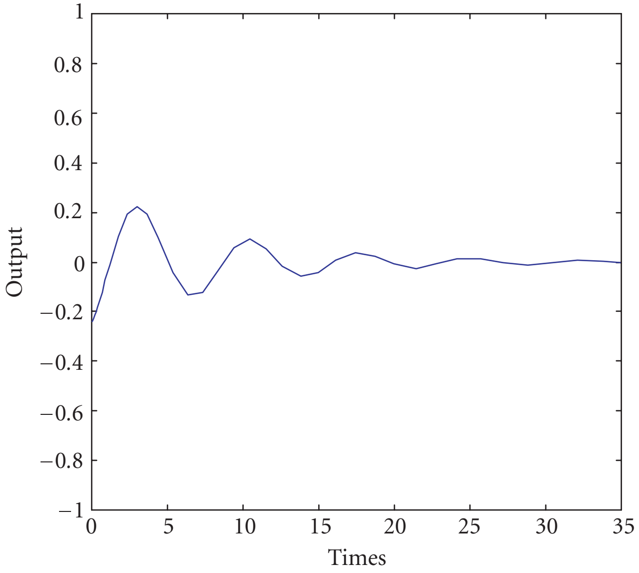


FIGURE 4.1. Trajectory of the output of this plant with initial values $x(0) = [-0.25 \quad 0.15]^T$.

5. Conclusions

In this paper, an LTI output feedback controller has been designed for LPV system to ensure that the closed-loop system achieves quadratic stability for all possible dependent parameters in a polytopic space. A heuristic iterative algorithm to solve such a controller has been presented in terms of LMI formulation. It also should be noted that the procedure is heuristic and the choice of initial value is important to ensure convergence to an acceptable solution. Finally, some numerical examples have been presented to illustrate the design method.

Acknowledgement

This work is supported by the Guangdong Natural Science Foundation (Grant no. 07006470).

References

[1] J. S. Shamma and M. Athans, "Analysis of gain scheduled control for nonlinear plants," *IEEE Transactions on Automatic Control*, vol. 35, no. 8, pp. 898–907, 1990.

[2] P. Apkarian and P. Gahinet, "A convex characterization of gain-scheduled $\mathcal{H}_\infty$ controllers," *IEEE Transactions on Automatic Control*, vol. 40, no. 5, pp. 853–864, 1995.

[3] P. Apkarian, P. Gahinet, and G. Becker, "Self-scheduled $\mathcal{H}_\infty$ control of linear parameter-varying systems: a design example," *Automatica*, vol. 31, no. 9, pp. 1251–1261, 1995.

[4] P. Gahinet, A. Nemirovski, A. J. Laub, and M. Chilali, *MATLAB LMI Control Toolbox*, The MathWorks, Natick, Mass, USA, 1995.

[5] D. J. Leith and W. E. Leithead, "Survey of gain-scheduling analysis and design," *International Journal of Control*, vol. 73, no. 11, pp. 1001–1025, 2000.

[6] W. Xie, “Quadratic L2 gain performance LPV system design by a LTI controller with ILMI algorithm,” *IEE Proceedings of Control Theory and Applications*, vol. 152, no. 2, pp. 125–128, 2005.

[7] W. J. Rugh and J. S. Shamma, “Research on gain scheduling,” *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

[8] G. Becker, A. Packard, D. Philbrick, and G. Balas, “Control of parametrically-dependent linear systems: a single quadratic Lyapunov approach,” in *Proceedings of the American Control Conference*, pp. 2795–2799, San Francisco, Calif, USA, June 1993.

[9] Y.-Y. Cao, J. Lam, and Y.-X. Sun, “Static output feedback stabilization: an ILMI approach,” *Automatica*, vol. 34, no. 12, pp. 1641–1645, 1998.

[10] L. El Ghaoui and V. Balakrishnan, “Synthesis of fixed-structure controllers via numerical optimization,” in *Proceedings of the 33rd IEEE Conference on Decision and Control (DC ’94)*, vol. 3, pp. 2678–2683, Lake Buena Vista, Fla, USA, December 1994.

[11] L. El Ghaoui, F. Oustry, and M. Aitrami, “A cone complementarity linearization algorithm for static output-feedback and related problems,” *IEEE Transactions on Automatic Control*, vol. 42, no. 8, pp. 1171–1176, 1997.

[12] K. M. Grigoriadis and R. E. Skelton, “Low-order control design for LMI problems using alternating projection methods,” *Automatica*, vol. 32, no. 8, pp. 1117–1125, 1996.

[13] T. Iwasaki and R. E. Skelton, “The XY-centring algorithm for the dual LMI problem: a new approach to fixed-order control design,” *International Journal of Control*, vol. 62, no. 6, pp. 1257–1272, 1995.

[14] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, vol. 15 of *SIAM Studies in Applied Mathematics*, SIAM, Philadelphia, Pa, USA, 1994.

[15] P. Gahinet and P. Apkarian, “A linear matrix inequality approach to $\mathcal{H}_\infty$ control,” *International Journal of Robust and Nonlinear Control*, vol. 4, no. 4, pp. 421–448, 1994.

[16] W. Xie and T. Eisaka, “Design of LPV control systems based on Youla parameterisation,” *IEE Proceedings of Control Theory and Applications*, vol. 151, no. 4, pp. 465–472, 2004.

Wei Xie: College of Automation Science and Technology, South China University of Technology,
Guangzhou 510641, China
Email address: weixie@scut.edu.cn

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 36261, 18 pages
doi:10.1155/2007/36261

Research Article

Modal Formulation of Segmented Euler-Bernoulli Beams

Rosemaira Dalcin Copetti, Julio C. R. Claeyssen, and Teresa Tsukazan

Received 5 September 2006; Revised 8 December 2006; Accepted 11 February 2007

Recommended by José Manoel Balthazar

We consider the obtention of modes and frequencies of segmented Euler-Bernoulli beams with internal damping and external viscous damping at the discontinuities of the sections. This is done by following a Newtonian approach in terms of a fundamental response of stationary beams subject to both types of damping. The use of a basis generated by the fundamental solution of a differential equation of fourth-order allows to formulate the eigenvalue problem and to write the modes shapes in a compact manner. For this, we consider a block matrix that carries the boundary conditions and intermediate conditions at the beams and values of the fundamental matrix at the ends and intermediate points of the beam. For each segment, the elements of the basis have the same shape since they are chosen as a convenient translation of the elements of the basis for the first segment. Our method avoids the use of the first-order state formulation also to rely on the Euler basis of a differential equation of fourth-order and it allows to envision how conditions will influence a chosen basis.

1. Introduction

The methodology introduced by Tsukazan [1] in terms of a fundamental response [2, 3] is applied here to a triple-span Euler-Bernoulli beam with internal damping of the type Kelvin-Voight and viscous external damping at the discontinuities of the sections.

In the literature, the study of free vibrations of beams of the type Euler-Bernoulli have been sufficiently studied [4–11]. However, the effects of the nonproportional damping has been little studied in terms of modal analysis. Friswell and Lees [12] considered the method of separation of variables for obtaining the eigenvalues of a double-span pinned-pinned nonhomogeneous damped beam without intermediate devices. Chang et al. [13]

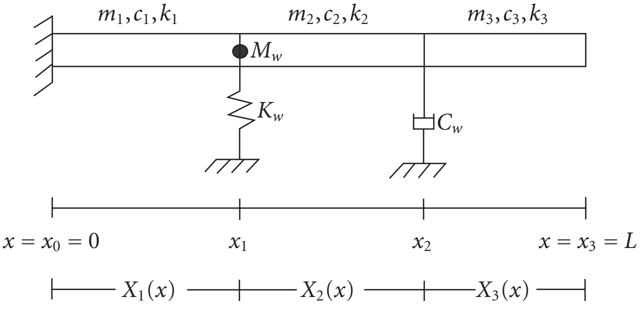


Figure 2.1. A triple-span discontinuous cantilever beam.

uses the Laplace transform for obtaining the natural frequencies of a pinned-pinned uniform Euler-Bernoulli beam, by considering masses, springs, and viscous dampers located in the middle of the beam. Sorrentino et al. [14] obtain the frequencies of the beam by using the state space formulation with a first-order transfer matrix. The obtention of the modes was accomplished by using the Euler basis in connection with fourth-order spatial differential equations, the Laplace transform and with the state-space methodology. Simulations were performed for double-span and four-span beams with several types of damping: internal, external, nonproportional, viscous damping.

Here, we consider the original Newtonian approach by keeping the formulation of a second-order system, that includes damping and stiffness, in each segment of the beam. The coefficients for the displacement boundary conditions and intermediate continuity conditions at discontinuity points of the beam are casted in a convenient block matrix that we refer to as being the coefficient matrix. The values that the elements of the basis at each segment take at the ends of the beam and intermediate discontinuity points give rise to another block matrix called the basis matrix. The introduction of these block matrices allows to formulate the eigenvalue problem in a compact matrix form. By choosing a basis that is generated by a fundamental solution of a fourth-order differential equation, the basis matrix becomes sparse. This approach can also be employed with double- or four-span beams subject to classical and nonclassical boundary conditions. In a forthcoming work, we will discuss multispan beams subject to a elastic coupling and discuss a reduction in the computation of the coefficients of a mode in each segment.

2. Statement of problem

We consider an Euler-Bernoulli beam of length L with two intermediate devices and two discontinuous cross sections, as in Figure 2.1. A flexural movement is represented in the beam by $v_j(t,x)$ in the jth segment $[x_{j-1}, x_j]$, $j = 1:3$ with $0 = x_0 \le x_1 \le x_2 \le x_3 = L$.

Here, M_w denotes value of the attached mass, C_w attached damping coefficient, K_w the attached stiffness.

In each segment of the beam, we have the governing equations [6, 15]

$$M_j \frac{\partial^2 v_j(t,x)}{\partial t^2} + C_j \frac{\partial v_j(t,x)}{\partial t} + K_j v_j(t,x) = 0, \quad x_{j-1} < x < x_j,\ j = 1:3, \tag{2.1}$$

where

$$M_j = m_j = \rho_j A_j,$$
$$K_j = \frac{\partial^2}{\partial x^2}\left[k_j(x)\frac{\partial^2}{\partial x^2}\right]. \tag{2.2}$$

The damping coefficient can be considered to be of the form

$$C_j = c_{0j}(x) + \frac{\partial^2}{\partial x^2}\left[c_{4j}(x)\frac{\partial^2}{\partial x^2}\right] \tag{2.3}$$

which includes the case of external viscous damping and internal Kelvin-Voigt damping. In the above, we have the following usual parameter description:

(i) ρ_j denotes density,
(ii) A_j denotes cross-sectional area,
(iii) c_{ij} denotes damping coefficients,
(iv) k_j denotes stiffness coefficients.

In what follows, we will consider the particular case of beams with uniform sections. Then the coefficients in the operators C_j, K_j become constants, that is,

$$K_j = k_j\frac{\partial^4}{\partial x^4} = E_jI_j\frac{\partial^4}{\partial x^4}, \qquad C_j = c_{0j} + c_{4j}\frac{\partial^4}{\partial x^4}, \qquad M_j = m_j, \tag{2.4}$$

where E_j denotes Young's modulus of elasticity, I_j denotes the area moment of inertia.

3. Modal analysis

Free vibrations whose spatial distribution amplitude in each segment is $X_j(x)$,

$$v_j = e^{\lambda t}X_j(x), \quad x \in [x_{j-1}, x_j],\ j = 1:3, \tag{3.1}$$

can be found by substituting them into the above system. It turns out the spatial modal differential equation

$$X_j^{(iv)}(x) - a_j^2(\lambda)\rho_j A_j X_j(x) = 0, \quad x \in [x_{j-1}, x_j],\ j = 1:3, \tag{3.2}$$

for each segment of the beam. Here,

$$a_j^2(\lambda) = -(\alpha_j + \lambda\beta_j)\lambda \tag{3.3}$$

with

$$\alpha_j = \frac{c_{0j}}{\rho_j A_j(E_jI_j + \lambda c_{4j})}, \quad \beta_j = \frac{1}{E_jI_j + \lambda c_{4j}}, \quad j = 1:3. \tag{3.4}$$

The solution for each segment (3.2) can be conveniently written as

$$X_j(x) = d_{1j}\phi_{1j}(x) + d_{2j}\phi_{2j}(x) + +d_{3j}\phi_{3j}(x) + d_{4j}\phi_{4j}(x) = \Psi_j(x)\mathbf{d_j}, \quad j = 1:3, \tag{3.5}$$

where

$$\Psi_j = \Psi_j(x,\lambda) = [\phi_{1,j}(x), \phi_{2,j}(x), \phi_{3,j}(x), \phi_{4,j}(x)] \tag{3.6}$$

is a solution basis of (3.2) in the segment $[x_{j-1}, x_j]$, $j = 1:3$, and $\mathbf{d_j}$ is the column vector with components d_{1j}, d_{2j}, d_{3j}, d_{4j}. Here we have emphasized that the solution matrix basis Ψ_j depend upon the parameter λ corresponding to a free vibration.

Generic boundary conditions of classical or nonclassical nature can be written as

$$\begin{aligned}
A_{11}X_1(0) + B_{11}X_1'(0) + C_{11}X_1''(0) + D_{11}X_1'''(0) &= 0,\\
A_{12}X_1(0) + B_{12}X_1'(0) + C_{12}X_1''(0) + D_{12}X_1'''(0) &= 0,\\
A_{21}X_3(L) + B_{21}X_3'(L) + C_{21}X_3''(L) + D_{21}X_3'''(L) &= 0,\\
A_{22}X_3(L) + B_{22}X_3'(L) + C_{22}X_3''(L) + D_{22}X_3'''(L) &= 0.
\end{aligned} \tag{3.7}$$

The continuity conditions for the displacement, the inertia moment, the bending moment, and the shear force at the discontinuity point x_j, $j = 1:2$ of the transversal section, including an intermediate device, can be written in general as follows:

$$\begin{aligned}
&E_{11}^{(j)}X_j(x_j) + F_{11}^{(j)}X_j'(x_j) + G_{11}^{(j)}X_j''(x_j) + H_{11}^{(j)}X_j'''(x_j)\\
&\qquad = E_{12}^{(j)}X_{j+1}(x_j) + F_{12}^{(j)}X_{j+1}'(x_j) + G_{12}^{(j)}X_{j+1}''(x_j) + H_{12}^{(j)}X_{j+1}'''(x_j),\\
&E_{21}^{(j)}X_j(x_j) + F_{21}^{(j)}X_j'(x_j) + G_{21}^{(j)}X_j''(x_j) + H_{21}^{(j)}X_j'''(x_j)\\
&\qquad = E_{22}^{(j)}X_{j+1}(x_j) + F_{22}^{(j)}X_{j+1}'(x_j) + G_{22}^{(j)}X_{j+1}''(x_j) + H_{22}^{(j)}X_{j+1}'''(x_j),\\
&E_{31}^{(j)}X_j(x_j) + F_{31}^{(j)}X_j'(x_j) + G_{31}^{(j)}X_1''(x_j) + H_{31}^{(j)}X_1'''(x_j)\\
&\qquad = E_{32}^{(j)}X_{j+1}(x_j) + F_{32}^{(j)}X_{j+1}'(x_j) + G_{32}^{(j)}X_{j+1}''(x_j) + H_{32}^{(j)}X_{j+1}'''(x_j),\\
&E_{41}^{(j)}X_j(x_j) + F_{41}^{(j)}X_j'(x_j) + G_{41}^{(j)}X_j''(x_j) + H_{41}^{(j)}X_j'''(x_j)\\
&\qquad = E_{42}^{(j)}X_{j+1}(x_j) + F_{42}^{(j)}X_{j+1}'(x_j) + G_{42}^{(j)}X_{j+1}''(x_j) + H_{42}^{(j)}X_{j+1}'''(x_j) + F_j, \quad j = 1:2,
\end{aligned} \tag{3.8}$$

where F_j denotes the force exerted by an external device.

Figure 2.1 shows a cantilever beam with intermediate continuity conditions at the points $x = x_1$ and $x = x_2$ and subject to a concentrated mass, spring, and a dashpot. The boundary conditions at $x = x_0 = 0$ and $x = x_3 = L$ are

$$X_1(0) = X_1'(0) = 0, \qquad X_3''(L) = X_3'''(L) = 0. \tag{3.9}$$

At the intermediate point $x = x_1$, we have

$$\begin{aligned} X_1(x_1) &= X_2(x_1), \\ X_1'(x_1) &= X_2'(x_1), \\ k_2^{-1}k_1X_1''(x_1) &= X_2''(x_1), \\ -k_2^{-1}(M_w\lambda^2 + K_w)X_1(x_1) + k_2^{-1}k_1X_1'''(x_1) &= X_2'''(x_1). \end{aligned} \tag{3.10}$$

Similarly, at the point $x = x_2$, we have

$$\begin{aligned} X_2(x_2) &= X_3(x_2), \\ X_2'(x_2) &= X_3'(x_2), \\ k_3^{-1}k_2X_2''(x_2) &= X_3''(x_2), \\ -k_3^{-1}(C_w\lambda)X_2(x_2) + k_3^{-1}k_2X_2'''(x_2) &= X_3'''(x_2). \end{aligned} \tag{3.11}$$

The substitution of (3.5) into (3.7) and (3.8), the boundary and continuity conditions leads to a linear algebraic system

$$\mathcal{U}(\lambda)\mathbf{d} = \mathbf{0}, \tag{3.12}$$

for the vector **d** of order 12 × 1,

$$\mathbf{d} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \end{bmatrix}, \qquad \mathbf{d_j} = \begin{bmatrix} d_{1j} \\ d_{2j} \\ d_{3j} \\ d_{4j} \end{bmatrix}, \quad j = 1:2. \tag{3.13}$$

Here, the matrix $\mathcal{U}$ is of order 12 × 12 and it has the form

$$\mathcal{U} = \mathcal{B}\Phi, \tag{3.14}$$

where $\mathcal{B}$ is a matrix of order 12 × 24 formed with the coefficients associated to the boundary and continuity conditions and Φ is a matrix of order 24 × 12 whose components are values of the solution basis at the ends and the conditions at the discontinuity. A detailed description of these block matrices is given in Section 4. Then nonzero solutions of (3.12) are obtained for frequency values λ real or complex that satisfy the characteristic equation

$$\det(\mathcal{U}) = 0. \tag{3.15}$$

In classical conservative mechanical vibration theory, modes are essential for performing a decoupling of the system. However, any real structure with or without intermediate devices is dissipative. This implies the existence of complex modes that not necessarily decouple a damped system [16]. On the other hand, any pair of complex conjugate modes represent a free vibration in which distributed coordinates oscillate and share the same decay rate and frequency but are not synchronous. This later is because it introduced a phase when writing the mode or amplitude was in polar form [15].

4. Block matrix formulation

A detailed description of the matrix $\mathcal{U}$ in terms of the boundary and basis block matrices is given in what follows for a triple-span beam subject to generic conditions. The matrix corresponding to the boundary values can be written as follows:

$$\mathcal{B}_0=\begin{bmatrix}A_{11} & B_{11} & C_{11} & D_{11}\\ A_{12} & B_{12} & C_{12} & D_{12}\end{bmatrix},\qquad \mathcal{B}_L=\begin{bmatrix}A_{21} & B_{21} & C_{21} & D_{21}\\ A_{22} & B_{22} & C_{22} & D_{22}\end{bmatrix}. \tag{4.1}$$

The matrix coefficients corresponding to the continuity conditions at $x=x_j$, $j=1:2$, can be described in terms of the matrices

$$\mathcal{B}_{1j}=\begin{bmatrix}E_{11}^{(j)} & F_{11}^{(j)} & G_{11}^{(j)} & H_{11}^{(j)}\\ E_{21}^{(j)} & F_{21}^{(j)} & G_{21}^{(j)} & H_{21}^{(j)}\\ E_{31}^{(j)} & F_{31}^{(j)} & G_{31}^{(j)} & H_{31}^{(j)}\\ E_{41}^{(j)} & F_{41}^{(j)} & G_{41}^{(j)} & H_{41}^{(j)}\end{bmatrix},\qquad \mathcal{B}_{2j}=\begin{bmatrix}E_{12}^{(j)} & F_{12}^{(j)} & G_{12}^{(j)} & H_{12}^{(j)}\\ E_{22}^{(j)} & F_{22}^{(j)} & G_{22}^{(j)} & H_{22}^{(j)}\\ E_{32}^{(j)} & F_{32}^{(j)} & G_{32}^{(j)} & H_{32}^{(j)}\\ E_{42}^{(j)} & F_{42}^{(j)} & G_{42}^{(j)} & H_{42}^{(j)}\end{bmatrix}. \tag{4.2}$$

The values of the basis solutions at the ends of the beam x_0, x_3, and at each discontinuity point x_k, $k=1,2$, can be written in terms of the Wronskian matrices in each segment

$$\mathbf{\Phi_j}(x)=\begin{bmatrix}\phi_{1j}(x) & \phi_{2j}(x) & \phi_{3j}(x) & \phi_{4j}(x)\\ \phi'_{1j}(x) & \phi'_{2j}(x) & \phi'_{3j}(x) & \phi'_{4j}(x)\\ \phi''_{1j}(x) & \phi''_{2j}(x) & \phi''_{3j}(x) & \phi''_{4j}(x)\\ \phi'''_{1j}(x) & \phi'''_{2j}(x) & \phi'''_{3j}(x) & \phi'''_{4j}(x)\end{bmatrix},\quad j=1:3. \tag{4.3}$$

For a triple-span beam, we will have the block matrices

$$\mathcal{B}=\begin{bmatrix}\mathcal{B}_0 & 0 & 0 & 0 & 0 & 0\\ 0 & \mathcal{B}_{11} & -\mathcal{B}_{21} & 0 & 0 & 0\\ 0 & 0 & 0 & \mathcal{B}_{12} & -\mathcal{B}_{22} & 0\\ 0 & 0 & 0 & 0 & 0 & \mathcal{B}_3\end{bmatrix}, \tag{4.4}$$

$$\mathbf{\Phi}=\begin{bmatrix}\mathbf{\Phi_1}(0) & 0 & 0\\ \mathbf{\Phi_1}(x_1) & 0 & 0\\ 0 & \mathbf{\Phi_2}(x_1) & 0\\ 0 & \mathbf{\Phi_2}(x_2) & 0\\ 0 & 0 & \mathbf{\Phi_3}(x_2)\\ 0 & 0 & \mathbf{\Phi_3}(x_3)\end{bmatrix}. \tag{4.5}$$

In the above, 0 denotes null matrices with appropriate dimensions, that is, 2×4 or 4×4.

4.1. A cantilever triple-span beam subject to damping. For the triple-span cantilever beam of Figure 2.1, the corresponding blocks for the coefficients of the boundary conditions are

$$\mathcal{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \qquad \mathcal{B}_L = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.6}$$

while the blocks for the continuity conditions at the intermediate discontinuous sections are

$$\mathcal{B}_{11} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & k_2^{-1}k_1 & 0 \\ -k_2^{-1}(M_w\lambda^2 + K_w) & 0 & 0 & k_2^{-1}k_1 \end{bmatrix}, \qquad \mathcal{B}_{21} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathcal{B}_{12} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & k_3^{-1}k_2 & 0 \\ -k_3^{-1}(C_w\lambda) & 0 & 0 & k_3^{-1}k_2 \end{bmatrix}, \qquad \mathcal{B}_{22} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.7}$$

5. The fundamental basis

The classical or spectral Euler basis of the fourth-order equation,

$$X^{(iv)}(x) - \varepsilon^4 X(x) = 0, \tag{5.1}$$

is constructed by using the roots $\pm\varepsilon$, $\pm i\varepsilon$ of the characteristic polynomial $s^4 - \varepsilon^4 = 0$, that is,

$$\Psi = \left[\sin(\varepsilon x), \cos(\varepsilon x), \sinh(\varepsilon x), \cosh(\varepsilon x)\right]. \tag{5.2}$$

However, among all possible bases that we can choose, it would be convenient to choose the basis that makes (4.5) as sparse as possible. In this work, this is accomplished by choosing in each segment a *fundamental* basis that is a translation of a fixed basis that is generated by an initial-value solution in the first segment. This later solution can be found in the work of Timoshenko et al. [17] literature without the systematic treatment considered in [2, 3, 18]. We will consider the basis for the first segment that is constituted by the solution $h(x)$ of the initial value problem

$$\begin{gathered} h^{(iv)}(x) - \varepsilon^4 h(x) = 0, \\ h(0) = 0, \qquad h'(0) = 0, \qquad h''(0) = 0, \qquad h'''(0) = 1, \end{gathered} \tag{5.3}$$

and its first three derivatives $h'(x)$, $h''(x)$, $h'''(x)$. With respect to the spectral Euler basis, the fundamental solution $h(x)$ has the following representation:

$$h(x) = \frac{\sinh(\varepsilon x) - \sin(\varepsilon x)}{2\varepsilon^3}. \tag{5.4}$$

By defining

$$\begin{gathered}\phi_{jk}^{(i-1)}(x) = h^{(j+i-2)}(x - x_{k-1}, \varepsilon_k), \quad i, j = 1:4,\ k = 1:3, \\ \varepsilon_k^4 = a_k^2(\lambda)\rho_k A_k,\end{gathered} \tag{5.5}$$

where we have emphasized the dependence of the solution of (5.3) upon the parameter ε in each segment of the beam, it turns out

$$\mathbf{\Phi}_j(x_{j-1}) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad j = 1:3. \tag{5.6}$$

By taking into account the initial values of $h(x, \epsilon)$, the matrix (4.5) becomes more sparse and it is given by

$$\Phi = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi_{11}(x_1) & \phi_{21}(x_1) & \phi_{31}(x_1) & \phi_{41}(x_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi'_{11}(x_1) & \phi'_{21}(x_1) & \phi'_{31}(x_1) & \phi'_{41}(x_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi''_{11}(x_1) & \phi''_{21}(x_1) & \phi''_{31}(x_1) & \phi''_{41}(x_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi'''_{11}(x_1) & \phi'''_{21}(x_1) & \phi'''_{31}(x_1) & \phi'''_{41}(x_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi_{12}(x_2) & \phi_{22}(x_2) & \phi_{32}(x_2) & \phi_{42}(x_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi'_{12}(x_2) & \phi'_{22}(x_2) & \phi'_{32}(x_2) & \phi'_{42}(x_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi''_{12}(x_2) & \phi''_{22}(x_2) & \phi''_{32}(x_2) & \phi''_{42}(x_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi'''_{12}(x_2) & \phi'''_{22}(x_2) & \phi'''_{32}(x_2) & \phi'''_{42}(x_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{13}(L) & \phi_{23}(L) & \phi_{33}(L) & \phi_{43}(L) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi'_{13}(L) & \phi'_{23}(L) & \phi'_{33}(L) & \phi'_{43}(L) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi''_{13}(L) & \phi''_{23}(L) & \phi''_{33}(L) & \phi''_{43}(L) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi'''_{13}(L) & \phi'''_{23}(L) & \phi'''_{33}(L) & \phi'''_{43}(L)
\end{bmatrix}. \tag{5.7}$$

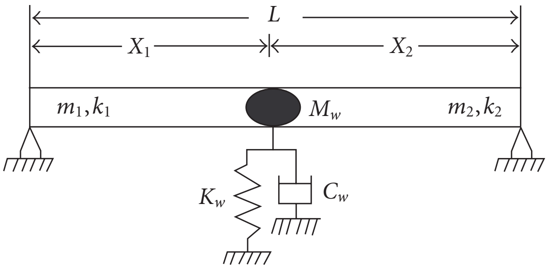


Figure 6.1. A double-span discontinuous cantilever beam.

The fundamental response $h(x,\varepsilon)$, has the same shape for each segment, but depends on different values for the involved physical parameters.

6. Numerical examples

6.1. Double-span beam. We first consider the case of a pinned-pinned double-span beam of length L as Figure 6.1 that was studied in Sorrentino et al. [14] and Chang et al. [13].

The spatial modal differential equation to double-span beam can be expressed in the form

$$X_j^{(iv)}(x) - a_j^2(\lambda)\rho_j A_j X_j(x) = 0, \quad x \in [x_{j-1}, x_j],\ j = 1:2, \tag{6.1}$$

for each segment of the beam, where a_j, $j = 1:2$ are given in (3.3).

The boundary conditions to beam above at $x = x_0 = 0$ and $x = x_2 = L$ are

$$X_1(0) = X_1''(0) = 0, \qquad X_2(L) = X_2''(L) = 0. \tag{6.2}$$

We have the intermediate continuity conditions at the point $x = x_1$,

$$\begin{gathered} X_1(x_1) = X_2(x_1), \\ X_1'(x_1) = X_2'(x_1), \\ k_2^{-1}k_1X_1''(x_1) = X_2''(x_1), \\ -k_2^{-1}(M_w\lambda^2 + C_w\lambda + K_w)X_1(x_1) + k_2^{-1}k_1X_1'''(x_1) = X_2'''(x_1). \end{gathered} \tag{6.3}$$

For a double-span beam, the blocks that correspond to the boundary conditions are

$$\mathcal{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \mathcal{B}_L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{6.4}$$

At the intermediate points, where continuity conditions are to be held, we have

$$\mathcal{B}_{11}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & k_2^{-1}k_1 & 0\\ -k_2^{-1}\left(M_w\lambda^2+C_w\lambda+K_w\right) & 0 & 0 & k_2^{-1}k_1\end{bmatrix},\qquad \mathcal{B}_{21}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}. \tag{6.5}$$

Thus, the coefficient block matrix of the given double-span beam is

$$\mathcal{B}=\begin{bmatrix}\mathcal{B}_0 & 0 & 0\\ 0 & \mathcal{B}_{11} & -\mathcal{B}_{21}\\ 0 & 0 & \mathcal{B}_L\end{bmatrix} \tag{6.6}$$

or expanded

$$\mathcal{B}=\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \dfrac{k_1}{k_2} & 0 & 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 0 & \Gamma & 0 & 0 & \dfrac{k_1}{k_2} & 0 & 0 & 0 & -1\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\end{bmatrix}, \tag{6.7}$$

where $\Gamma=-k_2(M_w\lambda^2+C_w\lambda+K_w)$.

For constructing the basis matrix, that carries the values of the generic solution basis at the ends of the beam and at the discontinuity points of a double-span beam, we consider

$$\phi_{jk}^{(i-1)}(x)=h^{(j+i-2)}\left(x-x_{k-1},\varepsilon_k\right),\quad i,j=1:4,\ k=1:2, \tag{6.8}$$

where $h(x)$ is the solution of (5.3). Then, the basis matrix is given by

$$\Phi=\begin{bmatrix}\mathbf{\Phi_1}(0) & 0\\ \mathbf{\Phi_1}(x_1) & 0\\ 0 & \mathbf{\Phi_2}(x_1)\\ 0 & \mathbf{\Phi_2}(L)\end{bmatrix} \tag{6.9}$$

Table 6.1. Parameter values of a double-span beam.

Parameter	Numeric value	Unit
$m_1 = m_2$	1.6363×10^4	kg/m
$k_1 = k_2$	1.6669×10^{11}	Nm^2
L	15.24	m

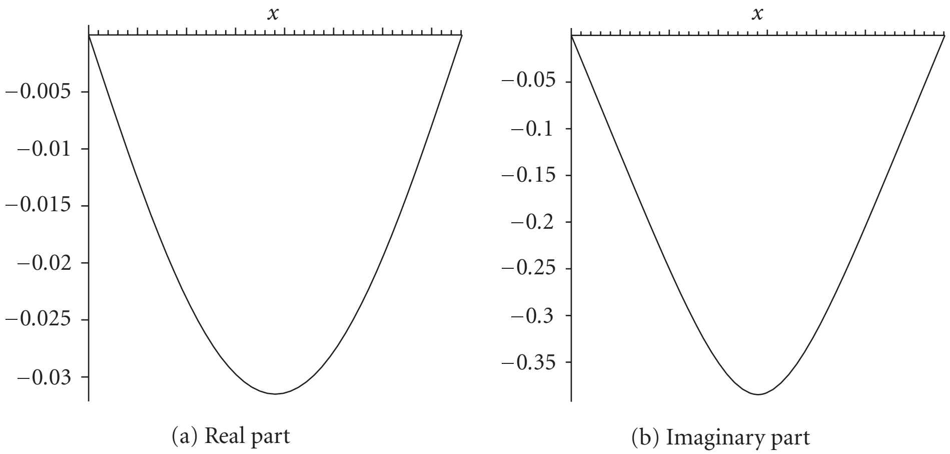
Table 6.2. Eigenvalues (rad/s) to double-span beam.

Mode (n)	Proposed method	[14]
1	−11.25426117 ± 135.0795544 I	−11.30627 ± 135.1799 I
2	.5512552857e-7 ±542.5166750 I	0 ± 542.5144 I
3	−8.442911066 ± 1128.708193 I	−8.482803 ± 1128.716 I

or expanded

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_{11}(x_1) & \phi_{21}(x_1) & \phi_{31}(x_1) & \phi_{41}(x_1) & 0 & 0 & 0 & 0 \\ \phi'_{11}(x_1) & \phi'_{21}(x_1) & \phi'_{31}(x_1) & \phi'_{41}(x_1) & 0 & 0 & 0 & 0 \\ \phi''_{11}(x_1) & \phi''_{21}(x_1) & \phi''_{31}(x_1) & \phi''_{41}(x_1) & 0 & 0 & 0 & 0 \\ \phi'''_{11}(x_1) & \phi'''_{21}(x_1) & \phi'''_{31}(x_1) & \phi'''_{41}(x_1) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_{12}(L) & \phi_{22}(L) & \phi_{32}(L) & \phi_{42}(L) \\ 0 & 0 & 0 & 0 & \phi'_{12}(L) & \phi'_{22}(L) & \phi'_{32}(L) & \phi'_{42}(L) \\ 0 & 0 & 0 & 0 & \phi''_{12}(L) & \phi''_{22}(L) & \phi''_{32}(L) & \phi''_{42}(L) \\ 0 & 0 & 0 & 0 & \phi'''_{12}(L) & \phi'''_{22}(L) & \phi'''_{32}(L) & \phi'''_{42}(L) \end{bmatrix}. \tag{6.10}$$

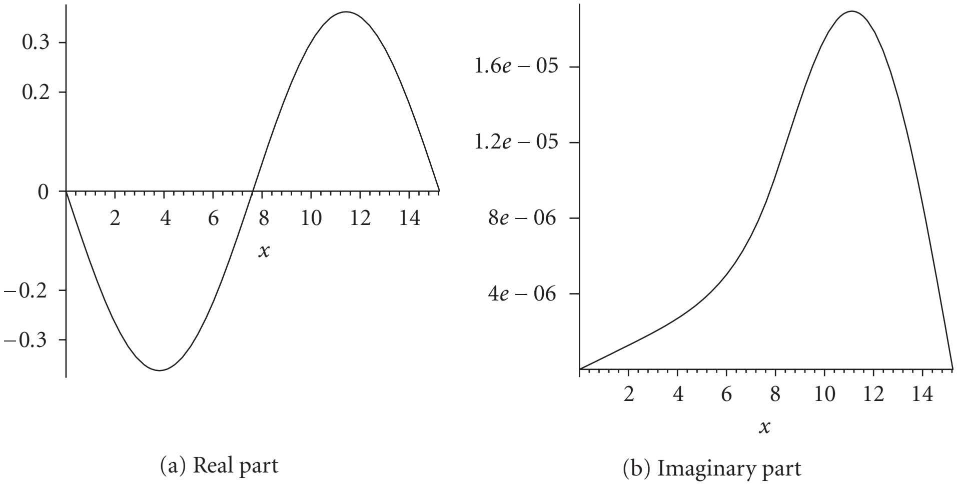
Numerical simulations with the proposed method are presented by using the data in Table 6.1. The parameter values at the discontinuity point $x = x_1 = (L/2)$ of beam used are $M_w = 0.1mL$, $K_w = 0.1mLw_1^2$, and $C_w = 0.1mLw_1$ where $m = m_1 = m_2$ and w_1 is the first natural frequency of the beam without added mass and spring [13]. In Table 6.2, the first three eigenvalues of the beam were obtained by solving the characteristic equation (3.15) with an approximation of $h(x)$ and compared with the ones obtained in [14]. We observed a good agreement among the two methods. In Figures 6.2, 6.3, and 6.4 are showed the modes shapes corresponding to the first three eigenvalues of the beam, where (a) indicates the real part of the mode and (b) the imaginary part of the mode.



(a) Real part

(b) Imaginary part

Figure 6.2. First mode to double-span beam.



(a) Real part

(b) Imaginary part

Figure 6.3. Second mode to double-span beam.

6.2. Triple-span beam. We now consider the triple-span beam given in Figure 2.1. First, we assume that the beam is uniform with parameter values given in Table 6.3. The viscous damping at the point of discontinuity $x = x_2$ is given by $Cw = 0.1mLw_1$, where $x_1 = 4m$, $x_2 = 10m$, and w_1 is the first natural frequency of the beam without added mass and spring [13].

In Table 6.4, we have the values of the first three eigenvalues of the beam and in Figures 6.5, 6.6, and 6.7 the correspondent modes shapes, where (a) it indicates the real part of the mode and (b) the imaginary part of the mode.

Table 6.3. System parameters to beam uniform triple-span.

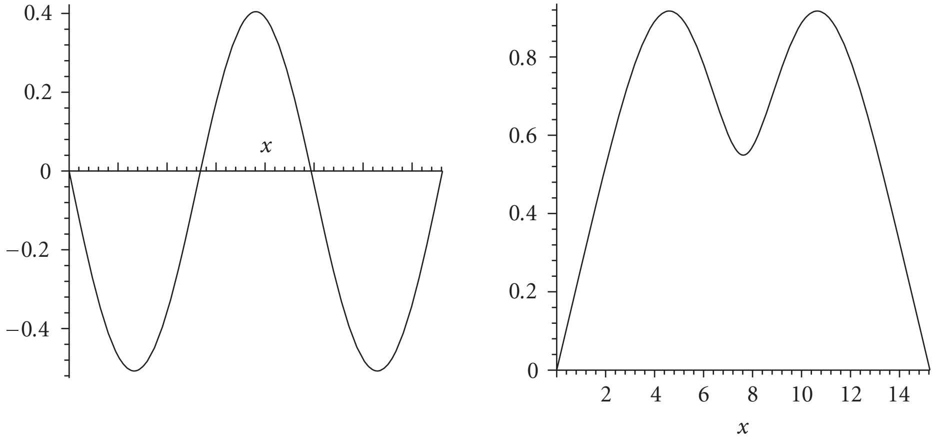
Parameter	Numeric value	Unit
$m_1 = m_2 = m_3 = m$	1.6363×10^4	kg/m
$k_1 = k_2 = k_3$	1.6669×10^{11}	Nm^2
L	15.24	m

Table 6.4. Eigenvalues of a uniform triple-span beam.

Mode (n)	Eigenvalues
1	$-1.355547843 \pm 48.31860604$ I
2	$-12.43829644 \pm 302.5538405$ I
3	$-8.226875796 \pm 847.5582997$ I

Table 6.5. System parameters to triple-span beam.

	Segment first	Segment second	Segment third	Unit
Mass	1.6363×10^4	$0.8 \times m_1$	$0.8 \times m_1$	kg/m
Stiffness	1.6669×10^{11}	$1.4 \times k_1$	$0.6 \times k_1$	Nm^2
Damping	5×10^{-1}	$0.5 \times c_1$	$11.7 \times c_1$	Ns/m^2
Length (L)	4	6	5.24	m



(a) Real part

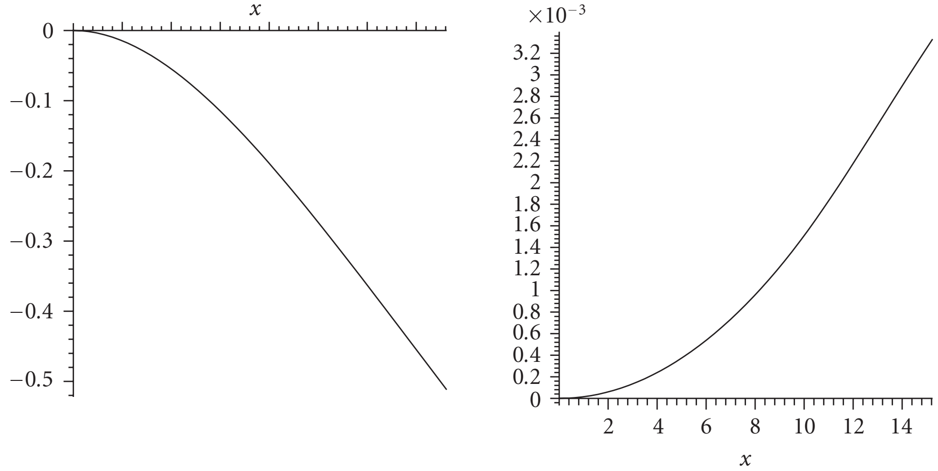
(b) Imaginary part

Figure 6.4. Third mode to double-span beam.

For the second case, we consider that the cantilever beam in Figure 2.1 is nonuniform. Its parameters values are given in Table 6.5. The first three eigenvalues of the beam are listed in Table 6.6.

Table 6.6. Eigenvalues (rad/s) to triple-span beam.

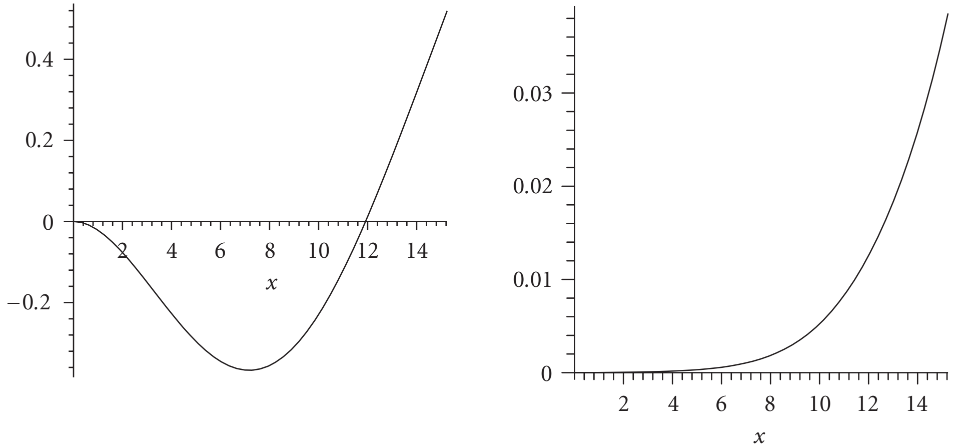
Mode (n)	Eigenvalues
1	$-.4314672830 \pm 49.72926784\,\mathrm{I}$
2	$-8.906552965 \pm 278.6470011\,\mathrm{I}$
3	$-16.73690547 \pm 797.5457311\,\mathrm{I}$



(a) Real part

(b) Imaginary part

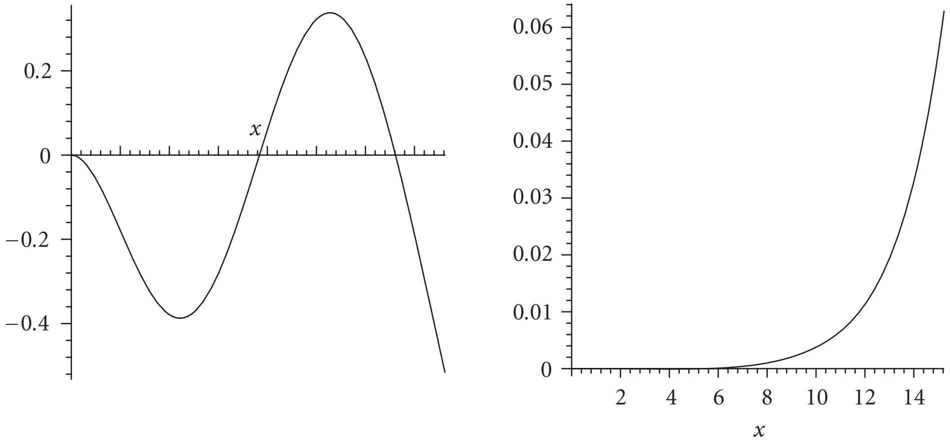
Figure 6.5. First mode of a uniform triple-span beam.



(a) Real part

(b) Imaginary part

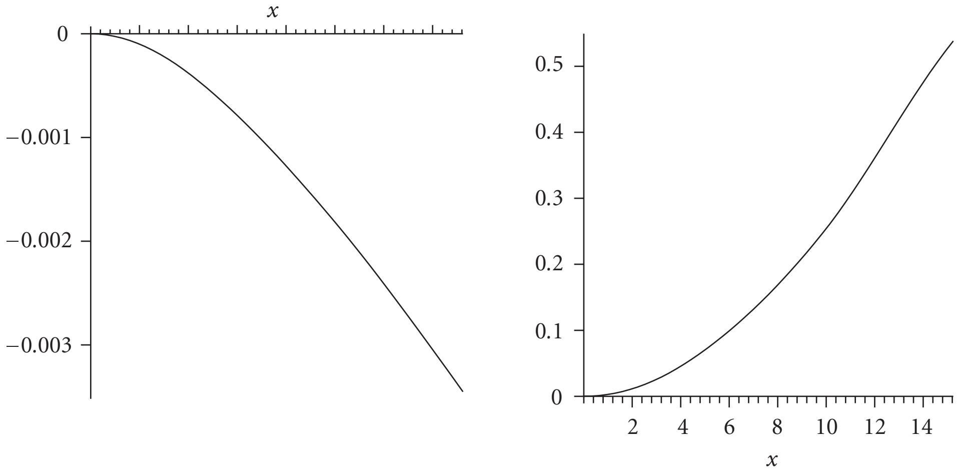
Figure 6.6. Second mode of a uniform triple-span beam.



(a) Real part

(b) Imaginary part

Figure 6.7. Third mode of a uniform triple-span beam.



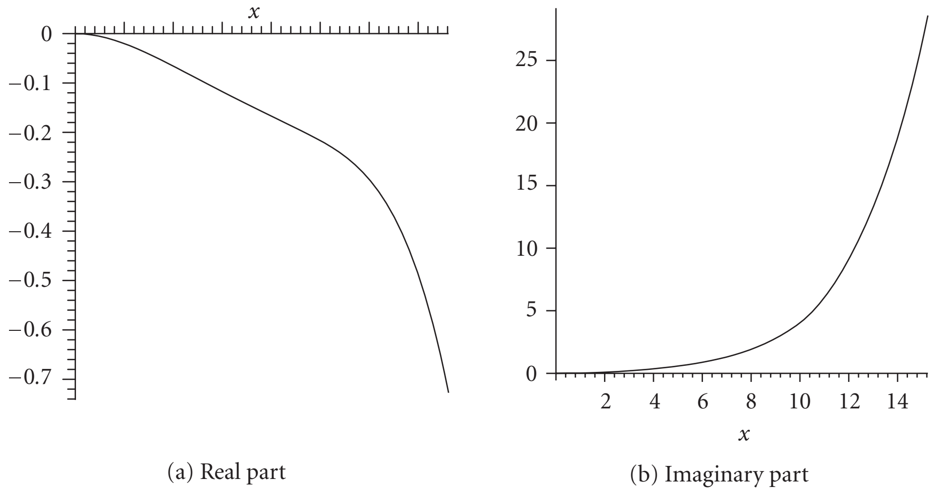
(a) Real part

(b) Imaginary part

Figure 6.8. First mode to triple-span beam.

In Figures 6.8, 6.9, and 6.10 are plotted the first three shape modes corresponding to the first three eigenvalues of the beam, where (a) it indicates the real part of the mode and (b) the imaginary part of the mode.

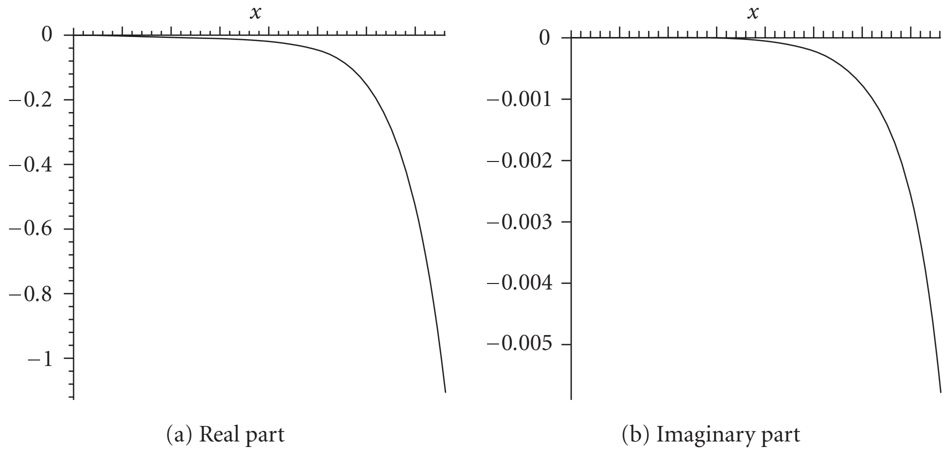
We can observe the effect of varying the parameters values in each segment of the beam on the modes shapes. The second and third modes are quite different from those of the uniform beam. This means that a beam with different sections some how influences



(a) Real part

(b) Imaginary part

Figure 6.9. Second mode to triple-span beam.



(a) Real part

(b) Imaginary part

Figure 6.10. Third mode to triple-span beam.

more the modes than external devices such as lumped mass, lumped stiffness, and lumped damping.

7. Conclusion

We have considered the study of the eigenanalysis of a triple-span Euler-Bernoulli beam subject to internal and external damping and to intermediate devices by keeping the original second-order Newtonian formulation. We also employed a matrix formulation that allows to observe the influence of the boundary and intermediate continuity conditions of the beam. Also, the values of a solution basis of the fourth-order differential equation for each segment. By choosing the elements of the basis in each segment as a convenient translation of the elements of a fundamental basis for the first segment, computations are reduced. This fundamental later is generated by a specific initial-value solution and its first three derivatives. The matrix method avoids the use of the first-order state formulation or to rely on the Euler basis of a differential equation of fourth order. It also allows to envision how conditions will influence a chosen basis.

References

[1] T. Tsukazan, “The use of a dynamical basis for computing the modes of a beam system with a discontinuous cross-section,” *Journal of Sound and Vibration*, vol. 281, no. 3–5, pp. 1175–1185, 2005.

[2] J. C. R. Claeyssen and R. A. Soder, “A dynamical basis for computing the modes of Euler-Bernoulli and Timoshenko beams,” *Journal of Sound and Vibration*, vol. 259, no. 4, pp. 986–990, 2003.

[3] J. C. R. Claeyssen and T. Tsukazan, “Dynamic solutions of linear matrix differential equations,” *Quarterly of Applied Mathematics*, vol. 48, no. 1, pp. 169–179, 1990.

[4] M. A. De Rosa, P. M. Belles, and M. J. Maurizi, “Free vibrations of stepped beams with intermediate elastic supports,” *Journal of Sound and Vibration*, vol. 181, no. 5, pp. 905–910, 1995.

[5] M. A. De Rosa, “Free vibrations of stepped beams with elastic ends,” *Journal of Sound and Vibration*, vol. 173, no. 4, pp. 563–567, 1994.

[6] D. Gorman, *Free Vibration Analysis of Beams and Shafts*, John Wiley & Sons, New York, NY, USA, 1975.

[7] B. G. Korenev and L. M. Reznikov, *Dynamic Vibration Absorbers*, John Wiley & Sons, New York, NY, USA, 1993.

[8] A. N. Krylov, “Ship vibrations,” in *Collected Works*, vol. 10, ANSSSR, Moscow, Russia, 1948.

[9] S. Naguleswaran, “Lateral vibration of a uniform Euler-Bernoulli beam carrying a particle at an intermediate point,” *Journal of Sound and Vibration*, vol. 227, no. 1, pp. 205–214, 1999.

[10] S. Naguleswaran, “Vibration of an Euler-Bernoulli beam on elastic end supports and with up to three step changes in cross-section,” *International Journal of Mechanical Sciences*, vol. 44, no. 12, pp. 2541–2555, 2002.

[11] H. V. Vu, A. M. Ordóñez, and B. H. Karnopp, “Vibration of a double-beam system,” *Journal of Sound and Vibration*, vol. 229, no. 4, pp. 807–822, 2000.

[12] M. I. Friswell and A. W. Lees, “The modes of non-homogeneous damped beams,” *Journal of Sound and Vibration*, vol. 242, no. 2, pp. 355–361, 2001.

[13] T.-P. Chang, F.-I. Chang, and M.-F. Liu, “On the eigenvalues of a viscously damped simple beam carrying point masses and springs,” *Journal of Sound and Vibration*, vol. 240, no. 4, pp. 769–778, 2001.

[14] S. Sorrentino, S. Marchesiello, and B. A. D. Piombo, “A new analytical technique for vibration analysis of non-proportionally damped beams,” *Journal of Sound and Vibration*, vol. 265, no. 4, pp. 765–782, 2003.

[15] J. Ginsberg, *Mechanical and Structural Vibrations*, John Wiley & Sons, New York, NY, USA, 2002.

[16] A. Sestieri and S. R. Ibrahim, “Analysis of errors and approximations in the use of modal coordinates,” *Journal of Sound and Vibration*, vol. 177, no. 2, pp. 145–157, 1994.

[17] S. P. Timoshenko, D. H. Young, and W. Weaver Jr., *Vibration Problems in Engineering*, John Wiley & Sons, New York, NY, USA, 1974.

[18] J. C. R. Claeyssen, G. Canahualpa, and C. Jung, "A direct approach to second-order matrix nonclassical vibrating equations," *Applied Numerical Mathematics*, vol. 30, no. 1, pp. 65–78, 1999.

Rosemaira Dalcin Copetti: Departamento de Matemática, Universidade Federal de Santa Maria, Avenida Roraima 1000, 97105-900 Santa Maria, RS, Brazil
Email address: rmaira@smail.ufsm.br

Julio C. R. Claeyssen: Instituto de Matemática-Promec, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500, 91509-900 Porto Alegre, RS, Brazil
Email address: julio@mat.ufrgs.br

Teresa Tsukazan: Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500, 91509-900 Porto Alegre, RS, Brazil
Email address: teresa@mat.ufrgs.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 82348, 25 pages
doi:10.1155/2007/82348

Research Article

Asymptotic Solution of the Theory of Shells Boundary Value Problem

I. V. Andrianov and J. Awrejcewicz

Received 2 November 2006; Revised 26 February 2007; Accepted 29 June 2007

This paper is dedicated to the memory of Professor J. J. Telega

Recommended by José Manoel Balthazar

This paper provides a state-of-the-art review of asymptotic methods in the theory of plates and shells. Asymptotic methods of solving problems related to theory of plates and shells have been developed by many authors. The main features of our paper are: (i) it is devoted to the fundamental principles of asymptotic approaches, and (ii) it deals with both traditional approaches, and less widely used, new approaches. The authors have paid special attention to examples and discussion of results rather than to burying the ideas in formalism, notation, and technical details.

1. Introduction

The theory of plates and shells is applied usually for technical purposes. However, a role of today's modern theory of plates and shells is certainly wider. In fact, in many important cases, the physical objects cannot be described by equations of 3D theory of elasticity. The examples can be biological membranes, liquid crystals, thin polymeric films, thin-walled objects made from materials with shape memory, as well as various nanostructure devices. Theory of plates and shells not only gives practically useful results, but outlines also a general methodology of the transition from 3D to 2D (or 1D) models. It is worth noting that development of mathematical physics in many cases has been motivated by theory of plates and shells problems, in particular we mean the problems associated with the application of variation and asymptotic methods. Note that a key (for singular asymptotics) concept of an edge effect appeared in the works of Lamb and Basset in 1890, while the concept of boundary layer occurred in fluid mechanics appeared only in 1904 [1]. The classical papers by Vishik and Lyusternik are a generalization of some results obtained earlier by Gol'denveĭzer [2]. On the other hand, theory of plates and shells problems associated with high-technology development of materials and constructions implied development of various homogenization procedures [3–12]. The investigation of rods stability yielded a linearization procedure, whereas Koiter's approach [13] has strongly influenced today's Catastrophe theory.

Generally, asymptotic methods are applied in the field of theory of plates and shells first for transition from 3D to 2D models, and then to solve 2D problems. Our attention is focused on the latter problem. A choice of discussed and illustrated asymptotic methods is mainly motivated by authors' subjective experience. Note that in this paper we do not concern up-to-date analysis of the existing linear and nonlinear models of shells, and a reader interested at this topic is invited to follow other known works [14–18].

We also omit here purely mathematical approaches regarding theory of shells and mainly developed by the French school (see, for instance, [19–21]).

2. On the parameter of asymptotic integration

Almost always while considering any asymptotic behavior, a term "small" or "large" parameter is applied. Since this traditional meaning may lead to confusion, we further apply the term of "asymptotic integration parameters," not restricted to be necessarily small (large). Notice that any asymptotic analysis should begin with normalization of the problem, that is defining it in terms of nondimensional variables whose typical scale is of the order of one, and the relative magnitude of different physical effects is measured by nondimensional parameters or dimensionless groups [22]. In particular, in theory of plates and shells, the following parameters are often used: h/R is the ratio of shell thickness to its characteristic size, that is, radius [2, 23]; a/b is the ratio of characteristic dimensions (i.e., a plate length to its width) [24]; ω^{-1}, where ω is the dimensionless frequency of vibrations [25]; A is the dimensionless amplitude of vibrations [26]; $\varepsilon = w/h$, where w is the normal displacement (the case $\varepsilon \ll 1$ belongs to Koiter's asymptotics [13], whereas the case $\varepsilon \gg 1$ is called Pogorolev's asymptotics [27]); B_1/B_2 is the ratio of bending stiffnesses of structurally orthotropic shell or the ratio of shear rigidity to membrane rigidity [28]; a small deviation of shell shape from canonical one [29] or a changeable thickness from a constant one; the ratio of shallow shell rise H to curvature radius R, and so on.

For periodically nonhomogeneous plates and shells, small parameter is the ratio of a period of nonhomogeneity to a characteristic size of considered structure [3–12].

If it is impossible to define a suitable real physical parameter, it can be introduced to equations in a purely formal manner (artificial parameter of asymptotic integration) [30].

"Let us try to find the asymptotics of some nontrivial solutions. First of all it is necessary to guess (no better word may be chosen) in what form this asymptotics must be sought. This stage—guessing the form of the asymptotics—of course, defines formalization. Analogies, experience, physical considerations, intuition, and "just lucky" guesses are the toolkit which is used by every investigator" [31], but after the introduction of the parameters of asymptotic integration and after the choice of an asymptotic method, it is not necessary to "reinvent the wheel"—it is better to use some well-known and well-worked out approach.

3. How to find parameters of asymptotic integration

One of the most peculiar aspects of theory of plates and shells is that associated with the existence of a few parameters of asymptotic integration yielding complexity of the problem being analyzed. In general, this fact is omitted in most studies. Therefore, a domain of application of the results is not clear enough. Gol'denveǐzer [2] indicated the importance of estimation of the order of coefficients of the partial differential equations and differential operators. In this reference, the index of variation of a function has been introduced and found to be very convenient. For example [2, 24, 32, 33], one can introduce estimations for the derivatives

$$w_x \sim \varepsilon^{\alpha} w; \qquad w_y \sim \varepsilon^{\beta} w; \qquad w_t \sim \varepsilon^{\gamma} w. \tag{3.1}$$

To compare the orders of several functions, their indices of intensity are introduced in the following way:

$$w \sim \varepsilon^{\delta}; \qquad w \sim \varepsilon^{\sigma} u. \tag{3.2}$$

Parameters of asymptotic integration $\alpha, \beta, \ldots$ are chosen in a way which yields a generalization of the Newton polygon. Notice that one gets finally not only simplified boundary value problem, but also the estimation of application domains for used asymptotic simplifications.

Let us introduce some remarks. Solutions of linear boundary value problem of theory of plates and shells usually include exponential and trigonometric functions, which causes efficiency of the described technique; but, for example, the solution of corner boundary layer type can contain powers of coordinates, and in this case the indices of variations should be applied carefully. In addition, it should be noted that the described technique gives local estimations.

Although Gol'denveǐzer's monograph [2] was published long time ago, some of the results reported there have been reconsidered again in the frame of the so-called power geometry [34].

Key steps of the method will be illustrated by the example of a membrane lying on an elastic support and governed by the equation

$$\varepsilon(w_{xx} + w_{yy}) + w = 0, \tag{3.3}$$

where $w(x, y)$ is the normal displacement of membrane, and ε is the small parameter.

The parameters of asymptotic integration α, β are introduced in the following way:

$$w_x \sim \varepsilon^{\alpha} w, \quad w_y \sim \varepsilon^{\beta} w, \quad -\infty < \alpha,\ \beta < \infty. \tag{3.4}$$

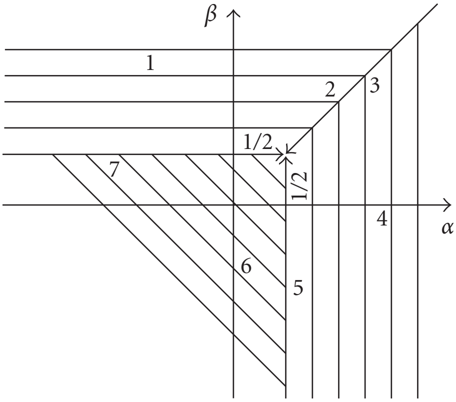


Figure 3.1. Newton polygon for (3.3).

Exponents of ε power for all terms of (3.3) follow:

$$1-2\alpha, \qquad 1-2\beta, \qquad 0. \tag{3.5}$$

Considering plane $\alpha\beta$ (see Figure 3.1), the areas corresponding to the smallest values of exponents associated with all terms of (3.3) are constructed.

Note that exponent $1-2\alpha$ is the smallest one under the choice of α and β values in area 4, exponent $1-2\beta$ in area 1, and exponent 0 in area 6 (areas 1, 4, 6 are open sets, i.e., their boundary lines are not included).

In areas 1, 4, 6 the limiting equations follow:

$$w_{yy}=0; \qquad w_{xx}=0; \qquad w=0. \tag{3.6}$$

The equations include only one term. The values of α and β associated with the equations with two terms are located on boundary lines (without point $\alpha=\beta=1/2$)

$$w_{xx}+w_{yy}=0, \qquad \varepsilon w_{xx}+w=0, \qquad \varepsilon w_{yy}+w=0. \tag{3.7}$$

Finally, for $\alpha=\beta=1/2$ in (3.3) all terms remain. Since there are no blank spaces on the $\alpha\beta$ plane, there are no other limiting systems.

Note that the occurrence of more than two parameters of the asymptotic integration results in an increase of the problem complexity. In [35, 36], effective algorithms to solve the occurring problems are introduced, whereas in [37], a generalization is proposed.

Simultaneous splitting of governing equations should be matched with an appropriate splitting of the associated boundary conditions. This complicated problem is discussed and illustrated in [2, 23, 24, 32, 33].

4. Timoshenko-type plate equations

Below, we consider an illustrative example showing the efficiency of asymptotic method [36]. According to Timoshenko, the effect of a shear deflection occurring for plate vibration is comparable to that of rotary inertia. However, the wave front sets are predicted incorrectly due to the Timoshenko theory. On the other hand, asymptotic method shows that a transverse compression effect is comparable with effects of rotary inertia and shear deflection. Correct asymptotic theory gives a proper location of wave fronts as well as averaged characteristics of stress-strain state in the vicinity of the mentioned fronts within two-dimensional equations of the form

$$\varphi_{1xx} + a_s^2\varphi_{1yy} + e\varphi_{2xy} + cW_x - 8a_s^2(w_x + \varphi_1) - \varphi_{1tt} = 0, \tag{4.1}$$

$$\varphi_{2yy} + a_s^2\varphi_{2xx} + e\varphi_{1xy} + cW_y - 8a_s^2(w_y + \varphi_2) - \varphi_{2tt} = 0, \tag{4.2}$$

$$a_s^2(w_{xx} + w_{yy}) + e(\varphi_{1x} + \varphi_{2y}) + W - w_{tt} = 0, \tag{4.3}$$

$$W + c(\varphi_{1x} + \varphi_{2y}) + 0.5w_{tt} + \frac{1}{16}W_{tt} = 0, \tag{4.4}$$

$$M_1 = \varphi_{1x} + c\varphi_{2y} + cW, \qquad M_2 = \varphi_{2y} + c\varphi_{1x} + cW, \tag{4.5}$$

$$N = W + c\varphi_{1x} + \varphi_{2y}, \tag{4.6}$$

$$H = a_s^2(\varphi_{2x} + \varphi_{1y}), \qquad Q_1 = w_x + \varphi_1 = \beta_1, \qquad Q_2 = w_y + \varphi_2 = \beta_2, \tag{4.7}$$

where: $e = 1/(2(1-\nu))$, $c = \nu/(1-\nu)$, $a_s^2 = (1-2\nu)/(1-\nu)^2$, w is the displacement of the middle plane of the plate, φ_1, φ_2 are the rotational angles of the normal to the middle plane of the plate in the x and y directions, W is the function of changing of the plate thickness, antisymmetric with respect to the middle plane of the plate.

Compare (4.1)–(4.7) with the equations of Timoshenko plate at the shear coefficient $k^2 = 2/3$,

$$\begin{gathered}
\varphi_{1xx} + \frac{1-\nu}{2}\varphi_{1yy} + \frac{1+\nu}{2}\varphi_{2xy} - 4(1-\nu)(w_x + \varphi_1) - \frac{1}{a_1^2}\varphi_{1tt} = 0,\\
\varphi_{2yy} + \frac{1-\nu}{2}\varphi_{2xx} + \frac{1+\nu}{2}\varphi_{1xy} - 4(1-\nu)(w_y + \varphi_2) - \frac{1}{a_1^2}\varphi_{2tt} = 0,\\
w_{xx} + w_{yy} + \varphi_{1x} + \varphi_{2y} - \frac{3}{2a_s^2}w_{tt} = 0,\\
M_1 = a_1^2(\varphi_{1x} + \nu\varphi_{2y}), \qquad M_2 = a_1^2(\varphi_{2y} + \nu\varphi_{1y}), \qquad H = a_s^2(\varphi_{2x} + \varphi_{1y}),\\
Q_1 = w_x + \varphi_1 = \beta_1, \qquad Q_2 = w_y + \varphi_2 = \beta_2, \qquad a_1^2 = \frac{1-2\nu}{(1-\nu)^2}.
\end{gathered} \tag{4.8}$$

Note that (4.1)–(4.7), contrary to (4.8), govern the velocities of all waves in comparison with the 3D case.

Equations (4.8) can be obtained from (4.1)–(4.7), but using the asymptotically inconsistent procedure: the last term of (4.4) as well as function N in (4.6) should be neglected; and expression $W = -c(\varphi_{1x} + \varphi_{2y})$ should be introduced to (4.2)–(4.4).

5. Dynamic edge effect method

Due to the main idea of this approach proposed by Bolotin [25], a continuous elastic system is separated into two parts. In one of them—an interior zone—solutions may be expressed by trigonometric functions with unknown constants. One can use exponential functions in the dynamic edge effect's zone. Then, a matching procedure permits to obtain unknown constants, and a complete solution of dynamic problem may be written in a relatively simple form. This approximate solution is very accurate for high-frequency vibrations, but even at low-frequency vibrations the error is not excessive. Dynamic edge effect method is naturally generalized for a nonlinear case [26, 32].

We should also emphasize that dynamic edge effect method works properly in connection with variation methods [26, 32]. This is due to the fact that the dynamic edge effect method gives good approximation of displacements. While finding the eigenvalues, the following general rule can be formulated: if you are looking for the eigenforms, then asymptotics should be used; if you need an eigenvalue, then the found asymptotic function can be used further by one of the variation methods.

6. Homogenization approach

Replacement of a nonhomogeneous shell by a homogeneous one with some reduced characteristics belongs to one of the most popular approximations in theory of plates and shells. We can mention structurally orthotropic theories of ribbed, corrugated, perforated plates and shells, plates and shells with many attached masses, and so forth. For many years, a design of similar simplifications depended fully on engineers' intuition, and the obtained quantities differed from each other depending on the theory used. Mathematical difficulties were caused by the occurrence of partial differential equations with rapidly changing coefficients. Beginning from the 1970s of the 20th century, the theory of homogenization of partial differential equations has been developed. It should be emphasized that a similar mathematical approach was proposed earlier in the theory of ribbed shells [12].

Using the homogenization approach, one must deal with two successively solvable problems: a local problem for periodically repeated element (cell) as well as the global homogeneous problem with some reduced parameters. As a rule, the fundamental difficulty is associated with solution of the cell problem. Although this problem can be solved numerically, an analytical solution is always highly required. The application of asymptotic methods to solve local problems allowed us to get homogenized solutions for various periodically nonhomogeneous plates and shells with correctly reduced coefficients. The areas of applicability of approximated theories are estimated, and full stress-strain states can be calculated. It is important that one can also predict boundary layers occurring in the vicinity of boundaries. Lack of this knowledge does not allow the shell stress-strain to be fully estimated. Using the homogenization procedure, one should take into account the relations between parameters of investigated structures. As an example, a deformation of a reinforced membrane governed by the following equation is analyzed:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = g_1(x, y), \quad kl \le y \le (k+1)l. \tag{6.1}$$

Conditions of conjugations of the neighboring parts of membrane are:

$$\lim_{y\to kl+0} u \equiv u^{+} \equiv \lim_{y\to kl-0} u \equiv u^{-}, \tag{6.2}$$

$$\left(\frac{\partial u}{\partial y}\right)^{+} - \left(\frac{\partial u}{\partial y}\right)^{-} = d_1 \frac{\partial^2 u}{\partial x^2}, \tag{6.3}$$

$$u = 0 \quad \text{for } x = 0, H. \tag{6.4}$$

Let a characteristic period of external load be $L \gg l$, $\varepsilon = l/L \ll 1$. We introduce the variables $\eta = y/l$, $y_1 = y/L$, and the following series:

$$\begin{aligned} u = {} & u_0(x,y) + \varepsilon^{\alpha_1}\left[u_{10}(x,y) + u_1(x,y,\eta)\right] \\ & + \varepsilon^{\alpha_2}\left[u_{20}(x,y) + u_2(x,y,\eta)\right] + \cdots, \quad 0 < \alpha_1 < \alpha_2 < \cdots. \end{aligned} \tag{6.5}$$

Substituting (6.5) into (6.1)–(6.4), the following recurrent system is obtained:

$$\begin{gathered} \frac{\partial^2 u_0}{\partial x^2} + \frac{\partial^2 u_0}{\partial y^2} + \varepsilon^{\alpha-2}\frac{\partial^2 u_1}{\partial \eta^2} + 2\varepsilon^{\alpha-1}\frac{\partial^2 u_0}{\partial y \partial \eta} + \varepsilon^{\alpha-2}\frac{\partial^2 u_2}{\partial \eta^2} + 2\varepsilon^{\alpha-1}\frac{\partial^2 u_0}{\partial y \partial \eta} + O(\varepsilon^{\alpha}) = q(x,y), \\ \left[u_0 + \varepsilon^{\alpha}(u_{10} + u_1) + \cdots\right]^{+} = \left[u_0 + \varepsilon^{\alpha}(u_{10} + u_1) + \cdots\right]^{-}, \\ \varepsilon^{\alpha-1}\left[\left(\frac{\partial u_1}{\partial \eta}\right)^{+} - \left(\frac{\partial u_1}{\partial \eta}\right)^{+}\right] + O(\varepsilon^{\alpha}) = d\left[\frac{\partial^2 u_0}{\partial x^2} + O(\varepsilon^{\alpha})\right], \end{gathered} \tag{6.6}$$

where: $q = L^2 q_1$, $d = d_1/L$.

The character of asymptotics depends essentially on the order of magnitude of d in comparison to ε. Let us introduce the estimation $d \sim \varepsilon^{\beta}$.

Depending on the value of β, one obtains the following limiting systems:

$$0 < \alpha < 2, \quad \frac{\partial^2 u_1}{\partial \eta^2} = 0, \tag{6.7}$$

$$\alpha = 2, \quad \frac{\partial^2 u_0}{\partial x^2} + \frac{\partial^2 u_0}{\partial y^2} + \frac{\partial^2 u_1}{\partial \eta^2} = q(x,y), \tag{6.8}$$

$$\alpha > 2, \quad \frac{\partial^2 u_0}{\partial x^2} + \frac{\partial^2 u_0}{\partial y^2} = q(x,y), \tag{6.9}$$

and the following conjugation conditions:

$$\begin{aligned} & \beta < \alpha - 1, \quad \frac{\partial^2 u_0}{\partial x^2} = 0, \\ & \beta = \alpha - 1, \quad \left[\left(\frac{\partial u_1}{\partial \eta}\right)^{+} - \left(\frac{\partial u_1}{\partial \eta}\right)^{+}\right] = d\varepsilon^{1-\alpha_1}\frac{\partial^2 u_0}{\partial x^2}, \\ & \beta > \alpha - 1, \quad \left(\frac{\partial u_1}{\partial \eta}\right)^{+} = \left(\frac{\partial u_1}{\partial \eta}\right)^{+}. \end{aligned} \tag{6.10}$$

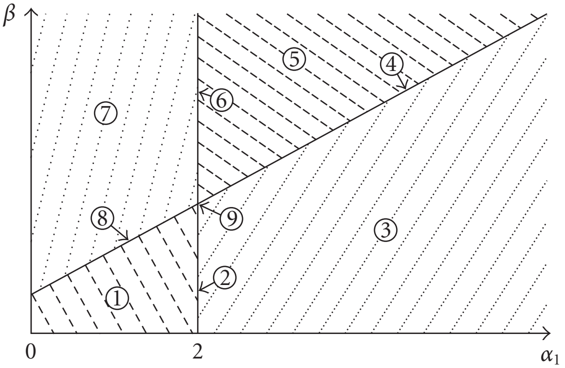


Figure 6.1. The plane of parameters $\beta > 0$, $\alpha > 0$.

The plane of parameters $\beta > 0$, $\alpha > 0$ is divided into nine parts (see Figure 6.1).

In zones 1–3, one has

$$\frac{\partial^2 u_1}{\partial \eta^2} = q(x, y). \tag{6.11}$$

In zones 4–6, the equation has the form of (6.9). For zones 7 and 9, the limiting systems are incorrect.

A particular role plays the case of $\alpha = 2$, $\beta = 1$ (zone 8). The corresponding limiting equation is (6.8) and

$$\begin{gathered} u^+ = u^-, \\ \left[\left(\frac{\partial u_1}{\partial \eta}\right)^+ - \left(\frac{\partial u_1}{\partial \eta}\right)^+\right] = d\varepsilon^{-1}\frac{\partial^2 u_0}{\partial x^2}. \end{gathered} \tag{6.12}$$

Solution of (6.8) can be written as follows.

$$u_1 = 0.5\left(\frac{\partial^2 u_0}{\partial x^2} + \frac{\partial^2 u_0}{\partial y^2} - q(x, y)\right)\eta^2 + C_1\eta + C_2. \tag{6.13}$$

Substituting solution (6.13) into conditions (6.12) yields the homogenized boundary value problem governed by the following equations:

$$\begin{gathered} \nabla^2 u_0 + \frac{d_1}{l}\frac{\partial^2 u_0}{\partial x^2} = q(x, y), \\ u_1 = \frac{d_1(x, y)}{l}\frac{\partial^2 u_0}{\partial x^2}\eta(\eta - l). \end{gathered} \tag{6.14}$$

Observe that boundary conditions (6.4) are not satisfied. In order to construct a boundary layer u_b, the new "fast" variable $\xi = x/l$ is introduced and the following series is applied:

$$u_n = \varepsilon^{\gamma_1} u_{11}(x,y,\xi,\eta) + \varepsilon^{\gamma_2} u_{22}(x,y,\xi,\eta) + \cdots, \quad 0 < \gamma_1 < \gamma_2 < \cdots. \tag{6.15}$$

Substituting series (6.15) into the governing boundary value problem yields the first approximation

$$\frac{\partial^2 u_{11}}{\partial \xi^2} + \frac{\partial^2 u_{11}}{\partial \eta^2} = 0, \qquad u_{11}|_{\eta = kl} = 0, \quad k = 0, \pm 1, \pm 2, \ldots. \tag{6.16}$$

Then, further construction of a boundary layer may be easily carried out using, for example, the Kantorovich procedure [38].

7. Distributional approach

Terms like $a(x/\varepsilon)$ often occur in the asymptotic problems. In order to introduce parameter ε explicitly, it is useful to apply the distributional approach [39]. As a model problem, we consider a transition from 2D ribs to 1D ones. The governing partial differential equations for bending deformation of an infinite plate on the elastic foundation, reinforced by periodic systems of ribs in two main directions, are as follows:

$$\begin{gathered} D\Delta\Delta w + Cw + D_1 F_1(x) w_{xxxx} + D_2 F_2(y) w_{yyyy} = q(x,y), \\ F_1(x) = \sum_{n=-\infty}^{\infty} \left[H(x + nl_1) - H(x + ml_1 + a) \right]; \\ F_2(y) = \sum_{n=-\infty}^{\infty} \left[H(y + nl_2) - H(y + ml_2 + a) \right], \end{gathered} \tag{7.1}$$

where H is the Heaviside function.

We suppose that the ribs are thin and choose their width a as the parameter of asymptotic integration. To introduce parameters a, b explicitly, we analyze function $f(x) = H(x) - H(x+a)$. Applying two-sided Laplace transformation, and using development into a Maclaurin series, one obtains

$$\overline{f}(p) = a + \frac{\sum_{n=1}^{\infty} (-1)^n a^{n+1} p^n}{(n+1)!}, \tag{7.2}$$

where $\overline{f}(p)$ is the Laplace transform of $f(x)$ $(x \to p)$.

The inverse Laplace transform leads to the following series:

$$f(x) = a\delta(x) + \frac{\sum_{n=1}^{\infty} (-1)^n a^{n+1} \delta^{(n)}(x)}{(n+1)!}, \tag{7.3}$$

where $\delta(x)$ is the Dirac function.

Functions $F_1(x)$ and $F_2(y)$ can be expanded in a similar way. As a result, we obtain the following equations:

$$D\Delta\Delta w + Cw + D_1\Phi_1(x)w_{xxxx} + D_2\Phi_2(y)w_{yyyy} = q(x,y),$$

$$\begin{aligned}\Phi_1(x) &= \Phi_{10}(x) + \Phi_{11}(x) + \Phi_{12}(x)\\ &= \sum_{n=-\infty}^{\infty} a\delta(x+nl_1) - 0.5\sum_{n=-\infty}^{\infty} a^2\delta'(x+nl_1) + \sum_{n=-\infty}^{\infty}\sum_{k=2}^{\infty}(-1)^k a^{k+1}\delta^{(n)}(x+nl_1),\\ \Phi_2(y) &= \Phi_{20}(y) + \Phi_{21}(y) + \Phi_{22}(y)\\ &= \sum_{n=-\infty}^{\infty} a\delta(y+nl_2) - 0.5\sum_{n=-\infty}^{\infty} a^2\delta'(y+nl_2) + \sum_{n=-\infty}^{\infty}\sum_{k=2}^{\infty}(-1)^k a^{k+1}\delta^{(n)}(y+nl_2).\end{aligned} \tag{7.4}$$

A solution to the equation can be sought in the form of the following series:

$$w = w_0 + \sum_{i=0}^{\infty} a^i w_i. \tag{7.5}$$

In the zero-order approximation, one gets a plate with 1D ribs governed by the following PDE:

$$D\Delta\Delta w_0 + Cw + D_1\Phi_{10}(x)w_{0xxxx} + D_2\Phi_{20}(y)w_{0yyyy} = q(x,y). \tag{7.6}$$

Note that an influence of the ribs width appears in the next approximations.

8. Real and asymptotic errors

Accuracy of asymptotic methods is usually estimated by an asymptotic error, that is, owing to the order of estimation of the last omitted term. However, a researcher engaged in theory of plates and shells is more interested in a real rather than asymptotic error. It may happen that in order to increase real accuracy of the obtained solution, one has to omit the asymptotic character of constructed solutions. Some methods for decreasing the real error of constructed approximate solutions follow.

(1) Asymptotically accurate semimembrane theory of cylindrical shells can be developed using the condition of absence of shear and torsion deformations in the shell middle surface. However, the condition of absence of shear deformations is realized with less accuracy than for torsion deformation. Although, a theory constructed on the basis of only rotary deformation absence is asymptotically inaccurate, practically it gives more accurate results [40–42].

(2) Donnell-Mushtari-Marguerre equations are good approximation of complete system of nonlinear dynamical shell equations except a case, when vibrations

form in circumferential direction can be modeled as $\cos(2\pi y/R)$. Shkutin [43] proposed a slight modification of the Donnell-Mushtari-Marguerre equations for overcoming this drawback, which is exhibited by the following equations:

$$\begin{gathered}\frac{D}{h}\nabla_1^4 W = W_{xx}\Phi_{yy} + W_{yy}\Phi_{xx} - 2W_{xy}\Phi_{xy} + \frac{1}{R}\Phi_{xx} - \rho h W_{tt} = 0,\\ \frac{1}{E}\nabla_1^4\Phi + \frac{1}{R}W_{xx} + W_{xx}W_{yy} - \left(W_{xy}\right)^2 = 0,\end{gathered} \tag{8.1}$$

where:

$$\nabla_1^4 = \frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2\partial y^2} + \frac{\partial^2}{\partial y^2}\left(\frac{\partial^2}{\partial y^2} + \frac{1}{R^2}\right). \tag{8.2}$$

(3) Owing to the asymptotic splitting of the boundary value problems, a fundamental error is introduced by simplification of the boundary conditions. In many cases, one may analytically obtain a general solution of edge effect equations. Using this solution, it is possible to exclude exactly the terms of the edge effect solution from the boundary conditions and hence avoid splitting of the boundary conditions [44].

(4) The method of composite equations is devoted to constructing uniformly suitable solutions on the basis of various limiting cases [45]. A fundamental idea of the method can be formulated in the following way. First, the components of the governing equations are detected, which, when neglected, lead to nonhomogeneity in a zero-order approximation. Second, the mentioned components are defined in a relatively simple way (they must include essential properties in the nonhomogeneous states). Matching of the limiting relations leads to uniformly suitable equations. In the theory of plates and shells, a composite equation of the stress-strain fundamental state has been obtained, unifying the semimembrane and membrane theories and a plane plate deformation. A simple edge effect and bending of the plate are included in a composite equation of the edge effect type. The obtained composite equations are of the fourth order because of a longitudinal variable and are applicable in the whole range of different loadings [26, 32, 46].

(5) For a posteriori error estimation of the asymptotic solutions, singular version of the Kantorovich theorem [46] can be successfully used [47].

9. Beyond the series locality

The principal shortcoming of asymptotic methods is the local nature of solutions yielded by them. Problems of elimination of the expansion locality, evaluation of the convergence domain, and construction of uniformly suitable solutions are highly expected.

There are many approaches to these problems [26, 30, 32, 48–50]: the method of analytic continuation, Borel summation procedure, Euler transformation, and Domb-Sykes diagram [45]. As a rule, they need a significant number of the expansion components.

Not diminishing the merits of the mentioned techniques, let us, however, note that in practice, only a few of the first components of the expansion of perturbations are usually known. Lately, the situation has indeed changed a little due to computer application. It may happen that a number of terms of asymptotic series can be increased without any serious problems. For instance, computing improvement terms with respect to an eigenvalue are usually successfully defined by eigenvalues and eigenfunctions. The knowledge of the nth eigenfunction allows us to define $2n+1$ eigenvalues [51]. However, until now, there are usually 3–5 components available in a perturbation series, and exactly from this segment of the series, we have to extract all available information. To this end, the method of Padé approximants may be very useful. Let us now define Padé approximants. Let

$$\begin{aligned} F(\varepsilon) &= \sum_{i=0}^{\infty} C_i \varepsilon^i, \\ F_{mn}(\varepsilon) &= \frac{\sum_{i=0}^{m} a_i \varepsilon^i}{\sum_{i=0}^{m} b_i \varepsilon^i}, \end{aligned} \tag{9.1}$$

where the coefficients a_i, b_i are determined from the following condition: the first $(m+n)$ components of the expansion of the rational function $F_{mn}(\varepsilon)$ in a Maclaurin series coincide with the first $(m+n+1)$ components of the series $F(\varepsilon)$. $F_{mn}(\varepsilon)$ is the Padé approximation of the function $F(\varepsilon)$.

Padé approximants perform meromorphic continuation of the function given in the form of the power series. If the Padé approximants sequence converges to a given function, then the roots of its denominators tend to singular points.

A wide application of the Padé approximants is observed due to its suitable properties. Among others, we should mention the effect of error autocorrection: even very significant errors in the coefficients of Padé approximants do not affect the accuracy of the approximation [52, 53]. This is because the errors in the numerator and the denominator of Padé approximants compensate each other. In other words, the errors in the coefficients of the Padé approximants are not distributed in an arbitrary way, but form the coefficients of a new approximant to the approximated function.

Padé approximants can be used for a heuristic evaluation of the domain of applicability of a perturbation series. The ε values, up to which the difference between calculations according to the truncated perturbation series and its diagonal Padé approximants do not exceed a given value (e.g., 5%), can be considered as a limiting value for applicability of the perturbation series.

10. Homotopy perturbation technique

Dorodnitzyn [54] proposed a method of introducing the parameter ε into the input boundary value problem in such a way that for $\varepsilon = 0$ the simplified problem is obtained, whereas for $\varepsilon = 1$ the input problem is governed. Then, the perturbation method can be used. Now, this approach is known as a homotopy perturbation technique [49, 55].

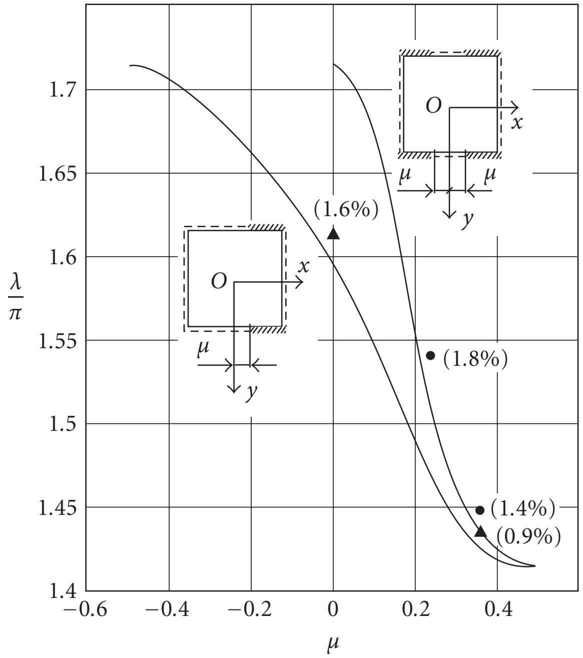


Figure 10.1. Relationship between the vibration frequency λ and the clamped segment length.

Unfortunately, perturbation series for $\varepsilon = 1$ usually diverges. In order to overcome this difficulty, the Padé approximants can be used effectively [26, 30, 32, 56].

Let us focus on the application of the homotopy perturbation method [26, 30, 32] when solving mixed BVP of the vibration of a rectangular plate ($-0.5k \leq x \leq 0.5k$, $-0.5 \leq y \leq 0.5$), simply supported at $x = \pm 0.5k$, and having mixed boundary conditions of the "clamped-simple supported" type, symmetrical to the y axis or the sides $y = \pm 0.5k$ (Figure 10.1). The governing equation is

$$\nabla^4 w - \lambda w = 0. \tag{10.1}$$

The boundary conditions after introducing a homotopy parameter have the following form:

$$\begin{gathered} w = 0, \quad w_{xx} = 0 \quad \text{for } x = \pm 0.5k, \\ w = 0, \quad w_{yy} = \overline{H}(x)\varepsilon\left(w_{yy} \pm w_y\right) \quad \text{for } y = \pm 0.5, \end{gathered} \tag{10.2}$$

where $\overline{H}(x) = -H(x-\mu) + H(-x-\mu)$.

Substituting w and λ in the form of ε-series

$$w = w_0 + \varepsilon w_1 + \cdots, \qquad \lambda = \lambda_0 + \varepsilon\lambda_1 + \cdots, \tag{10.3}$$

and after applying the usual perturbation procedure to the boundary value problem (10.1), (10.2), one obtains

$$\begin{aligned}
&\lambda_0 = \pi^4\psi^2, \qquad \lambda_1 = 4\pi^2 n^2 \gamma_{mm},\\
&\lambda_2 = 4\pi^2 n^2 \gamma_{mm}\left\{1 - \frac{\gamma_{mm}}{\pi^2\psi}\left[\frac{\pi\alpha}{2}\operatorname{cth}^{(-1)^m}\left(\frac{\pi\alpha}{2}\right) + \frac{n^2}{\psi} - \frac{3}{2}\right]\right\}\\
&\quad - \frac{2n^2}{\psi}\sum_{\left\{\substack{i=1,3,5,\dots\\ i=2,4,6,\dots}\right\}}^{\infty} \gamma_{im}\left[\alpha_i \operatorname{cth}^{(-1)^i}\left(\frac{\alpha_i}{2}\right) + \left\{\begin{matrix} -\phi_i \operatorname{cth}^{(-1)^i}(\phi_i/2) \\ \beta_i \operatorname{cth}^{(-1)^i}(\beta_i/2)\end{matrix}\right\}\right], \quad \left\{\begin{matrix} i^2 > m^2 + n^2 k \\ i^2 < m^2 + n^2 k\end{matrix}\right\},
\end{aligned} \tag{10.4}$$

where

$$\psi = n^2 + \frac{m^2}{k^2}, \qquad \alpha = \sqrt{2\frac{m^2}{k^2} + n^2}, \qquad \alpha_i = \sqrt{\frac{i^2 + m^2}{k^2} + n^2}, \qquad \beta_i = \pi\sqrt{\frac{m^2 - i^2}{k^2} + n^2},$$

$$\gamma_{im} = \begin{cases} 2(0.5 - \mu) + \dfrac{(-1)^m}{\pi m}\sin(2\pi\mu m), & \text{for } i = m\\[2ex] \dfrac{4}{\pi}\dfrac{1}{(m^2 - i^2)}\left[\left\{\begin{matrix} i \\ m\end{matrix}\right\}\sin(\pi\mu i)\cos(\pi\mu m) - \left\{\begin{matrix} m \\ i\end{matrix}\right\}\sin(\pi\mu m)\cos(\pi\mu i)\right], & \text{for } i \neq m,\end{cases} \tag{10.5}$$

and $\sum'$ is the sum without the component $i = m$.

Truncated perturbation series for $\mu = 0$ (both sides $y = \pm 0.5$ are completely clamped) for the square plate gives $(1.4783\pi)^4$. Padé approximants are

$$\lambda_p(\varepsilon) = \frac{a_0 + a_1\varepsilon}{1 + b_1\varepsilon}, \quad a_0 = \lambda_0,\ a_1 = \lambda_1 + b_1\lambda_0,\ b_1 = \frac{-\lambda_2}{\lambda_1}, \tag{10.6}$$

and for $\varepsilon = 1$ one obtains $\lambda_p = (1.7081\pi)^4$, while numerical value $\lambda = (1.7050\pi)^4$. Figure 10.1 presents the relation of λ versus μ and some experimental data (dots and triangles).

11. Theories of higher-order approximations

In order to increase approximation accuracy, the terms of higher order may remain in the input equations, but such an approach can increase the order of the approximate partial differential equation. This problem can be overcome by Padé approximants. Let us consider vibrations of a stretched beam modeled by the following equations:

$$w_{tt} - w_{\xi\xi} + \varepsilon w_{\xi\xi\xi\xi} = 0, \tag{11.1}$$

$$w = w_{\xi\xi} = 0 \quad \text{for } \xi = 0, 1. \tag{11.2}$$

Note that one may obtain a string-type model from (11.1) for $\varepsilon = 0$, namely,

$$w_{tt} - w_{\xi\xi} = 0, \tag{11.3}$$

$$w = 0 \quad \text{for } \xi = 0, 1. \tag{11.4}$$

In (11.1), instead of the differential operator $-\partial^2/\partial\xi^2 + \varepsilon\partial^4/\partial\xi^4$, one can use the following approximation:

$$\frac{-\partial^2}{\partial\xi^2} + \frac{\varepsilon\partial^4}{\partial\xi^4} \approx \frac{-\partial^2/\partial\xi^2}{(1+\varepsilon\partial^2/\partial\xi^2)}. \tag{11.5}$$

Finally one obtains

$$\left(1+\varepsilon\frac{\partial^2}{\partial\xi^2}\right)w_{tt} - w_{\xi\xi} = 0. \tag{11.6}$$

The associated boundary conditions have the form (11.4). Observe that if the model (11.3), (11.4) approximates eigenvalues of the initial problem up to the order of ε, then model (11.6), (11.4) includes second-order approximation of ε^2 preserving the equation order with respect to the spatial coordinates.

12. Matching of limiting asymptotics

It happens often that solutions related to two limiting values of a certain parameter can be easily constructed. In this case, one can define a solution valid for all parameter values with a help of two-point Padé approximants [26, 30, 32]. Let

$$F(\varepsilon) = \begin{cases} \sum\limits_{i=0}^{\infty} a_i\varepsilon^i & \text{when } \varepsilon \longrightarrow 0, \\ \sum\limits_{i=0}^{\infty} b_i\varepsilon^{-i} & \text{when } \varepsilon \longrightarrow A. \end{cases} \tag{12.1}$$

The two-point Padé approximation is represented by the following rational function:

$$F(\varepsilon) = \frac{\sum_{k=0}^{m} a_k\varepsilon^k}{\sum_{k=0}^{n} b_k\varepsilon^k}, \tag{12.2}$$

where $k+1$ $(k = 0,1,\dots,n+m+1)$ are the coefficients of a Taylor expansion if $\varepsilon \to 0$, and $m+n+1-k$ are the coefficients of a Laurent series, and for $\varepsilon \to A$ they coincide with the corresponding coefficients of the series (12.1).

As an example, we consider the problem of nonlinear deformation of a sphere. The solution

$$Q = 0.42\varepsilon + 0.3\varepsilon^3 + 0(\varepsilon^5),$$
$$\varepsilon = 2(w/h)\sqrt{3\sqrt{1-\nu^2}}, \qquad Q = \frac{0.5qR^2 3\sqrt{1-\nu^2}}{Eh^2} \tag{12.3}$$

has been obtained by means of the asymptotic methods for a closed sphere subjected to the uniform external pressure q [27]. In the above, w is the amplitude of post-buckling axially symmetric equilibrium form.

In the region of small displacements, the Koiter approach [13] holds, and hence

$$Q = 1 + 0(\varepsilon^{-4}). \tag{12.4}$$

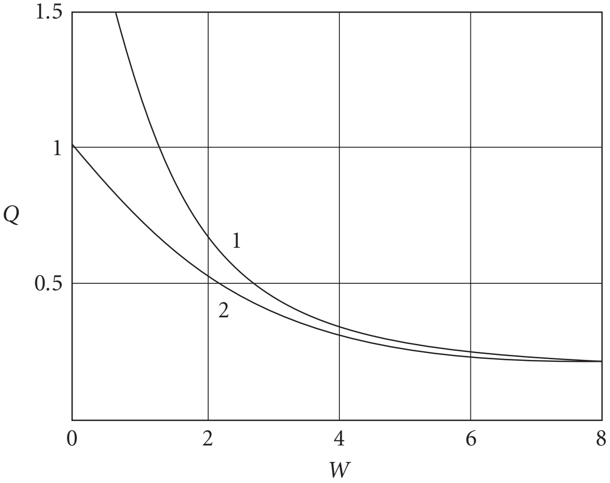


FIGURE 12.1. Matching of quasilinear and essentially nonlinear asymptotics.

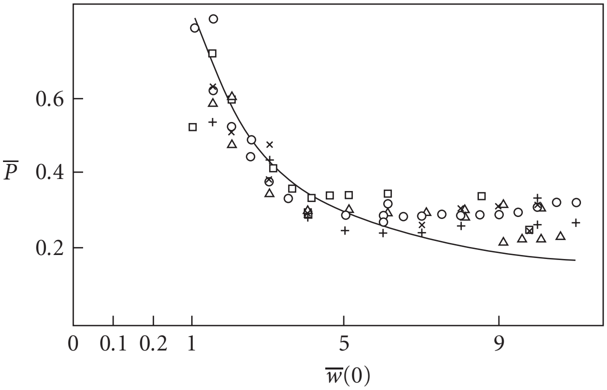


FIGURE 12.2. Comparison of two-point Padé approximation solution (solid line) with experimental results.

By matching expansions (12.3) and (12.4) with the two-point Padé approximation, one obtains the following solution [27]:

$$Q = \frac{A}{A + 2.19}, \quad A = \varepsilon^4 + 0.082\varepsilon^3 + 0.386\varepsilon^2 + 0.92\varepsilon. \tag{12.5}$$

Curves 1 and 2 in Figure 12.1 correspond to solutions (12.3), (12.5), respectively. Accuracy of solution (12.5) is confirmed by comparison with the precise numerical solution.

In Figure 12.2, results of comparison of experimental data for post-buckling equilibrium states of shallow elliptic parabolic-shaped shells under external pressure [57] with the solution based on two-point Padé approximation [27] are shown, where $\overline{w} = w/h$; $\overline{P} = (0.5qR_1R_2 3\sqrt{1-\nu^2})/Eh^2$.

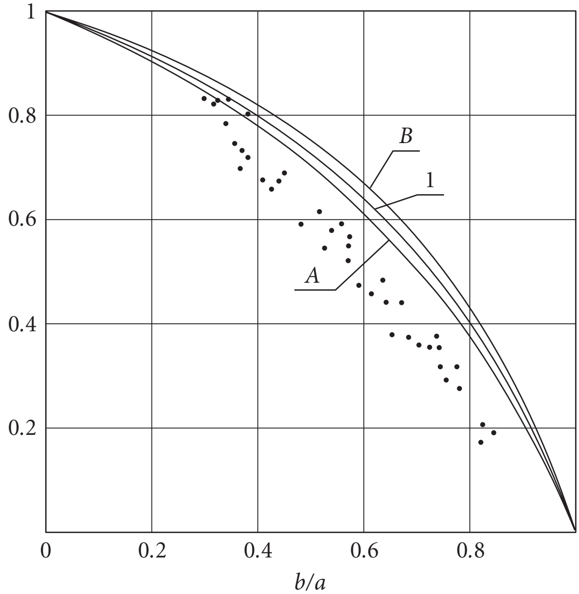


Figure 12.3. Homogenized coefficients of perforated plates.

The second example is associated with homogenization of a rectangular plate with circular perforations. Analytical solutions for small and large holes were obtained [12] by using the AM perturbation of the domain and boundary form. Coefficients A and B of the homogenized equation

$$A\left(W_{xxxx}+W_{yyyy}\right)+2BW_{xxyy}=q(x,y) \tag{12.6}$$

are yielded by the following expressions (for $\nu = 0.3$):

$$A=\frac{1-\lambda}{1-0.5785\lambda},\qquad B=\frac{1-\lambda}{1-0.6701\lambda}, \tag{12.7}$$

where $\lambda = b/a$ (b is the diameter of the hole, a is the length of the square cell side).

Figure 12.3 shows the numerical results for A and B.

The values of coefficients are compared with both theoretical results obtained by means of the two-period elliptic functions (curve 1 in Figure 12.3) and experimental results (points in Figure 12.3).

Evidently, the two-point Padé approximants are not a panacea. As a rule, one of the limit expansions ($\varepsilon \to 0$ or $\varepsilon \to A$) contains logarithmic or exponential terms. In this case, one can use the method of asymptotically equivalent functions. Suppose that we have a perturbation approach in powers of ε for $\varepsilon \to 0$ and asymptotic expansions $F(\varepsilon)$ containing logarithm for $\varepsilon \to A$. By definition, an asymptotically equivalent function is the ratio with unknown coefficients a_i, b_i, containing both powers of ε and function $F(\varepsilon)$. The coefficients a, b are chosen in such a way, that the expansion of a ratio in powers of

ε matches the corresponding perturbation expansion and the asymptotic behavior of the ratio for $\varepsilon \to \infty$ coincides with $F(\varepsilon)$.

13. Nonlinear problems

Although asymptotic techniques regarding linear problems of theory of plates and shells are relatively good developed, there are still many unsolved tasks related to nonlinear problems. In particular, nonlinear systems with distributed parameters exhibit various internal resonance between modes. It may happen that neglection of higher modes may yield also erroneous results [58]. In [59], the asymptotic method has been proposed, where all modes of vibrations can be approximately applied. In order to show main features of the proposed approach, let us consider free vibrations of a membrane attached to a nonlinear foundation. The governing equation follows:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - \beta_1 u - \varepsilon\beta_2 u^3 = 0, \tag{13.1}$$

where ε is a nondimensional parameter ($\varepsilon \ll 1$).

The boundary conditions have the following form:

$$u|_{x=0,l_1} = u|_{y=0,l_2} = 0. \tag{13.2}$$

A being sought periodic solution satisfies the following periodicity requirement:

$$u(t) = u(t+T), \tag{13.3}$$

where $T = 2\pi/\omega$ is a period, and ω is a natural vibration frequency.

We are going to find natural frequencies of vibrations associated with such fundamental modes that in the associated linear case (for $\varepsilon = 0$) only half waves appear in both x- and y-directions. Now we proceed in a usual way, using Lindstedt-Poincaré procedure [60]. Namely, we scale time

$$\tau = \omega t, \tag{13.4}$$

and the following series are introduced:

$$\begin{aligned} u &= u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots, \\ \omega &= \omega_0 + \varepsilon\omega_1 + \varepsilon^2\omega_2 + \cdots. \end{aligned} \tag{13.5}$$

Substitution of relations (13.5) into (13.1), (13.2), (13.3) and comparison of the terms standing by the same powers of ε yields the following set of linear boundary value problems:

$$\frac{\partial^2 u_0}{\partial x^2} + \frac{\partial^2 u_0}{\partial y^2} - \omega_0^2\frac{\partial^2 u_0}{\partial \tau^2} - \beta_1 u_0 = 0, \tag{13.6}$$

$$\frac{\partial^2 u_1}{\partial x^2} + \frac{\partial^2 u_1}{\partial y^2} - \omega_1^2\frac{\partial^2 u_1}{\partial \tau_1^2} - \beta_1 u_1 = 2\omega_0\omega_1\frac{\partial^2 u_0}{\partial \tau^2} + \beta_2 u_0^3. \tag{13.7}$$

Both boundary conditions (13.2) and periodicity relations (13.3) are cast to the following form:

$$u_i|_{x=0,l_1} = u_i|_{y=0,l_2} = 0; \qquad u_i(\tau) = u_i(\tau + 2\pi). \tag{13.8}$$

A solution to (13.6) has the following form

$$u_{0,0} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_{m,n} \sin\left(\frac{\omega_{m,n}^{\text{lin}}}{\omega_{1,0}}\tau\right)\sin\left(\frac{\pi m}{l_1}x\right)\sin\left(\frac{\pi n}{l_2}y\right), \tag{13.9}$$

where $A_{1,1}$ is the amplitude of the principal mode; $A_{m,n}$, $m,n = 1,2,3,\ldots$, $(m,n) \neq (1,1)$ are the amplitudes of the successive modes: $\omega_{m,n}^{\text{lin}} = \sqrt{(\pi^2 m^2/l_1^2) + (\pi^2 n^2/l_2^2) + \beta_1}$, $m,n = 1,2,3,\ldots$ are the natural frequencies of the linear system, and $\omega_0 = \omega_{1,1}^{\text{lin}}$.

Next approximation regarding ε is found owing to solution of the boundary value problem governed by (13.7), (13.8). In order to cancel the secular terms, the coefficient standing by the terms of the form $\sin((\omega_{m,n}^{\text{lin}}/\omega_0)\tau)\sin((\pi m/l_1)x)\sin((\pi n/l_2)y)$, $m,n = 1,2,3,\ldots$, occurred in the right-hand side of (13.7) are assigned to zero. This approach yields the following infinite system of algebraic equations:

$$\frac{2A_{m,n}\omega_1}{\beta_2\omega_0}\left(\omega_{m,n}^{\text{lin}}\right)^2 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{s=1}^{\infty} C_{m,n}^{(ijklps)} A_{i,j}A_{k,l}A_{p,s}, \quad m,n = 1,2,3,\ldots. \tag{13.10}$$

The coefficients $C_{m,n}^{(ijklps)}$ are found via substitution of relations (13.9) into the right-hand side of (13.7).

Note that system (13.10) can be solved via a reduction method. However, owing to introduction of many equations the essential difficulties regarding efficient computation may appear. Besides, the mentioned approach does not include higher modes interaction. In order to omit the mentioned problem, one may introduce a new parameter μ such that for $\mu = 0$ the studied system is essentially simplified. Then, the solution as a series regarding that parameter is constructed and finally $\mu = 1$ is assumed.

In our study in the right-hand side of each (m,n)th equation of system (13.10), a parameter μ is introduced before the terms $A_{i,j}A_{r,l}A_{p,s}$ for which the following condition holds: $(i > m) \cup (k > m) \cup (p > m) \cup (j > n) \cup (l > n) \cup (s > n)$. Note that now system (13.10) takes "triangle" form for $\mu = 0$, whereas for $\mu = 1$ the system takes the initial form. The solution is sought further in the following series form:

$$\begin{gathered}\omega_1 = \omega^{(0)} + \mu\omega^{(1)} + \mu^2\omega^{(2)} + \cdots, \\ A_{m,n} = A_{m,n}^{(0)} + \mu A_{m,n}^{(1)} + \mu^2 A_{m,n}^{(2)} + \cdots, \quad m,n = 1,2,3,\ldots, (m,n) \neq (1,1),\end{gathered} \tag{13.11}$$

and then one assumes $\mu = 1$. Finally, let us emphasize that the mentioned approach allows to contain in systems (13.10) arbitrary number of equations.

14. Merits and demerits of the asymptotic methods

Advantages of asymptotic methods follow.

(1) Essentially simplified solutions, which in many cases can be obtained in an analytical way.

(2) Asymptotic methods are easily matched with other approaches, that is, numerical, variational ones, and so forth. Owing to the introduced simplification of the input boundary value problem and separation of the associated peculiarities of the considered problem, one may effectively apply numerical approaches. Asymptotic methods allow us to exhibit the structure of solution and the type of approximating functions in the Bubnov-Galerkin, Rayleigh-Ritz, Trefftz and Kantorovich approaches. Owing to the construction of zero order solution, it can be applied as a starting solution for other iteration processes like the Newton-Kantorovich method.

(3) Asymptotic methods are strictly associated with a physical aspect of the analyzed problem allowing for it easier understanding.

(4) Asymptotic methods allow us to explain mathematical and physical bases of approximated engineering methods, increasing their accuracy and reliability of obtained results.

(5) Asymptotic methods give a possibility of a unified approach to various different problems exhibiting their common aspects and internal unity.

However, the main drawback of asymptotic methods is generated by insufficiently accurate results of low approximations, since a construction of successive approximations is not always easy. Also an accuracy of the estimation of asymptotic methods and intervals of their applicability in many cases causes serious difficulties.

15. Concluding remarks

Many important methods like WKB [24] or matched asymptotic expansion [61] are omitted in our review. Other interesting problems such as junction of plates and shells with 1D and 3D bodies or junction of two shells [62, 63], solutions of shell problems in singular domains [60, 64], have not been considered either.

In addition, an application of the asymptotic methods in the localization problems [24, 65] are also omitted in our paper.

One can also add to this list the problems of bonding, which arise in laminated plates and shells and which attracted many researches in the recent past [66, 67].

Important results related to the so-called first-order accuracy problems in theory of plates and shells have been reported by Gol’denveĭzer et al. [68] and Nazarov [69]. They show, among others, that inclusion into consideration of 3D boundary layers may improve accuracy order of the being modeled systems.

In our paper, we are mainly focused on linear problems. On the other hand there is no doubt that the development regarding application of asymptotic methods devoted to analysis of nonlinear problems plays a key role in nonlinear problems of both dynamics and stability of continuous systems [70, 71].

It is expected that further development of asymptotic methods is associated with combined numerical-analytical approaches and includes them in standard codes. This is important because an accurate numerical computation of shells with arbitrarily small thickness is impossible in practice. Standard finite-element codes usually fail to give accurate results for $h/R \sim 0.01$ or 0.001.

Nowadays, in order to compute thin-walled structures, the standard finite-element codes are used. It seems that asymptotic information is rather rarely applied. On the other hand, asymptotic methods belong to fundamental ones during the construction of mathematical models of physical processes [22, 72]. "Design of computational or experimental schemes without the guidance of asymptotic information is wasteful at best, dangerous at worst, because of possible failure to identify crucial (stiff) features of the process and their localization in coordinate and parameter space. Moreover, all experience suggests that asymptotic solutions are useful numerically far beyond their nominal range of validity, and can often be used directly, at least at a preliminary product design stage, for example, saving the need for accurate computation until the final design stage, where many variables have been restricted to narrow ranges" [72].

Finally, there are many books and papers devoted to the considered problems, and therefore only some of them are cited. However, a reader may find additional references in [24, 30, 32, 33, 50, 60, 65, 73–76] to extend knowledge associated with asymptotic approaches to plates and shells modeling.

Acknowledgments

Authors are grateful to Professor A.D. Shamrovskii and Professor V.N. Pilipchuk, whose valuable suggestions and comments helped to improve this paper.

References

[1] C. R. Calladine, "The theory of thin shell structures 1888–1988," *Proceedings of the Institution of Mechanical Engineers A*, vol. 202, no. 42, pp. 141–149, 1988.

[2] A. L. Gol'denveĭzer, *Theory of Elastic Thin Shells*, International Series of Monographs on Aeronautics and Astronautics, Pergamon Press, New York, NY, USA, 1961.

[3] G. Allaire, *Shape Optimization by the Homogenization Method*, vol. 146 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 2002.

[4] N. Bakhvalov and G. Panasenko, *Homogenisation: Averaging Processes in Periodic Media. Mathematical Problems in the Mechanics of Composite Materials*, vol. 36 of *Mathematics and Its Applications*, Kluwer Academic, Dordrecht, The Netherlands, 1989.

[5] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, vol. 5 of *Studies in Mathematics and Its Applications*, North-Holland, Amsterdam, The Netherlands, 1978.

[6] D. Caillerie, "Thin elastic and periodic plates," *Mathematical Methods in the Applied Sciences*, vol. 6, no. 2, pp. 159–191, 1984.

[7] D. Cioranescu and J. S. J. Paulin, *Homogenization of Reticulated Structures*, vol. 136 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1999.

[8] A. L. Kalamkarov, *Composite and Reinforced Elements of Constructions*, John Wiley & Sons, Chichester, UK, 1992.

[9] A. L. Kalamkarov and A. G. Kolpakov, *Analysis, Design and Optimization of Composite Structures*, John Wiley & Sons, Chichester, UK, 1997.

[10] V. Kozlov, V. Maz'ya, and A. Movchan, *Asymptotic Analysis of Fields in Multi-Structures*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, NY, USA, 1999.

[11] T. Lewiński and J. J. Telega, *Plates, Laminates and Shells*, vol. 52 of *Series on Advances in Mathematics for Applied Sciences*, World Scientific, River Edge, NJ, USA, 2000.

[12] L. I. Manevitch, I. V. Andrianov, and V. G. Oshmyan, *Mechanics of Periodically Heterogeneous Structures*, Foundations of Engineering Mechanics, Springer, Berlin, Germany, 2002.

[13] W. T. Koiter, "Elastic stability and post-buckling behavior," in *Nonlinear Problems (Madison, Wis, 1962)*, pp. 257–275, University of Wisconsin Press, Madison, Wis, USA, 1963.

[14] E. L. Axelrad, *Theory of Flexible Shells*, vol. 28 of *North-Holland Series in Applied Mathematics and Mechanics*, North-Holland, Amsterdam, The Netherlands, 1987.

[15] H. Altenbakh and P. A. Zhilin, "General theory of elastic simple shells," *Advances in Mechanics*, vol. 11, no. 4, pp. 107–148, 1988 (Russian).

[16] W. Pietraszkiewicz, "Geometrically nonlinear theories of thin elastic shells," *Advances in Mechanics*, vol. 12, no. 1, pp. 51–130, 1989.

[17] A. Libai and J. G. Simmonds, "The nonlinear elastic shell theory," *Advances in Applied Mechanics*, vol. 23, pp. 271–371, 1983.

[18] V. A. Eremeyev, "Micropolar shells: theory and applications," in *Shell Structures:Theory and Applications: Proceedings of the 8th SSTA Conference, Jurata, Poland*, W. Pietraszkiewicz and C. Szymczak, Eds., pp. 11–18, Taylor & Francis/Balkema, London, UK, 2005.

[19] P. G. Ciarlet, *Introduction to Linear Shell Theory*, vol. 1 of *Series in Applied Mathematics*, Gauthier-Villars, Paris, France, 1998.

[20] P. G. Ciarlet, *Mathematical Elasticity. Vol. III: Theory of Shells*, vol. 29 of *Studies in Mathematics and Its Applications*, North-Holland, Amsterdam, The Netherlands, 2000.

[21] P. G. Ciarlet, L. Trabucho, and J. M. Viaño, Eds., *Asymptotic Methods for Elastic Structures: Proceedings of the International Conference, Lisbon, Portugal, October, 1993*, Walter de Gruyter, Berlin, Germany, 1995.

[22] A. B. Tayler, *Mathematical Models in Applied Mechanics*, Oxford Texts in Applied and Engineering Mathematics, The Clarendon Press, Oxford University Press, New York, NY, USA, 2001.

[23] A. L. Gol'denveĭzer, V. B. Lidskiĭ, and P. E. Tovstik, *Free Oscillations of Thin Elastic Shells*, Nauka, Moscow, Russia, 1979.

[24] S. M. Bauer, S. B. Filippov, A. L. Smirnov, and P. E. Tovstik, "Asymptotic methods in mechanics with applications to thin shells and plates," in *Asymptotic Methods in Mechanics*, vol. 3 of *CRM Proc. Lecture Notes*, pp. 3–140, American Mathematical Society, Providence, RI, USA, 1993.

[25] V. V. Bolotin, *Random Vibrations of Elastic Systems*, vol. 8 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics of Elastic Stability*, Martinus Nijhoff, Dordrecht, The Netherlands, 1984.

[26] J. Awrejcewicz, I. V. Andrianov, and L. I. Manevitch, *Asymptotic Approaches in Nonlinear Dynamics: New Trends and Applications*, Springer Series in Synergetics, Springer, Berlin, Germany, 1998.

[27] A. Y. Evkin, "A new approach to asymptotic integration of equations in the theory of shallow convex shells in the post-critical stage," *Journal of Applied Mathematics and Mechanics*, vol. 53, no. 1, pp. 92–96, 1989.

[28] L. I. Manevich, A. V. Pavlenko, and S. G. Koblik, *Asymptotic Methods in the Elasticity Theory of an Orthotropic Body*, Vishcha Shkola, Kiev, Ukraine, 1982.

[29] D. Henry, *Perturbation of the Boundary in Boundary-Value Problems of Partial Differential Equations*, vol. 318 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, UK, 2005.

[30] I. V. Andrianov and J. Awrejcewicz, "New trends in asymptotic approaches: summation and interpolation methods," *Applied Mechanics Reviews*, vol. 54, no. 1, pp. 69–92, 2001.

[31] M. V. Fedoryuk, "Asymptotic methods in analysis," in *Analysis I*, R. V. Gamkrelidze, Ed., vol. 13 of *Encyclopaedia of Mathematical Sciences*, Springer, Berlin, Germany, 1989.

[32] I. V. Andrianov, J. Awrejcewicz, and L. I. Manevitch, *Asymptotical Mechanics of Thin-Walled Structures: A Handbook*, Springer, Berlin, Germany, 2004.

[33] J. D. Kaplunov, L. Yu. Kossovich, and E. V. Nolde, *Dynamics of Thin Walled Elastic Bodies*, Academic Press, San Diego, Calif, USA, 1998.

[34] A. D. Bruno, *Power Geometry in Algebraic and Differential Equations*, Elsevier, Amsterdam, The Netherlands, 2000.

[35] A. D. Shamrovskii, "Asymptotic integration of static equation of the theory of elasticity in Cartesian coordinates with automated search of integration parameters," *Journal of Applied Mathematics and Mechanics*, vol. 43, no. 5, pp. 925–934, 1979.

[36] A. D. Shamrovskii, *Asymptotic Group Analysis of the Differential Equations of the Theory of Elasticity*, Zaporozhie State Engineering Academy, Zaporozhie, Ukraine, 1997.

[37] A. D. Shamrovskii, I. V. Andrianov, and J. Awrejcewicz, "Asymptotic-group analysis of algebraic equations," *Mathematical Problems in Engineering*, vol. 2004, no. 5, pp. 411–451, 2004.

[38] L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis*, Noordhoff, Groningen, The Netherlands, 1958.

[39] R. Estrada and R. P. Kanwal, *A Distributional Approach to Asymptotics. Theory and Applications*, Birkhäuser Advanced Texts: Basel Textbooks, Birkhäuser, Boston, Mass, USA, 2nd edition, 2002.

[40] A. F. Feofanov, *Structural Mechanics of Thin-Walled Structures*, Mashinostroenie, Moscow, Russia, 1964.

[41] I. F. Obraztsov, *Variational Methods of Calculations of Thin-Walled Aircraft Space Structures*, Mashinostroenie, Moscow, Russia, 1966.

[42] I. V. Andrianov and A. N. Pasechnik, "Equations of higher-order approximations for investigation of the stressed-deformed state of cylindrical shells," *Doklady Akademii Nauk Ukrainskoj SSR*, vol. 4, pp. 34–38, 1990 (Russian).

[43] L. I. Shkutin, "Introduction of two resolution functions to the equations of nonshallow shells," *Doklady Akademii Nauk SSSR*, vol. 204, no. 4, pp. 809–811, 1972.

[44] Ju. M. Vahromeev and V. M. Kornev, "Boundary-value problems involving a small parameter for ordinary differential equations," *Differential Equations*, vol. 13, no. 7, pp. 803–808, 1977.

[45] M. van Dyke, *Perturbation Methods in Fluid Mechanics*, The Parabolic Press, Stanford, Calif, USA, 1975.

[46] I. V. Andrianov and A. N. Pasechnik, "Method of composite equations in the theory of cylindrical shells," *Soviet Physics. Doklady*, vol. 31, pp. 353–354, 1986.

[47] V. I. Yudovich, "Vibrodynamics and vibrogeometry of the mechanical systems with constraints—III," *Advances in Mechanics*, vol. 3, pp. 130–158, 2006 (Russian).

[48] A. Elhage-Hussein, M. Potier-Ferry, and N. Damil, "A numerical continuation method based on Padé approximants," *International Journal of Solids and Structures*, vol. 37, no. 46-47, pp. 6981–7001, 2000.

[49] R. El Mokhtari, J.-M. Cadou, and M. Potier-Ferry, "A two grid algorithm based on perturbation and homotopy methods," *Comptes Rendus Mecanique*, vol. 330, no. 12, pp. 825–830, 2002.

[50] I. I. Vorovich, *Nonlinear Theory of Shallow Shells*, vol. 133 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1999.

[51] G. I. Marchuk, V. I. Agoshkov, and V. P. Shutyaev, *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems*, CRC Press, Boca Raton, Fla, USA, 1996.

[52] G. L. Litvinov, “Approximate construction of rational approximations and the effect of error autocorrection. Applications,” *Russian Journal of Mathematical Physics*, vol. 1, no. 3, pp. 313–352, 1993.

[53] G. L. Litvinov, “Error autocorrection in rational approximation and interval estimates,” *Central European Journal of Mathematics*, vol. 1, no. 1, pp. 36–60, 2003.

[54] A. A. Dorodnitzyn, *Using of Small Parameter Method for Numerical Solution of Mathematical Physics Equations*, Numerical Methods for Solving of Continuum Mechanics Problems, VZ Akad. Nauk SSSR, Moscow, Russia, 1969.

[55] S. Liao, *Beyond Perturbation. Introduction to the Homotopy Analysis Method*, vol. 2 of *CRC Series: Modern Mechanics and Mathematics*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2004.

[56] I. V. Andrianov, V. V. Danishevs’kyy, and J. Awrejcewicz, “An artificial small perturbation parameter and nonlinear plate vibrations,” *Journal of Sound and Vibration*, vol. 283, no. 3–5, pp. 561–571, 2005.

[57] V. I. Babenko, V. M. Koshelev, and V. Sh. Avedyan, “Experimental investigation of postbuckling equilibrium states of shallow elliptic parabolic-shaped shells under external pressure,” *Doklady NAN Ukraine*, vol. 8, pp. 48–51, 2000 (Russian).

[58] J. Kevorkian and J. D. Cole, *Multiple Scale and Singular Perturbation Methods*, vol. 114 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1996.

[59] I. V. Andrianov and V. V. Danishevs’kyy, “Asymptotic approach for non-linear periodical vibrations of continuous structures,” *Journal of Sound and Vibration*, vol. 249, no. 3, pp. 465–481, 2002.

[60] A. H. Nayfeh, *Perturbation Methods*, Wiley Classics Library, John Wiley & Sons, New York, NY, USA, 2000.

[61] S. A. Nazarov, *Asymptotic Theory of Thin Plates and Rods. Vol. 1. Dimension Reduction and Integral Estimates*, Nauchnaya Kniga, Novosibirsk, Russia, 2001.

[62] J. Sanchez-Hubert and E. Sanchez Palencia, *Coques Elastiques Minces, Propriétes Asymptotiques*, Masson, Paris, France, 1997.

[63] I. Titeux and E. Sanchez-Palencia, “Junction of thin plates,” *European Journal of Mechanics: A/Solids*, vol. 19, no. 3, pp. 377–400, 2000.

[64] V. Maz’ya, S. Nazarov, and B. Plamenevsky, *Asymptotic Theory of Elliptic Boundary Problems in Singular Perturbated Domain*, Birkhäuser, Basel, Switzerland, 2000.

[65] I. V. Andrianov and J. Awrejcewicz, “Edge-localized effects in buckling and vibrations of a shell with free in circumferential direction ends,” *Acta Mechanica*, vol. 173, no. 1–4, pp. 41–47, 2004.

[66] U. Edlund and A. Klarbring, “Analysis of elastic and elastic-plastic adhesive joints using a mathematical programming approach,” *Computer Methods in Applied Mechanics and Engineering*, vol. 78, no. 1, pp. 19–47, 1990.

[67] G. Geymonat, F. Krasucki, and S. Lenci, “Mathematical analysis of a bonded joint with a soft thin adhesive,” *Mathematics and Mechanics of Solids*, vol. 4, no. 2, pp. 201–225, 1999.

[68] A. L. Gol’denveĭzer, J. D. Kaplunov, and E. V. Nolde, “On Timoshenko-Reissner type theories of plates and shells,” *International Journal of Solids and Structures*, vol. 30, no. 5, pp. 675–694, 1993.

[69] S. A. Nazarov, “Localization effects for eigenfunctions near to the edge of a thin domain,” *Mathematica Bohemica*, vol. 127, no. 2, pp. 283–292, 2002.

[70] I. V. Andrianov and J. Awrejcewicz, “On the improved Kirchhoff equation modelling nonlinear vibrations of beams,” *Acta Mechanica*, vol. 186, no. 1–4, pp. 135–139, 2006.

[71] I. V. Andrianov, V. M. Verbonol, and J. Awrejcewicz, “Buckling analysis of discretely stringer-stiffened cylindrical shells,” *International Journal of Mechanical Sciences*, vol. 48, no. 12, pp. 1505–1515, 2006.

[72] D. G. Crighton, "Asymptotics—an indispensable complement to thought, computation and experiment in applied mathematical modeling," in *Proceedings of the 7th European Conference on Mathematics in Industry*, A. Fasano and M. Primicerio, Eds., pp. 3–19, B. G.Teubner, Stuttgart, Germany, 1994.

[73] L. A. Agalovyan, *Asymptotic Theory of Anisotropic Plates and Shells*, Nauka, Moscow, Russia, 1997.

[74] I. V. Andrianov, J. Awrejcewicz, and R. G. Barantsev, "Asymptotic approaches in mechanics: new parameters and procedures," *Applied Mechanics Reviews*, vol. 56, no. 1, pp. 87–109, 2003.

[75] P. G. Ciarlet, *Plates and Junctions in Elastic Multi-Structures: An Asymptotic Analysis*, vol. 14 of *Research in Applied Mathematics*, Masson, Paris, France; Springer, Berlin, Germany, 1990.

[76] R. P. Gilbert and K. Hackl, Eds., *Asymptotic Theories for Plates and Shells*, vol. 319 of *Pitman Research Notes in Mathematics Series*, Longman Scientific & Technical, Harlow, UK; John Wiley & Sons, New York, NY, USA, 1995.

I. V. Andrianov: Institute of General Mechanic, RWTH Aachen, Templergraben 64,
52056 Aachen, Germany
Email address: igor_andrianov@inbox.ru

J. Awrejcewicz: Department of Automatics and Biomechanics, Technical University of Łódź,
1/15 Stefanowski Street, 90-924 Łódź, Poland
Email address: awrejcew@p.lodz.pl

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 62157, 17 pages
doi:10.1155/2007/62157

Research Article

Dynamic Stationary Response of Reinforced Plates by the Boundary Element Method

Luiz Carlos Facundo Sanches, Euclides Mesquita, Renato Pavanello, and Leandro Palermo Jr.

Received 1 October 2006; Revised 7 February 2007; Accepted 26 February 2007

Recommended by José Manoel Balthazar

A direct version of the boundary element method (BEM) is developed to model the stationary dynamic response of reinforced plate structures, such as reinforced panels in buildings, automobiles, and airplanes. The dynamic stationary fundamental solutions of thin plates and plane stress state are used to transform the governing partial differential equations into boundary integral equations (BIEs). Two sets of uncoupled BIEs are formulated, respectively, for the in-plane state (membrane) and for the out-of-plane state (bending). These uncoupled systems are joined to form a macro-element, in which membrane and bending effects are present. The association of these macro-elements is able to simulate thin-walled structures, including reinforced plate structures. In the present formulation, the BIE is discretized by continuous and/or discontinuous linear elements. Four displacement integral equations are written for every boundary node. Modal data, that is, natural frequencies and the corresponding mode shapes of reinforced plates, are obtained from information contained in the frequency response functions (FRFs). A specific example is presented to illustrate the versatility of the proposed methodology. Different configurations of the reinforcements are used to simulate simply supported and clamped boundary conditions for the plate structures. The procedure is validated by comparison with results determined by the finite element method (FEM).

1. Introduction

Reinforced panel systems are widely used in buildings, bridges, ships, aircrafts, and machines. These structural systems are efficient, economical, and readily constructed from common materials. The panels are usually built by the association of plates (or shells) with orthogonally displaced beams, which are the reinforcements. The main advantage of applying these structural elements is the increase of structural rigidity without considerable increase in weight.

The static analysis of the reinforced plate systems has been performed using solution strategies such as methodologies based on energy principles [1], semi-analytical methods [2], or the differential quadrature methods [3]. Also it is possible to model the behavior of these structures by the finite element method (FEM) [4, 5], the boundary element method (BEM) [6–10], or a combination of these numerical methods [11]. A rather limited amount of technical literature is available on the dynamic analysis of stiffened plate systems. On the other hand, a significant research effort is under way in both the academia and the industry to improve the numerical models and to develop new modeling methods for the dynamic analysis [12]. Finite and boundary elements have some limitations to obtain vibration responses at middle and upper frequency ranges due to the necessity of intense mesh refining. The use of very fine meshes in the finite element analysis results in large algebraic systems. An alternative is posed by the BEM. If formulated with the proper auxiliary state, the BEM only requires boundary discretization, leading to considerable smaller algebraic systems.

Direct boundary element subregion formulations based on Kirchhoff's plate theory has been applied to the dynamic analysis of thin-walled structures formed by assembling folded plate models using the so-called static fundamental solution [13, 14]. Assembled plate structures were also analyzed by BEM and comparisons with FEM are given to demonstrate the accuracy of this methodology [15]. Another dynamic analysis of elastic plates reinforced with beams takes into account the resulting in-plane forces and deformations in the plate, as well as the axial forces and deformations in the beam, due to combined response of the system [16]. The method presented in [16] employs the static solution similar to the models described previously. The consequence of these formulations is that the inertia forces lead to domain integrals. In these previous articles, it was necessary to develop a procedure to deal with the domain integral. An alternative way to derive the governing integral equation for the problem is to use a stationary dynamic fundamental solution [17–19]. If this fundamental solution is applied, the resulting integral equation requires only the discretization of the boundary of the single-folded plate being analyzed.

The present paper analyzes the dynamic stationary response of reinforced panels subjected to time harmonic loadings using the BEM. In the proposed methodology, the panels are considered as assembled folded-plate structures [20]. The formulation is built by coupling BE formulations of plate bending and two-dimensional plane stress elasticity. These uncoupled systems are joined to form a macro-element. The plate structure is divided into several regions, and equilibrium and compatibility equations along the interface boundaries are imposed. The boundaries are discretized by means of linear continuous and discontinuous isoparametric elements. Four displacement integral equations are written for every boundary node. The stationary dynamic responses are characterized by modal quantities, that means by eigenfrequencies and eigenvalues. These quantities are obtained by analyzing the numerically synthesized frequency response functions (FRFs) of the reinforced structures. A harmonic force of constant amplitude excites the structure at a given point and the resulting displacement is measured (calculated) at another point. From the resonances or peaks of the FRFs, the operational eigenfrequencies may be determined. The operational eigemodes (vibration mode shapes) are determined by calculating the folded-plate structure displacement field at the determined operational eigenfrequencies. In the present article, an example is presented to illustrate the proposed methodology where different configurations of the reinforcements are used to simulate simply supported and clamped boundary conditions. The implementation is validated by comparison with numerical results determined by FE solutions. The results obtained by the present BEM are shown to be in good agreement with those obtained by the FEM. The proposed scheme may be seen as an accurate methodology to analyze free and forced stationary vibrations of structures assembled by folded plates, like plate structures and reinforced panels. This methodology may be regarded as an extension of the previous article that analyzed the stationary dynamic behavior of frame structures by the BEM [21].

2. Boundary integral formulations

The dynamic equilibrium equations for plane stress and thin plate theory will be presented next with Latin indices taking values {1, 2, and 3} and Greek indices assuming the range {1, 2}. In the plane macro-element formulation, the membrane displacements u_1 and u_2 are in the x_1-x_2 plane. The thin plate transversal displacement w is in the x_3 directions. The equilibrium equations for the dynamic plane stress problem in the domain Ω is given by

$$\sigma_{\alpha\beta,\beta} + \rho F_\alpha = \rho \ddot{u}_\alpha, \tag{2.1}$$

where σ_{ij} represents the stresses components, ρ is the mass density, F_α $(\alpha = 1,2)$ are the body forces components in the x_1-x_2 plane and dots over the quantities indicate differentiation with respect to time.

The equilibrium equations for an infinitesimal thin plate element under a dynamical transverse loading g and in absence of a body forces are given by

$$\begin{gathered} q_{\alpha,\alpha} + g = \rho h \ddot{w}, \\ m_{\alpha\beta,\beta} - q_\alpha = 0, \end{gathered} \tag{2.2}$$

where ρh is the mass density per unit area, h is the thickness, $q_\alpha(q_1, q_2)$ represent the shear forces, $m_{\alpha\alpha}(m_{11}, m_{22})$ represent the bending moments and $m_{\alpha\beta}(m_{12}, m_{21})$ represent the twisting moments.

Now consider the plane element occupying the area Ω, bounded by the contour Γ, in the plane x_1-x_2. The displacement boundary integral equation for the plane stressproblem (membrane) and smooth boundaries is given by

$$\frac{1}{2}\delta_{\alpha\beta}u_{\beta}(P) = \int_{\Gamma} U^{*}_{\alpha\beta}(P,Q)t_{\beta}(Q)d\Gamma(Q) \\ - \int_{\Gamma} T^{*}_{\alpha\beta}(P,Q)u_{\beta}(Q)d\Gamma(Q) + \int\limits_{\Omega} U^{*}_{\alpha\beta}(P,Q)F_{\beta}(Q)d\Omega(Q), \tag{2.3}$$

where $\delta_{\alpha\beta}$ is the *Kronecker* delta; $d\Gamma$ and $d\Omega$ denote boundary and domain differentials, respectively; $u_{\beta}(Q)$ and $t_{\beta}(Q)$ are displacement and traction boundary values associated with a boundary point Q, respectively. The term $U^{*}_{\alpha\beta}(Q,P)$ represents a displacement fundamental solution and may be interpreted as the displacement at point Q in the direction α due to a harmonic unit point force applied at the point P in the direction β. Analogosly, the term $T^{*}_{\alpha\beta}(Q,P)$ represents the traction fundamental solution and may also be interpreted as the traction at point Q in the direction α due to a harmonic unit point load applied at P in the direction β.

Considering that all variables are undergoing a time harmonic displacement with circular frequency ω, the displacement and traction fundamental solutions are given, respectively, by the expressions [22]

$$U^{*}_{\alpha\beta} = \frac{1}{2\pi\rho c_2^2}\left[\psi\delta_{\alpha\beta} - \chi r_{,\alpha}r_{,\beta}\right], \tag{2.4}$$

where

$$\begin{gathered} \psi = K_0(k_2 r) + \frac{1}{k_2 r}\left[K_1(k_2 r) - \frac{c_2}{c_1}K_1(k_1 r)\right], \\ \chi = K_2(k_2 r) - \frac{c_2^2}{c_1^2}K_2(k_1 r), \\ T^{*}_{\alpha\beta} = \frac{1}{2\pi}\left[\left(\frac{d\psi}{dr} - \frac{1}{r}\chi\right)\left(\delta_{\alpha\beta}\frac{\partial r}{\partial n} + r_{,\beta}n_{\alpha}\right) - \frac{2}{r}\chi\left(n_{\beta}r_{,\alpha} - 2r_{,\alpha}r_{,\beta}\frac{\partial r}{\partial n}\right)\right. \\ \left. - 2\frac{d\chi}{dr}r_{,\alpha}r_{,\beta}\frac{\partial r}{\partial n} + \left(\frac{c_1^2}{c_2^2} - 2\right)\left(\frac{d\psi}{dr} - \frac{d\chi}{dr} - \frac{1}{r}\chi\right)r_{,\alpha}n_{\beta}\right], \end{gathered} \tag{2.5}$$

where $\delta_{\alpha\beta}$ is again the *Kronecker* delta, n is the normal vector, K_0 and K_1 are the zero and first-order modified *Bessel* function of second kind, r is the distance between load and displacement point, $k_1 = i(\omega/c_1)$ and $k_2 = i(\omega/c_2)$, $i = \sqrt{-1}$, ω is the circular frequency, $c_1 = (\lambda + 2\mu/\rho)^{1/2}$, $c_2 = (\mu/\rho)^{1/2}$, λ and μ are the *Lamé's* constants which can be written in terms of the *Young* Modulus E and the Poisson ratio ν.

Additionally, the integral equation for the thin plate theory is employed to describe the bending action:

$$\begin{aligned}&\frac{1}{2}\delta(P)w(P)+\int_{\Gamma}\left[V_n^*(P,Q)w(Q)-M_n^*(P,Q)w_{,n}(Q)\right]d\Gamma(Q)+\sum_{k=1}^{N_c}R_{ck}^*(P,c)w_{ck}(P,c)\\&\quad=\int_{\Gamma}\left[w^*(P,Q)V_n(Q)-w_{,n}^*(P,Q)M_n(Q)\right]d\Gamma(Q)\\&\qquad+\sum_{k=1}^{N_c}w_{ck}^*(P,Q)R_{ck}(Q)+\int_{\Omega}w^*(P,q)g(q)d\Omega(q),\end{aligned} \tag{2.6}$$

where $\delta(P)$ is equal to *Kronecker* delta for a smooth boundary, w is the out-of-plane displacement, $w_{,n}$ is the rotation in the direction of outward normal to the boundary Γ, V_n is the equivalent shear, M_n is the bending moment, and R_c is the corner reaction. The classical theory makes use of the equivalent shear (V_n) in boundary integrals and a corner reaction (R_c) at each corner when polygonal plates are considered,

$$V_n=Q_n+\frac{\partial M_{ns}}{\partial s}=-D\left(w_{,\gamma\gamma\alpha}\cdot n_\alpha+(1-\nu)w_{,nss}\right) \tag{2.7}$$

$$R_{ck}=\left(M_{ns}^F-M_{ns}^B\right)_k, \tag{2.8}$$

where Q_n is the shear in the direction of outward normal and M_{ns} is the twisting moment in the direction normal and tangential to the boundary Γ. The expression (2.8) presents the corner reaction (R_c) at corner k as the difference between the twisting moments at the corner neighborhood on the forward side (M_{ns}^F) and the backward side (M_{ns}^B).

Considering again that all variables are undergoing a time harmonic displacement, $u(t)=\hat{u}\exp(i\omega t)$ with circular frequency ω. Under this circumstance, load g and deflections w will also vary harmonically and the fundamental solution for (2.6) has the form [23, 24]

$$w^*=-iC_1J_0(\eta r)+C_1Y_0(\eta r)+C_2K_0(\eta r) \tag{2.9}$$

with

$$\begin{gathered}C_1=\frac{1}{8\eta^2},\qquad C_2=\frac{1}{4\pi\eta^2},\\\eta^4=\frac{\rho h\omega^2}{D}.\end{gathered} \tag{2.10}$$

In (2.9) to (2.10), the flexural rigidity D is equal to $Eh^3/[12(1-\nu^2)]$, E is the *Young* Modulus and ν is the Poisson ratio. The variables J_0 and Y_0 are the zero-order *Bessel* functions of the first and second kind, respectively, K_0 is the zero-order modified *Bessel* function of the second kind. Explicit expressions for derivatives of fundamentals solutions, rotations $w_{,n}^{*}$, moments M_n^*, and shear forces V_n^* are as follows [23]:

$$\begin{aligned}
w_{,n}^{*} &= iC_1\eta J_1(\eta r)\cos\overline{\beta} - \eta\left[C_1Y_1(\eta r) + C_2K_1(\eta r)\right]\cos\overline{\beta},\\
M_n^{*} &= -i\left\{C_1\frac{D}{2}\left[1+\nu+(1-\nu)\cos 2\overline{\beta}\right]\eta^2J_0(\eta r) - C_1D\eta(1-\nu)\frac{J_1(\eta r)}{r}\cos 2\overline{\beta}\right\}\\
&\quad+\frac{D}{2}\left\{\eta^2\left[1+\nu+(1-\nu)\cos 2\overline{\beta}\right]\left[C_1Y_0(\eta r) - C_2K_0(\eta r)\right]\right.\\
&\qquad\left.-2\eta(1-\nu)\frac{1}{r}\left[C_1Y_1(\eta r)+C_2K_1(\eta r)\right]\cos 2\overline{\beta}\right\},\\
V_n^{*} &= iC_1D\left\{J_1(\eta r)\left[\eta^3\cos\overline{\beta} + \frac{\eta^3(1-\nu)}{2}\operatorname{sen}2\overline{\beta}\operatorname{sen}\overline{\beta} + \frac{2\eta(1-\nu)}{r}\left(\frac{\cos 3\overline{\beta}}{r} - \frac{\cos 2\overline{\beta}}{R}\right)\right]\right.\\
&\qquad\left.+(1-\nu)\eta^2J_0(\eta r)\left(\frac{\cos 2\overline{\beta}}{R} - \frac{\cos 3\overline{\beta}}{r}\right)\right\} - D\eta^3\left[C_1Y_1(\eta r) - C_2K_1(\eta r)\right]\cos\overline{\beta}\\
&\quad+D(1-\nu)\left\{\frac{\eta^2}{r}\left[C_1Y_0(\eta r) - C_2K_0(\eta r)\right] - \frac{2\eta}{r^2}\left[C_1Y_1(\eta r)+C_2K_1(\eta r)\right]\cos 3\overline{\beta}\right\}\\
&\quad-D(1-\nu)\left\{\frac{\eta^2}{R}\left[C_1Y_0(\eta r) - C_2K_0(\eta r)\right] - \frac{2\eta}{rR}\left[C_1Y_1(\eta r)+C_2K_1(\eta r)\right]\cos 2\overline{\beta}\right\}\\
&\quad-\frac{D(1-\nu)}{2}\eta^3\left[C_1Y_1(\eta r) - C_2K_1(\eta r)\right]\operatorname{sen}2\overline{\beta}\operatorname{sen}\overline{\beta},
\end{aligned}\tag{2.11}$$

where J_1 and Y_1 are the first-order *Bessel* functions of the first and second kind, respectively; K_1 is the first-order modified *Bessel* function of the second kind and $\overline{\beta}$ the angle formed between r and n.

3. Algebraic formulation of the macro-elements

In this session, the plane macro-element will be assembled by superposition of the membrane and thin plate effects. The plane stress boundary integral equation (2.3) representing the membrane may be discretized leading to the following algebraic system of equations:

$$\begin{bmatrix} H_{11}^m & H_{12}^m \\ H_{21}^m & H_{22}^m \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{bmatrix} G_{11}^m & G_{12}^m \\ G_{21}^m & G_{22}^m \end{bmatrix}\begin{Bmatrix} t_1 \\ t_2 \end{Bmatrix}. \tag{3.1}$$

Analogosly, the BIE (2.6) describing the out-of-plane bending effect (thin plate) may be discretized as follows:

$$\begin{bmatrix} H_{11}^p & H_{12}^p \\ H_{21}^p & H_{22}^p \end{bmatrix}\begin{Bmatrix} w \\ w_{,n} \end{Bmatrix} = \begin{bmatrix} G_{11}^p & G_{12}^p \\ G_{21}^p & G_{22}^p \end{bmatrix}\begin{Bmatrix} V_n \\ M_n \end{Bmatrix}. \tag{3.2}$$

In (3.1) and (3.2), the upper indices m and p on the coefficient matrices H and G stand, respectively, for membrane and plate mechanisms. Furthermore, u_1 and u_2 represent the

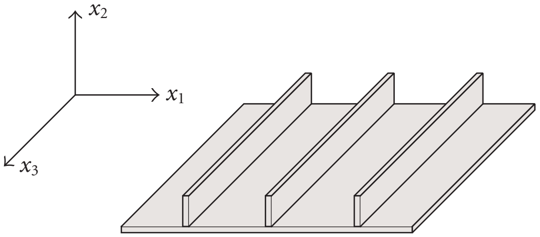


Figure 3.1. Global coordinate system and macro-elements interfaces.

in-plane membrane displacements associated with the in-plane tractions t_1 and t_2. The plate displacement normal to the x_1-x_2 plane is w and its derivative with respect to the boundary normal n is $w_{,n}$. The corresponding generalized forces are the shear forces V_n and the bending moment M_n. Equations (3.1) and (3.2) may be superposed to form the plane macro-element in which membrane and bending mechanisms are uncoupled:

$$\begin{bmatrix} H_{11}^m & H_{12}^m & 0 & 0 \\ H_{21}^m & H_{22}^m & 0 & 0 \\ 0 & 0 & H_{11}^p & H_{12}^p \\ 0 & 0 & H_{21}^p & H_{22}^p \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ w \\ w,n \end{Bmatrix} = \begin{bmatrix} G_{11}^m & G_{12}^m & 0 & 0 \\ G_{21}^m & G_{22}^m & 0 & 0 \\ 0 & 0 & G_{11}^p & G_{12}^p \\ 0 & 0 & G_{21}^p & G_{22}^p \end{bmatrix} \begin{Bmatrix} t_1 \\ t_2 \\ V_n \\ M_n \end{Bmatrix}. \tag{3.3}$$

The interface boundaries between macro-elements must be parallel to a single axis. In a global coordinate system, this axis is called x_3, as shown in Figure 3.1. Figure 3.1 also shows a plate with reinforcements. It can be noticed that the reinforcements are all aligned parallel to the x_3 axis.

The plane macro-element given by (3.3) is written in terms of a local coordinate system. To perform the coupling of distinct macro-elements, as the ones shown in Figure 3.1, it is necessary to transform (3.3) from a local to a global coordinate system. This is done by means of an intermediate coordinate system and a set of two coordinate transformation matrices. After the macro-element equations have been written in terms of the global coordinate system, the assemblage may take place. The vector of generalized displacements and forces may now be subdivided into ones belonging or not to a common interface. For the case of two macro-elements, the individual equations for every macro-element may be written as

$$\begin{gathered} \begin{bmatrix} H_{11}^1 & H_{1i}^1 \\ H_{i1}^1 & H_{ii}^1 \end{bmatrix} \begin{Bmatrix} U_1 \\ U_i^1 \end{Bmatrix} = \begin{bmatrix} G_{11}^1 & G_{1i}^1 \\ G_{i1}^1 & G_{ii}^1 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_i^1 \end{Bmatrix}, \\ \begin{bmatrix} H_{11}^2 & H_{1i}^2 \\ H_{i1}^2 & H_{ii}^2 \end{bmatrix} \begin{Bmatrix} U_2 \\ U_i^2 \end{Bmatrix} = \begin{bmatrix} G_{11}^2 & G_{1i}^2 \\ G_{i1}^2 & G_{ii}^2 \end{bmatrix} \begin{Bmatrix} T_2 \\ T_i^2 \end{Bmatrix}. \end{gathered} \tag{3.4}$$

The coupling of the macro-elements is performed by considering kinematic compatibility and equilibrium at the interface nodes. Considering T the vector of external loads applied at the elements interface, compatibility and equilibrium is given by

$$\begin{gathered} U_i^1 = U_i^2 = U_i, \\ T_i^1 + T_i^2 + T = 0. \end{gathered} \tag{3.5}$$

After (3.5) has been applied to (3.4), the basic system of equation for two coupled macro-elements is given by

$$\begin{bmatrix} H_{11}^1 & H_{1i}^1 & 0 & -G_{1i}^1 & 0 \\ H_{i1}^1 & H_{ii}^1 & 0 & -G_{ii}^1 & 0 \\ 0 & H_{ii}^2 & H_{i2}^2 & 0 & -G_{ii}^2 \\ 0 & H_{2i}^2 & H_{22}^2 & 0 & -G_{2i}^2 \\ 0 & 0 & 0 & I & I \end{bmatrix} \begin{Bmatrix} U_1 \\ U_i \\ U_2 \\ T_i^1 \\ T_i^2 \end{Bmatrix} = \begin{bmatrix} G_{11}^1 & 0 & 0 \\ G_{1i}^1 & 0 & 0 \\ 0 & 0 & G_{i2}^2 \\ 0 & 0 & G_{22}^2 \\ 0 & I & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T \\ T_2 \end{Bmatrix}. \tag{3.6}$$

In (3.6), U_1 and U_2 are generalized displacement vectors (bending and stretching) related to subregions Ω_1 and Ω_2, respectively. T_1 and T_2 are the corresponding generalized forces. The displacement vector U_i and the corresponding forces vector T_i stand for the values at the interface; T_i^1 and T_i^2 represent forces vectors at the interfaces for each one of the macro-elements.

4. BEM formulations

In this paper, the macro-elements coupled by (3.6) were discretized by rectilinear boundary elements described by linear shape functions. Considering B_1 and B_2 the initial and final coordinates of the elements, the element geometry may be expressed in terms of intrinsic coordinates, ς:

$$b(\varsigma) = B_1 \frac{1-\varsigma}{2} + B_2 \frac{1+\varsigma}{2}. \tag{4.1}$$

This same interpolation is used for the field variables of the boundary elements possessing no corners, leading to an isoparametrical formulation. For elements with corners, the field variables were discretized by discontinuous elements. The corner nodes were displaced towards the interior by one-fourth of the element length ($0.25\,L_e$). Four integral equations were written for every boundary node. The collocation points were placed outside the plane element (macro-element) domains. When collocation point P is placed outside the plate domain ($P \notin \Omega$), the integration free-term disappears, $\delta(P) = 0$. Moreover, the corner reactions R_{ck} can be written in terms of neighbor node rotations using a finite difference scheme. Although this is the correct way to treat corner reactions, in the present implementation these terms were neglected.

A final algebraic system $[A]\{X\} = \{B\}$ is obtained once the equations are assembled and the prescribed boundary conditions applied. The solution of this system, the vector X, contains all unknown boundary quantities. The system matrix $[A(\omega)]$ contains frequency dependent terms. After the vector X is determined, the displacement at the assembled folded plate domains may be readily obtained by the nonsingular integrations indicated in (2.3) and (2.6).

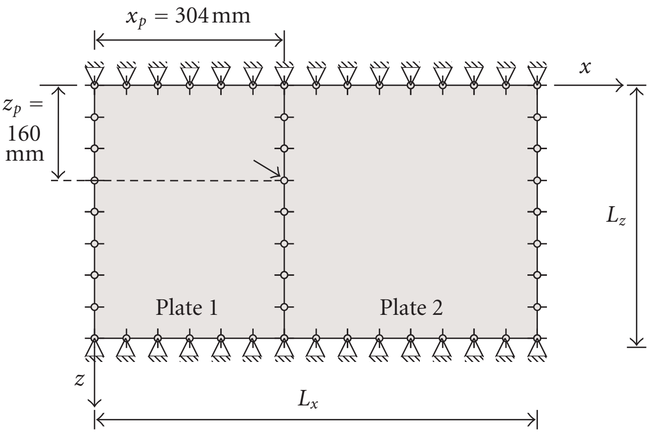


Figure 5.1. Two plates, S-F-S-F, no reinforcements.

5. Numerical analysis

This session applies the previously described strategy to analyze reinforced panels. The strategy is simple. Consider two joined rectangular plates, simply supported (S) in two edges $z = 0$ and $z = L_z$, and freely supported (F) at the remaining edges, $x = 0$ and $x = L_x$, as shown in Figure 5.1. The plates are excited by a concentrated force applied at the point with coordinates x_p and z_p. The frequency of excitation is continuously changed within a preestablished range. The displacement response at some point of the plates to this frequency dependent excitation is the so-called frequency response function (FRF).

In the sequence, reinforcements are placed at the boundaries $x = 0$ and $x = L_x$. If the reinforcement is very thin but very high in the y-direction, then the bending effect of the support is very small compared to the membrane effect. This should simulate a simply supported (S) boundary condition. On the other hand, if the plate thickness is increased, then the clamped (C) boundary condition should be simulated. The validation strategy is composed of these three steps. In the first step, the FRF of S-F-S-F plate is determined and the natural frequencies of the present methodology compared to results from a finite element (FE) analysis. In the second step, high but thin reinforcements are placed at the originally free boundaries, giving rise to a model that simulates completely simply supported plates S-S-S-S. Again the operational eigenfrequencies are obtained from the FRF. Comparisons are also made with the FE solution. Finally, thicker reinforcements are placed at the free boundaries, simulating the clamped (C) boundary condition. The operational eigenfrequencies for this C-S-C-S plate are compared to the FE results. This strategy is sketched in Figure 5.2.

Take initially the two plates loaded by a unit harmonic normal excitation on the interface between the two plates at distances $x_1 = 304\,\text{mm}$ and $x_3 = 160\,\text{mm}$ (X-Z plane), as shown in Figure 5.1. The two plates are assembled and are simply supported (S) at their edges $z = 0$ and $z = L_z$, andfree (F) at their boundaries $x = 0$ and $x = L_x$. Each

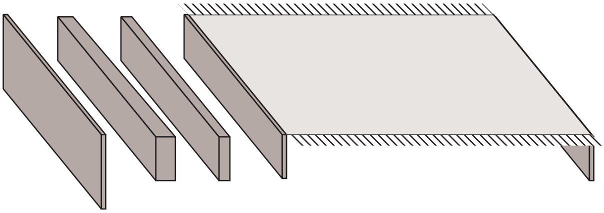


Figure 5.2. Reinforcements as replacements for distinct boundary conditions.

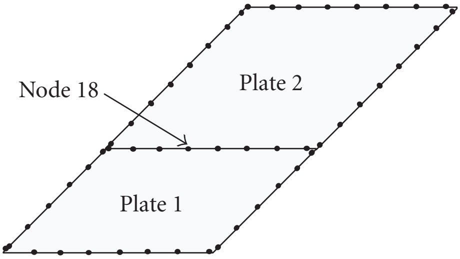


Figure 5.3. Example of the discretization for the two plates.

Table 5.1. First six natural frequencies of the two SFSF plates [Hz].

Method	Mesh	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
BEM	1	6056	7618	12 110	19 922	24 415	26 172
	2	6056	7618	12 305	20 313	24 415	26 172
FEM	1	5951	7459	12 013	19 709	23 878	25 575

assembled plate is made of same constitutive properties with Young's modulus $E = 6.9 \times 10^{10}\,\text{kN/m}^2$, thickness $h = 4\,\text{mm}$, density $\rho = 2700\,\text{Kg/m}^3$, length $L_x = 704\,\text{mm}$, width $L_z = 400\,\text{mm}$, and Poisson ratio $\nu = 0.3$.

Computations by the BEM are carried out for the following boundary discretization (two macro-elements) using linear micro-elements: Mesh 1 : 18 and 20 boundary elements per macro-element (plate 1 and 2, resp.) and Mesh 2 : 28 and 30 boundary elements per macro-element (plate 1 and 2, resp.). An example of the discretization of boundary is shown in Figure 5.3.

Figure 5.4 shows the $\text{FRF}_{18\text{-}18}$ for the first BEM mesh. The $\text{FRF}_{18\text{-}18}$ is obtained by exciting node 18 (see Figure 5.3) and measuring the response at the same node. In this FRF, the resonances and antiresonances can be clearly recognized. The system operational eigenfrequencies (natural frequencies) are determined from the frequencies at which resonances in the FRF occur.

The values of the first six eigenfrequencies taken from the FRF of the two assembled plates given in Figure 5.4 are reproduced in Table 5.1. These valuesare compared with

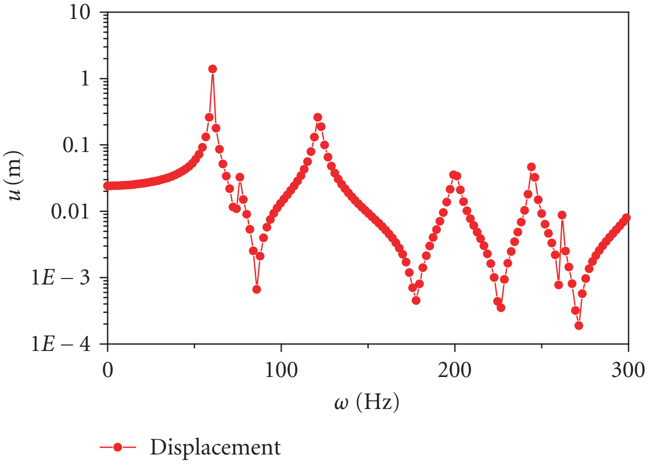


Figure 5.4. $\mathrm{FRF}_{18\text{-}18}$ for the first BE discretization of the two SFSF plates.

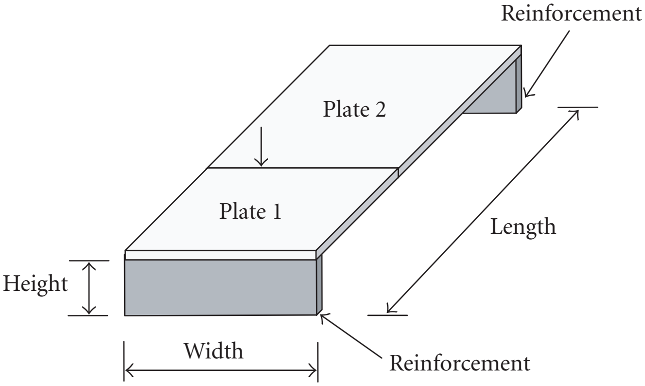


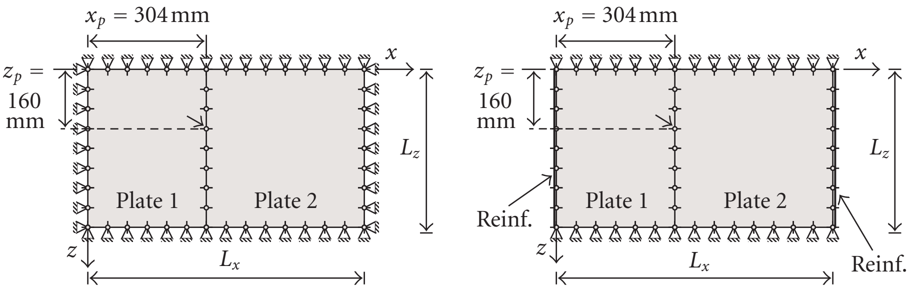
Figure 5.5. Reinforced panel structure subjected to concentrated time-harmonic load.

similar discretization of the FEM by ANSYS® using SHELL63® elements. The discretization of de FEM by ANSYS® consisted of the 18 × 30 finite elements.

Now the reinforcements are included in the originally freely supported boundary conditions (F). The resulting reinforced panel structure is shown in Figure 5.5. In the reinforcements, only the central nodes of the sides are submitted to simply supported boundary conditions. The remaining nodes are free.

To simulate a simply supported boundary condition (S) as shown in Figure 5.6, the reinforcement is a thin and high macro-element. In this case, the plates and reinforcements are made of same material properties described in the previous example.

The reinforced panel is discretized with 4 macro-elements, 2 elements for the plates and 2 elements for the reinforcements (see Figure 5.7). Two BE meshes are used to perform the calculations. In the first mesh, all 4 macro-elements are discretized with 18 × 20



(a) Boundary conditions (S-S-S-S)

(b) Equivalent thin reinforcements

Figure 5.6. Schematic illustration of the reinforced panel structure.

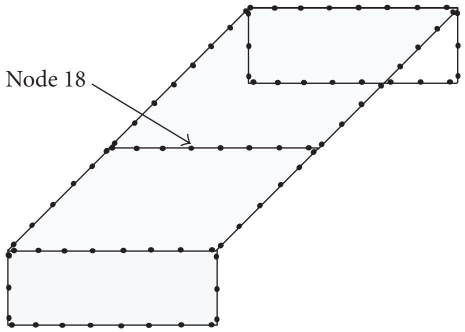


Figure 5.7. BEM for the reinforced panel structure (Mesh 1).

Table 5.2. First six natural frequencies of the SSSS structure [Hz].

Method	Mesh	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
BEM	1	8203	14063	24024	26563	32617	37891
	2	8203	14 063	24 024	26 563	32 617	37 891
FEM	1	7911	13 638	23 174	25 771	31 292	36 528

linear elements. In the second mesh, the 4 macro-elements have been discretized with 28×30 linear elements. The geometric properties of the reinforcements are thickness 0.4 mm, height 400 mm, and width 400 mm.

The $\text{FRF}_{18\text{-}18}$, of the reinforced panel structure is shown in Figure 5.8, for the discretization (Mesh 1) mentioned above.

The values of the first six eigenfrequencies of the reinforced panel structure are reproduced in Table 5.2. These values are compared with results obtained by the FEM commercial code ANSYS® using 18×30 SHELL63® elements.

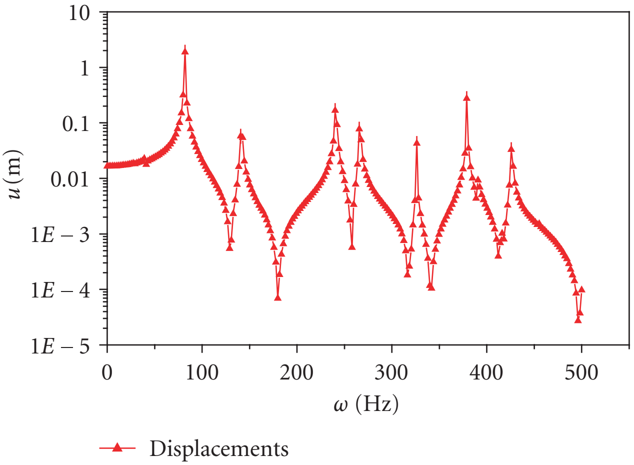
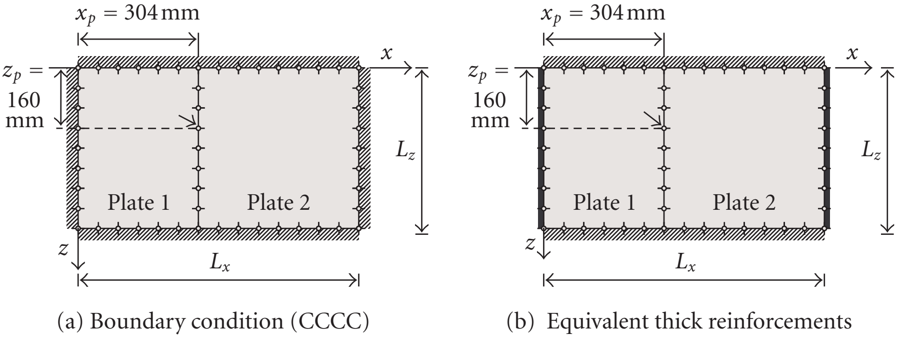


Figure 5.8. FRF$_{18\text{-}18}$ for the first BEM discretization of the reinforced SSSS panel.



(a) Boundary condition (CCCC)

(b) Equivalent thick reinforcements

Figure 5.9. Scheme to reproduce clamped boundary conditions on the plate by thick reinforcements.

Let us now consider the plate structure with its entire contour under clamped (C) boundary conditions (Figure 5.9(a)). The intention is to simulate a clamped boundary condition of the two plates on the boundaries $z = 0$ and $z = L_z$ using a thick reinforcement. The idea was to introduce thick reinforcements to increase the bending rigidity. This strategy is illustrated in Figure 5.9(b). In the reinforcement, only the central nodes of the sides are considered clamped. The others are considered free. In this case, the plates are made of same material properties and geometry described in the previous example. However, the thickness of the two reinforcements is increased to 40 mm. This thickness is increased by a factor 100, compared to the previous case.

The FRF$_{18\text{-}18}$, of the reinforced (campled) panel structure discretized by BE is shown in Figure 5.10. The discretization utilized is the same as the previous case (mesh 1).

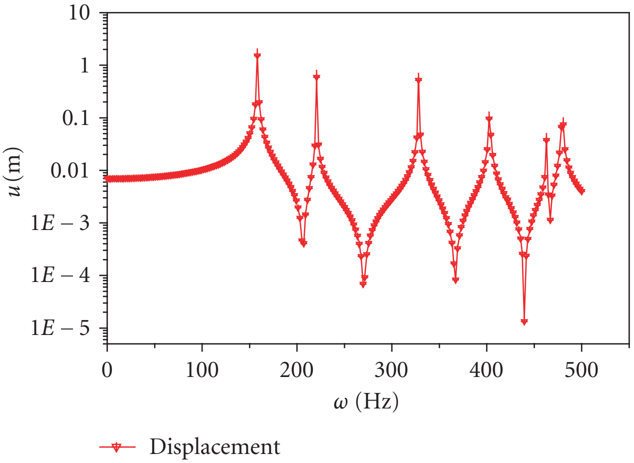
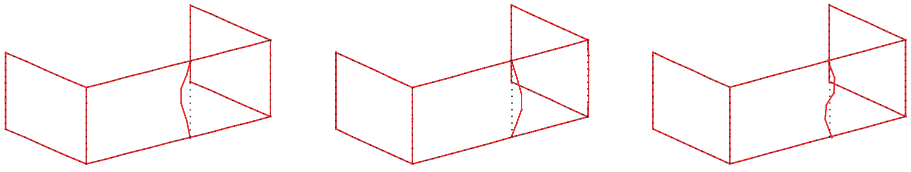


Figure 5.10. $\mathrm{FRF}_{18\text{-}18}$ for the first BEM discretization of the reinforced CCCC panel.

Table 5.3. First six natural frequencies of the CCCC structure [Hz].

Method	Mesh	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
BEM	1	15 821	22 071	32 813	40 235	46 289	48 046
FEM	1	15 456	21 485	32 019	39 328	45 023	46 860



(a) $\omega_1 = 15821$ Hz

(b) $\omega_2 = 22071$ Hz and $\omega_3 = 32813$ Hz

(c) $\omega_4 = 40235$ Hz

Figure 5.11. Four lower operational eigenmodes for the CCCC structure.

The values of the first six (operational) eigenfrequencies are reproduced in Table 5.3. The eigenvalues obtained by the commercial FEM code ANSYS® using 18 × 30 SHELL63® elements are also given in Table 5.3.

The other modal quantity necessary to characterize the stationary dynamic behavior of the reinforced panel structure is given by the eigenmodes or the natural modes of vibration. For the last case (clamped bc), the operational eigenmodes are obtained by calculating the displacement field at boundary of the structure at each resonance frequency present in the FRF.

Figure 5.11 shows the boundary displacements corresponding to the first four eigenmodes of the excited structure. In this case, all external boundary nodes of the structure are campled. It should be noticed that the second and third modes present the same boundary displacement.

6. Concluding remarks

An implementation of the direct version of the boundary element method has been presented to analyze the stationary dynamic behavior of the reinforced panel structures. The dynamic stationary fundamental solution has been used to transform the differential equations governing the thin plate and membrane behavior into boundary-only integral equations. The proposed scheme is used, exemplarily, to obtain modal data, that is, operational eigenfrequencies and eigenmodes of the assembled plates and reinforced panel structures with different boundary conditions. The formulation was shown to be capable of modelling plates subjected to varied boundary conditions and out-of-plane loadings. Frequency response functions may be determined for every boundary or domain point of the structure. In the reported examples, the FRF of a node on an interface boundary is used to recover eigenfrequencies. The eigenfrequencies are determined from the resonances of the FRF. At these resonance frequencies, the displacement fields of the structure furnish the operational eigenmodes. The presented results agree well with numerical solutions obtained by a FEM commercial code. The proposed scheme may be seen as an accurate methodology to analyze free and forced stationary vibrations of structures assembled by folded plates, plate structures, and also reinforced panels which only require the discretization of the folded plate boundary. The simplicity of the BE mesh generation presents some advantages over other domain methods.

Acknowledgment

The research leading to this article has been supported by FAPESP. This is gratefully acknowledged.

References

[1] A. R. Kukreti and E. Cheraghi, “Analysis procedure for stiffened plate systems using an energy approach,” *Computers & Structures*, vol. 46, no. 4, pp. 649–657, 1993.

[2] M. Mukhopadhyay, “Stiffened plates in bending,” *Computers & Structures*, vol. 50, no. 4, pp. 541–548, 1994.

[3] Z. A. Siddiqi and A. R. Kukreti, “Analysis of eccentrically stiffened plates with mixed boundary conditions using differential quadrature method,” *Applied Mathematical Modelling*, vol. 22, no. 4-5, pp. 251–275, 1998.

[4] A. Deb and M. Booton, “Finite element models for stiffened plates under transverse loading,” *Computers & Structures*, vol. 28, no. 3, pp. 361–372, 1988.

[5] G. S. Palani, N. R. Iyer, and T. V. S. R. Appa Rao, “An efficient finite element model for static and vibration analysis of eccentrically stiffened plates/shells,” *Computers & Structures*, vol. 43, no. 4, pp. 651–661, 1992.

[6] M. Tanaka and A. N. Bercin, “A boundary element method applied to the elastic bending problem of stiffened plates,” in *Proceedings of the 19th International Conference on Boundary Element Methods (BEM ’97)*, pp. 203–212, Rome, Italy, September 1997.

[7] E. J. Sapountzakis and J. T. Katsikadelis, "Analysis of plates reinforced with beams," *Computational Mechanics*, vol. 26, no. 1, pp. 66–74, 2000.

[8] M. Tanaka, T. Matsumoto, and S. Oida, "A boundary element method applied to the elastostatic bending problem of beam-stiffened plates," *Engineering Analysis with Boundary Elements*, vol. 24, no. 10, pp. 751–758, 2000.

[9] P. H. Wen, M. H. Aliabadi, and A. Young, "Boundary element analysis of shear deformable stiffened plates," *Engineering Analysis with Boundary Elements*, vol. 26, no. 6, pp. 511–520, 2002.

[10] L. de Oliveira Neto and J. B. de Paiva, "A special BEM for elastostatic analysis of building floor slabs on columns," *Computers & Structures*, vol. 81, no. 6, pp. 359–372, 2003.

[11] S. F. Ng, M. S. Cheung, and T. Xu, "A combined boundary element and finite element solution of slab and slab-on-girder bridges," *Computers & Structures*, vol. 37, no. 6, pp. 1069–1075, 1990.

[12] J. R. F. Arruda and K. M. Ahmida, "The structural dynamics mid-frequency challenge: bridging the gap between FEA and SEA," in *Proceedings of the 10th International Symposium on Dynamic Problems of Mechanics (DINAME '03)*, pp. 159–164, ABCM, Ubatuba, Brazil, March 2003.

[13] M. Tanaka, K. Yamagiwa, K. Miyazaki, and T. Ueda, "Free vibration analysis of elastic plate structures by boundary element method," *Engineering Analysis*, vol. 5, no. 4, pp. 182–188, 1988.

[14] M. Tanaka, T. Matsumoto, and A. Shiozaki, "Application of boundary-domain element method to the free vibration problem of plate structures," *Computers & Structures*, vol. 66, no. 6, pp. 725–735, 1998.

[15] T. Dirgantara and M. H. Aliabadi, "Boundary element analysis of assembled plate structures," *Communications in Numerical Methods in Engineering*, vol. 17, no. 10, pp. 749–760, 2001.

[16] E. J. Sapountzakis and J. T. Katsikadelis, "Dynamic analysis of elastic plates reinforced with beams of doubly-symmetrical cross section," *Computational Mechanics*, vol. 23, no. 5, pp. 430–439, 1999.

[17] D. E. Beskos, "Boundary element methods in dynamic analysis," *Applied Mechanics Reviews*, vol. 40, no. 1, pp. 1–23, 1987.

[18] D. E. Beskos, "Dynamic analysis of plates," in *Boundary Element Analysis of Plates and Shells*, D. E. Beskos, Ed., pp. 35–92, Springer, Berlin, Germany, 1991.

[19] D. E. Beskos, "Boundary element methods in dynamic analysis—part II (1986–1996)," *Applied Mechanics Reviews*, vol. 50, no. 3, pp. 149–197, 1997.

[20] L. C. F. Sanches, E. Mesquita, and L. Palermo Jr., "The dynamic of thin-walled structures by the boundary element method," in *Proceedings of the 25th CILAMCE: 25nd Iberian Latin-American Congress on Computational Methods in Engineering*, Campinas, Brazil, 2004.

[21] E. Mesquita, S. F. A. Barretto, and R. Pavanello, "Dynamic behavior of frame structures by boundary integral procedures," *Engineering Analysis with Boundary Elements*, vol. 24, no. 5, pp. 399–406, 2000.

[22] J. Dominguez, *Boundary Elements in Dynamics*, International Series on Computational Engineering, Computational Mechanics, Southampton; Elsevier Applied Science, London, UK, 1993.

[23] J. Vivoli and P. Filippi, "Eigenfrequencies of thin plates and layer potentials," *Journal of the Acoustical Society of America*, vol. 55, no. 3, pp. 562–567, 1974.

[24] Y. Niwa, S. Kobayashi, and M. Kitahara, “Eigenfrequencies analysis of a plate by the integral equation method,” *Theoretical and Applied Mechanics*, vol. 29, pp. 287–307, 1981.

Luiz Carlos Facundo Sanches: Department of Mathematics, Paulista State University,
Al. Rio de Janeiro s/n, 15385-000 Ilha Solteira, SP, Brazil
Email address: luiz@mat.feis.unesp.br

Euclides Mesquita: Department of Computational Mechanics, State University of Campinas,
Rua Mendeleiev s/n, 13083-970 Campinas, SP, Brazil
Email address: euclides@fem.unicamp.br

Renato Pavanello: Department of Computational Mechanics, State University of Campinas,
Rua Mendeleiev s/n, 13083-970 Campinas, SP, Brazil
Email address: pava@fem.unicamp.br

Leandro Palermo Jr.: Department of Structures, State University of Campinas,
Avenida Albert Einstein 951, 13083-970 Campinas, SP, Brazil
Email address: leandro@fec.unicamp.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 49781, 8 pages
doi:10.1155/2007/49781

Research Article

Simple Orbit Determination Using GPS Based on a Least-Squares Algorithm Employing Sequential Givens Rotations

Rodolpho Vilhena de Moraes, Aurea Aparecida da Silva, and Helio Koiti Kuga

Received 21 September 2006; Revised 4 June 2007; Accepted 13 July 2007

Recommended by José Manoel Balthazar

A low-cost computer procedure to determine the orbit of an artificial satellite by using short arc data from an onboard GPS receiver is proposed. Pseudoranges are used as measurements to estimate the orbit via recursive least squares method. The algorithm applies orthogonal Givens rotations for solving recursive and sequential orbit determination problems. To assess the procedure, it was applied to the TOPEX/POSEIDON satellite for data batches of one orbital period (approximately two hours), and force modelling, due to the full JGM-2 gravity field model, was considered. When compared with the reference Precision Orbit Ephemeris (POE) of JPL/NASA, the results have indicated that precision better than 9 m is easily obtained, even when short batches of data are used.

1. Introduction

The Global Positioning System (GPS) provides a powerful and quick means to compute orbits for low Earth artificial satellites. Theoretically, four GPS satellites, simultaneously tracked, are enough for geometrical positioning of an artificial satellite carrying an onboard GPS receiver. Dynamical orbit determination is a nonlinear problem where the perturbing factors are not easily modelled. The GPS satellites transmit signals such that accurate measurements of distances are performed based on the comparison between received signals and template signals generated by the receiver [1]. Through a GPS receiver onboard of an artificial satellite it is possible to obtain such measurements (pseudoranges) that can be processed to estimate a state vector (e.g., position and velocity vectors and parameters referred to the receiver clock bias). Using our knowledge about the dynamics of the motion and, at the same time, assuming statistics for the measurement errors, the state vector can be computed based on a set of observations. To this, the differences between the observed and modelled observations are minimised according to the least squares criterion [2–4].

The aim of this work is to determine the orbit of an artificial satellite carrying a GPS receiver, by a recursive least squares method based on L1-code GPS satellite-to-satellite tracking observations, and to make use of Givens rotations [5] in order to solve the problem in a recursive way. The Givens rotations algorithm is as stable as other orthogonalization algorithms (such as Cholesky decomposition, Gram-Schmidt, or Householder) and allows the processing for each observation epoch, avoiding the storage of large matrices. Advantages of the use of Givens rotations were already point out in the seventies of the last century in an improvement of a computational form of the discrete Kalman filter [6]. Givens rotations have been successively used at Center for Space Research, University of Texas at Austin, Austin, Tex, USA [7].

Thus, this procedure for near real time positioning has been considered to be used in the next two Chinese-Brazilian remote sensing satellites CBERS-3 and CBERS-4 scheduled to fly in 2008–2010, according to a Brazil-China protocol agreement. These missions carry a main CCD camera payload with 20 m resolution and, therefore, a simple orbit determination scheme with 10 m accuracy is a desirable requirement for the image processing users. Due to onboard memory limitations it is likely that only a short batch of GPS data (one revolution) around the image scene will be available.

Initially, a simple dynamical model (geopotential perturbations only) was used to preliminary assess the procedure, by using the TOPEX/POSEIDON satellite as a test bed. However, the final purpose is to evolve to applications of the procedure considering real and suitable models for the perturbations well suited for the orbital geometry and physical characteristics of the CBERS mission satellites [8].

2. Dynamical model

The problem of dynamical orbit determination is essentially nonlinear. The orbital motion is described in an inertial frame by a set of ordinary differential equations:

$$\ddot{\mathbf{r}} = -\mu\frac{\mathbf{r}}{r^3} + \mathbf{P}, \tag{2.1}$$

where $\mathbf{r} = (x, y, z)$ is the position vector, μ is the gravity parameter, and $\mathbf{P}$ stands for modelled perturbations. Formally, we include an additional equation $\dot{\mathbf{b}} = \mathbf{0}$ with $\mathbf{b} = [b_0, b_1, b_2]$ solve-for parameters referred to the receiver clock bias, (bias, drift, and drift rate). Thus, the state vector to be estimated is defined by $\mathbf{x} \equiv [\mathbf{r} \quad \mathbf{v} \quad \mathbf{b}]^t$ with $\mathbf{v}$ being the velocity vector. The state transition matrix which relates the state between times t_k and t_{k+1} can be computed by $\dot{\mathbf{\Phi}}(t, t_k) = \mathbf{F}(\mathbf{x}, t)\mathbf{\Phi}(t, t_k)$ with initial condition $\mathbf{\Phi}(t_k, t_k) = \mathbf{I}$ [3, 9, 10]. The Jacobian matrix $\mathbf{F}(\mathbf{x},t)$ contains partial derivatives of the differential equations of motion. In general, the transition matrix is numerically integrated together with the orbit. In our case, where only the geopotential perturbation is modelled, the matrix

F can be written as

$$\mathbf{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{A}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \end{bmatrix}, \tag{2.2}$$

where **f** is the acceleration acting on the satellite, $\mathbf{0}_{3\times 3}$ is a 3 × 3 matrix of zeros, $\mathbf{I}_{3\times 3}$ is a 3 × 3 identity matrix and, $\mathbf{A}_{3\times 3}$ is the 3 × 3 matrix of gravity gradient 3 × 3 given by $\mathbf{A}_{3\times 3} = \partial \dot{\mathbf{v}}/\partial \mathbf{r}$. In this work, the modelled force includes the geopotential, taking into account the spherical harmonic coefficients up to 50th order and degree of the JGM-2 model [11]. The geopotential acceleration and its gradient matrix $\mathbf{A}_{3\times 3}$ were computed using basically Pine's universal recursive formulation [12], however we use the improved numerical recursions as stated in Lundberg and Schutz [13]. For assessing the proposed approach, we decided to consider only the geopotential perturbation, applied to TOPEX/Poseidon satellite for which results are broadly available for comparison purposes. Considering that the aimed application would be the CBERS satellite, the final procedure, as developed for the CBERS mission, will include also other relevant perturbations such as radiation pressure forces, lunisolar attraction, and atmospheric drag.

3. Measurement model

The general nonlinear equation representing the scalar model of observations at epoch. is given by $\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, t) + \mathbf{v}_k$, where $\mathbf{y}_k$ is the vector of m observations, $\mathbf{h}_k(\mathbf{x}_k)$ is the m-dimensional nonlinear vector function of the state vector $\mathbf{x}_k$, $\mathbf{v}_k$ is the m-dimensional vector of observation errors modelled as white noise. Pseudoranges are measurements of the distance between the GPS satellites and the receiver's antenna, referring to the epochs of emission and reception of the signals. Pseudorange is a typical type of measurement in an orbit determination process using GPS, and can be written as [1] $P_i = \rho_i + c(dt - dT_i) + D_{\text{ion}} + D_{\text{trop}} + \nu$, where P_i is the pseudorange measured by the user with respect to the ith GPS satellite, ρ_i is the geometric distance, c is the speed of light, dt is the user clock offset, dT_i is the ith GPS satellite clock offset, D_{ion} are ionospheric delays (not considered in this work), D_{trop} are tropospheric delays (not applicable), and ν denotes random measurement noise. Tests computing the first-order ionosphere effect using P_1 and P_2 measurements, were performed. They show that, given the aimed for precision in this work (order of 10 m), ionospheric delays can be neglected. The user clock offset is modelled as $cdt = b_0 + b_1\Delta t + b_2\Delta t^2$, where $\mathbf{b} = [b_0, b_1, b_2]$ is a vector of parameters to be estimated representing the clock bias, bias rate, and rate of bias rate. To estimate the state vector parameters using the least squares algorithm (e.g., [14]), the measurement equation is linearized and the sensitivity matrix of partial derivatives with respect to the state vector is derived.

4. Recursive least squares method based on givens rotations

The numerical stability of a state vector estimator based on the least square method should be assured. To validate the method for orbit determination using a recursive approach, namely the recursive least squares method, one is tempted to use the Kalman form [2] in straightforward way. Extensive analyses of such algorithms have shown that they are unstable numerically and sensitive to error accumulation [2, 3]. On the other hand, the nonrecursive solution of normal equations requires a matrix inversion, another source of numerical errors, which should be avoided. In order to overcome such problems, several methods can be found in the literature using orthogonal transformations. They yield best numerical performance with respect to problems due to error propagation or uncertainty in the information. The main goal of applying orthogonal transformations to matrices and vectors in the least squares algorithms is the substitution of the conventional algebra based on matrix inversions by a method that is more robust and less prone to numerical errors. Among the several existing orthogonal transformations (e.g., Givens, Gram-Schmidt, Householder), the Givens rotations algorithm is the most adequate transformation to selectively annihilate elements of a matrix, making it easy to implement the least squares method in a recursive way. For a batch least squares method, Gram-Schmidt or Householder transformations are in general computationally less costly than Givens rotations. However, from a recursive formulation point of view, Givens rotations algorithm allows us to implement the processing of one measurement at a time, avoiding the need of large matrices for storage (see, e.g., [6, 9, 15]).

Indeed, the direct solution of the normal equations can be quite sensitive to small errors in a sensitivity matrix $\mathbf{H}$ that are inevitable when forming the product $(\mathbf{H}^T\mathbf{H})$, with a limited numerical machine accuracy. In order to avoid the normal equations of the form $(\mathbf{H}^T\mathbf{H})x = \mathbf{H}^T y$, the minimization of the least squares cost function using $\mathbf{Q}$-R factorisation [6, 16] is recommended:

$$\mathbf{H}_{m\times n} = \mathbf{Q}_{m\times m}\begin{pmatrix}\mathbf{R}_{n\times n}\\ \mathbf{0}_{(m-n)\times n}\end{pmatrix}, \tag{4.1}$$

where the matrix $\mathbf{H}$ is factored into an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$.

Methods for performing the $\mathbf{Q}$-$\mathbf{R}$ factorisation have been proposed [5, 6, 16, 17], involving orthogonal transformations that successively annihilate the sub-diagonal elements of $\mathbf{H}$. Nevertheless we concentrated our approach on the Givens rotations method for the reasons stated above.

The complete transformation can be written as the product of the sequence of orthogonal rotations required to transform Matrix $\mathbf{H}$ into the form of (4.1):

$$\mathbf{Q}^T = (\mathbf{U}_m\mathbf{U}_{m-1}\cdots\mathbf{U}_3\mathbf{U}_2) \tag{4.2}$$

denoting the sequence of orthogonal rotations required to put matrix $\mathbf{H}$ into the form of (4.1).

Orthogonal transformations of matrices play a considerable role in the numerical computations of the least squares problems [18]. In fact, the Euclidian norm of a vector does not change, the same accuracy is obtained with single computer floating point arithmetic that otherwise would require double precision and the problem is solved in a numerically robust fashion.

Several programming tests were carried out before choosing the Givens rotation algorithm for implementation in the recursive least squares method. The Kalman type algorithm led to truncation errors for long batches of data. The Gram-Schmidt and Householder method were not as well suited for recursive implementation. Thus the Givens algorithm was finally selected. The final combination of a simple orbit model and robust numerical properties of the Givens orthogonal transformation resulted in a very compact computer program with small requirements of storage, paving the way for further improvements in terms of more complex models for orbit and GPS measurements.

5. Results

Actual TOPEX/POSEIDON (T/P) satellite flight data are used to validate the proposed method. The T/P satellite was launched on August 10, 1992, and it orbits the Earth at an altitude of 1336 km in an orbit with an inclination of 66°, with near-zero eccentricity, and an orbital period of about 1.87 hours. It has a GPS receiver onboard as experimental equipment to verify several precision methods for orbit determination. The receiver can track up to 6 GPS satellites simultaneously on two frequencies if antispoofing is inactive. RINEX format for T/P observation data, GPS group data, and navigation messages are provided by International GNSS Service (IGS), NASA, DC, USA [19], Goddard Space Flight Center (GSFC), NASA, DC, USA, and Jet Propulsion Laboratory (JPL), NASA, DC, USA. Here, the following files were used [20].

(1) T/P observation files: Pseudorange codes measured at two frequencies in GPS time steps of 10 seconds, and made available by the GPS data processing facility of JPL in RINEX format.
(2) Files with the Precise Orbit Ephemeris (POE): generated by JPL in one minute UTC time steps in inertial true of date coordinates.
(3) GPS navigation message files in RINEX format made available by the Crustal Dynamics Data Information System (CDDIS) of the GSFC.

In this work, the estimated position and velocities were compared with the T/P Precise Orbit Ephemeris (POE) generated by JPL which provides position estimates with an estimated precision of 15 cm or better. The test conditions for the investigated problem are the following:

(1) actual topex/poseidon pseudorange measurement dataset collected by the onboard GPS receiver on November 18 1993;
(2) geopotential perturbations taking into account the spherical harmonic coefficients up to 50th order and degree of JGM-2 model;
(3) GPS antenna offset neglected, as its effects is about 5 m;
(4) pseudorange measurements at frequency L_1 (Code);
(5) recursive least squares method with Givens rotations algorithm;
(6) short dataset of 2 hours (about 1 orbital period).

The estimates obtained by our method were compared with the POE reference and produced position error of better than 9 m, starting with the JPL/POE values at first epoch, but clock parameters at the 100 m error level.

Figure 5.1 shows the position error components compared with the JPL/POE reference orbit and Figure 5.2 shows the pseudorange residuals versus time, where the final

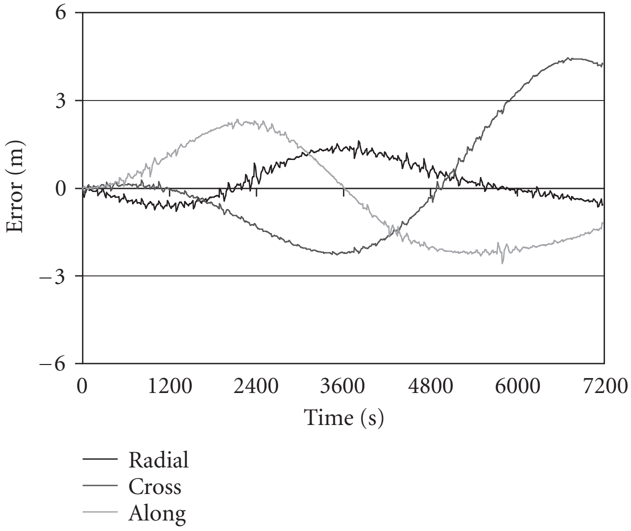


Figure 5.1. Position differences to the POE/JPL orbit.

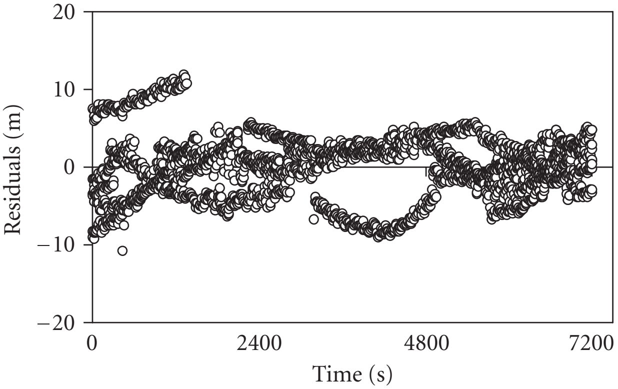


Figure 5.2. Pseudorange residuals.

residuals are clearly unbiased with respect to the apriori residuals. These are the plots of pseudorange data residuals (observed minus computed) for different satellites, seen at different times around the orbit. The apriori covariance matrix was set to 1000 m and 10 m/s for each component of the position and velocity vectors, respectively, and 5 m standard deviation was adopted for the measurement error.

6. Conclusions

The aim of this work was to assess a recursive least squares method based on Givens rotations to determine the orbit of an artificial satellite using an onboard GPS receiver. A short batch of a two hours pseudorange measurement dataset (about one orbital period), collected by the GPS receiver onboard of the T/P satellite, was used. The modelled forces included the geopotential taking into account the spherical harmonic coefficients up to 50th order and degree of the JGM-2 model. Results of this work, using T/P dataset from November 18, 1993, compared against the postprocessed GPS ephemeris POE/JPL, has demonstrated an accuracy better than 9 m, without requiring much computational effort. In accordance with the obtained results, we can conclude that the recursive least squares procedure implemented with Givens rotations algorithm is a simple, reliable, and numerically stable method for orbit determination by using GPS measurements. The attained precision can, of course, be improved, and the authors are working on it, considering other relevant perturbing effects [21]. Finally, it is important to point out that for the target application, the CBERS-3 and 4 satellites, the perturbations due to atmospheric drag, solar radiation pressure, lunisolar gravity field and tides, as well as measurement errors (e.g., ionosphere) must be taken into account for precision applications.

Acknowledgments

The authors wish to express their appreciation for the support provided by The São Paulo State Research Foundation (FAPESP), —Brazil, under contracts no. 98/14358-2 and no. 02/07129-4, and by Brazilian Council of Scientific Research and Development (CNPq) contracts no. 300141/2002-5 and no. 303143/2004-5.

References

[1] A. Leick, *GPS Satellite Surveying*, John Wiley & Sons, New York, NY, USA, 1995.

[2] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, vol. 128 of *Mathematics in Science and Engineering*, Academic Press, New York, NY, USA, 1977.

[3] H. K. Kuga, “Orbit determination of Earth artificial satellites through estimation techniques combined with state smoothing techniques,” Ph.D. Dissertation, (INPE-4959-TDL/388), INPE, São José dos Campos, SP, Brazil, November 1989.

[4] A. P. M. Chiaradia, H. K. Kuga, and A. F. B. A. Prado, “Single frequency GPS measurements in real-time artificial satellite orbit determination,” *Acta Astronautica*, vol. 53, no. 2, pp. 123–133, 2003.

[5] W. Givens, “Computation of plane unitary rotations transforming a general matrix to triangular form,” *SIAM Journal on Applied Mathematics*, vol. 6, no. 1, pp. 26–50, 1958.

[6] C. L. Thornton, *Triangular covariance factorization for Kalman filtering*, Ph.D. thesis, Jet Propulsion Laboratory, University of California, Pasadena, Calif, USA, 1972, NASA-CR- 149147.

[7] H. D. Rim and B. E. Schutz, “Precision orbit determination (POD),” Algorithm Theoretical Basis Document, Center for Space Research, The University of Texas at Austin, Austin, Tex, USA, October 2002.

[8] INPE, www.cbers.inpe.br/en/programas/aplicacoes2.htm, March 2007.

[9] O. Montenbruck and E. Gill, *Satellite Orbits. Models, Methods and Applications*, Springer, Berlin, Germany, 2000.

[10] A. Gelb, Ed., *Applied Optimal Estimation*, The MIT Press, Cambridge, Mass, USA, 1974.

[11] R. S. Nerem, F. J. Lerch, J. A. Marshall, et al., “Gravity model development for TOPEX/POSEIDON: joint gravity models 1 and 2,” *Journal of Geophysical Research*, vol. 99, no. C12, pp. 24421–24448, 1994.

[12] S. Pines, “Uniform representation of the gravitational potential and its derivatives,” *AIAA Journal*, vol. 11, no. 11, pp. 1508–1511, 1973.

[13] J. B. Lundberg and B. E. Schutz, "Recursion formulas of legendre functions for use with nonsingular geopotential models," *Journal of Guidance, Control, and Dynamics*, vol. 11, no. 1, pp. 31–38, 1988.

[14] P. W. Binning, "Satellite orbit determination using GPS pseudoranges under SA (AAS 97-111)," in *Spaceflight Mechanics*, vol. 95 of *Advances in the Astronautical Sciences*, pp. 183–193, American Astronautical Society, San Diego, Calif, USA, 1997.

[15] O. Montenbruck and M. Suarez, "A modular Fortran library for sequential least squares estimation using QR factorization," DLR-GSOC-IB 94-05, German Space Operations Center, Oberpfaffenhofen, Germany, 1995.

[16] G. H. Golub and C. Van Loan, *Matrix Computations*, vol. 3 of *Johns Hopkins Series in the Mathematical Sciences*, Johns Hopkins University Press, Baltimore, Md, USA, 2nd edition, 1989.

[17] A. S. Householder, "Unitary triangularization of a nonsymmetric matrix," *Journal of the Association for Computing Machinery*, vol. 5, no. 4, pp. 339–342, 1958.

[18] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall Series in Automatic Computation, Prentice-Hall, Englewood Cliffs, NJ, USA, 1974.

[19] J. M. Dow, R. E. Neilan, and G. Gendt, "The international GPS service: celebrating the 10th anniversary and looking to the next decade," *Advances in Space Research*, vol. 36, no. 3, pp. 320–326, 2005.

[20] A. P. M. Chiaradia, E. Gill, O. Montenbruck, H. K. Kuga, and A. F. B. A. Prado, "Algorithms for on-board determination using, GPS-OBODE-GPS," DLR-GSOC TN 00-04, German Space Operations Center, Oberpfaffenhofen, Germany, 2000.

[21] P. C. P. Raimundo, "Orbit determination using GPS considering solar radiation pressure model for the TOPEX/POSEIDON satellite," M.S. Dissertation, INPE, São José dos Campos, SP, Brazil, 2007.

Rodolpho Vilhena de Moraes: Departamento de Matemática, UNESP Campus de Guaratinguetá,
12516-410 Guaratinguetá, SP, Brazil
Email address: rodolpho@feg.unesp.br

Aurea Aparecida da Silva: Divisão de Mecânica Espacial e Controle,
Instituto Nacional de Atividades Espaciais, 12227-010 São José dos Campos, SP, Brazil
Email address: aurea@dem.inpe.br

Helio Koiti Kuga: Divisão de Mecânica Espacial e Controle,
Instituto Nacional de Atividades Espaciais, 12227-010 São José dos Campos, SP, Brazil
Email address: hkk@dem.inpe.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 32514, 11 pages
doi:10.1155/2007/32514

Research Article

Evaluation of Tropospheric and Ionospheric Effects on the Geographic Localization of Data Collection Platforms

C. C. Celestino, C. T. Sousa, W. Yamaguti, and H. K. Kuga

Received 29 September 2006; Accepted 23 April 2007

Recommended by José Manoel Balthazar

The Brazilian National Institute for Space Research (INPE) is operating the Brazilian Environmental Data Collection System that currently amounts to a user community of around 100 organizations and more than 700 data collection platforms installed in Brazil. This system uses the SCD-1, SCD-2, and CBERS-2 low Earth orbit satellites to accomplish the data collection services. The main system applications are hydrology, meteorology, oceanography, water quality, and others. One of the functionalities offered by this system is the geographic localization of the data collection platforms by using Doppler shifts and a batch estimator based on least-squares technique. There is a growing demand to improve the quality of the geographical location of data collection platforms for animal tracking. This work presents an evaluation of the ionospheric and tropospheric effects on the Brazilian Environmental Data Collection System transmitter geographic location. Some models of the ionosphere and troposphere are presented to simulate their impacts and to evaluate performance of the platform location algorithm. The results of the Doppler shift measurements, using the SCD-2 satellite and the data collection platform (DCP) located in Cuiabá town, are presented and discussed.

1. Introduction

The current Brazilian Environmental Data Collection System is composed of the SCD-1, SCD-2, and CBERS-2 satellite constellations (space segment), a network of more than 700 data collection platforms (DCP) spread out in Brazil, the Reception Stations of Cuiabá and Alcântara, and the Data Collection Mission Center. Figure 1.1(a) illustrates the Brazilian Environmental Data Collection System and Figure 1.1(b) the system space segment. In this system, the satellite works as a message retransmitter (bent pipe transponder).

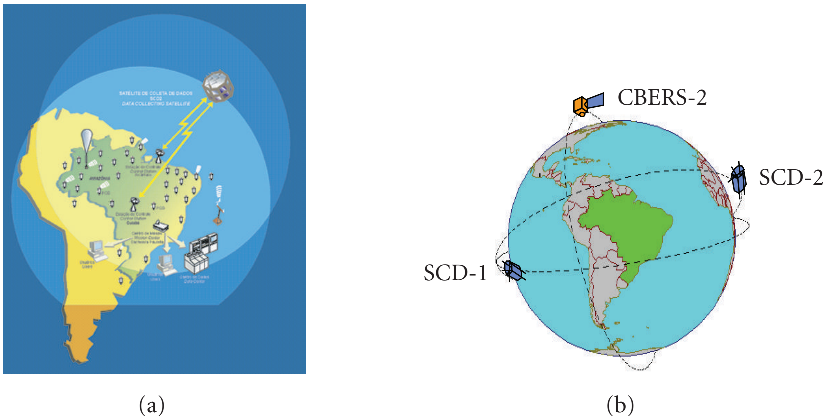


Figure 1.1. (a) Brazilian Environmental Data Collection System with Cuiabá and Alcântara station visibility circles and (b) space segment composed of SCD-1, SCD-2, and CBERS-2 satellites.

A communication link between a data collection platform (DCP) and a reception station is established through one of the satellites. The platforms installed on ground (fixed or mobile) are configured for transmission intervals spanning from 40 to 220 seconds. Each message may have up to 32 bytes of useful data that correspond to a 1-second transmission burst. The DCP messages retransmited by the satellites and received by the Cuiabá or Alcântara stations are sent to the Data Collection Mission Center located at Cachoeira Paulista (São Paulo state) for processing, storage, and dissemination to the users. The users receive the processed data through Internet, at most 30 minutes after being received at a station. The DCP geographical location could be determined by using the Doppler effect or by the use of a GPS receiver. Considering the Doppler shift as the platform location method, the coordinates of a platform are obtained from the Doppler shift measurements of the transmitter frequency carrier signal [1]. As these signals spread in the terrestrial atmosphere, among other factors, they are influenced by the electrochemical elements that compose the atmosphere layers, which generate a propagation delay, and cause errors in the final coordinates supplied to the system users. The signal propagation delay due to the atmospheric effects consists, essentially, of the ionospheric and tropospheric effects.

Zenithal delays due to ionosphere can range from a few meters up to dozens of meters, while that due to troposphere is usually around three meters [2]. To evaluate the impacts on geographical location due to the ionospheric and tropospheric effects, the SCD-2 satellite and DCP #32590, located in Cuiabá, with latitude 15.55293°S and longitude 56.06875°W were considered in this work.

The content of this article is the following: the effects in the geographic location and characteristics of the ionosphere and troposphere are shown in Section 2; the models are described in Section 3; in Section 4, qualitative analyses of the ionospheric and tropospheric effects are presented; Section 5 presents the results of the evaluation of tropospheric and ionospheric effects, and Section 6 presents the conclusions and final remarks.

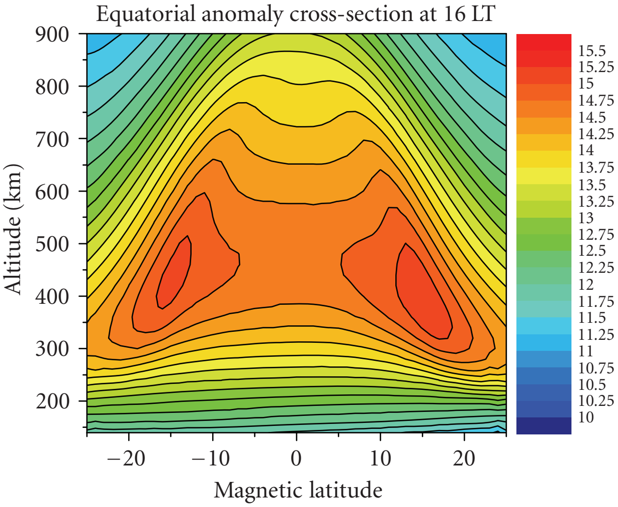


Figure 2.1. Ionospheric electron density (expressed as ln in units of cm^{-3}) in the magnetic meridional plane for the Brazilian longitudinal sector calculated by Sheffield University Plasmasphere Ionosphere Model (Communication: Jonas R. Souza and G. J. Bailey).

2. Ionosphere and troposphere characteristics

This section presents a brief description of the ionosphere and troposphere characteristics.

2.1. Ionosphere. The ionosphere layer is located between 50 and 1.000 kilometers above the Earth surface [3], and is composed of ions and electrons, being thus named ionosphere. The key agent of ionization is the Sun, whose radiation in the X-ray and ultraviolet bands inserts a great amount of free electrons in the environment. In the ionosphere the density of free electrons is variable in close connection with the hour of the day, season, and chemical composition of the high atmosphere. The refraction in the ionosphere depends on the signal frequency and is proportional to the total electron content (TEC) along the path traveled by the signal between platform and satellite.

Figure 2.1 shows a sample of the ionospheric electron density. The data of the ionosphere in Brazil are obtained using rockets, satellites, modeling systems, and simulation of the ionospheric and thermospheric processes.

2.2. Troposphere. The effect of the troposphere depends on the density of the atmosphere and on the satellite line of sight elevation angle. This effect can be observed from ground up to approximately 50 km height. Tropospheric effect on signal propagation at frequencies below 30 GHz do not depend on the transmitted frequency [3]. The influence of the gaseous mass can be divided in two parts: (a) composed of dry gases, called dry or hydrostatic component and, (b) composed of water vapor, called wet or humid component. The tropospheric delay is generated by these components: hydrostatic and wet. The delay due to hydrostatic component can correspond to approximately 2 to 3 m in the zenith and it varies with the temperature and the local atmospheric pressure [3]; the delay for the wet component is of approximately 1 to 30 cm at the zenith [4], however its variation is very large. Its prediction with good accuracy becomes a difficult task.

3. Ionospheric and tropospheric models

The models used in this work are described below.

3.1. Ionospheric model. The ionospheric signal delay is given by [5]

$$R_I = \frac{40.3\ \mathrm{VTEC}\sec Z}{f^2}, \tag{3.1}$$

where VTEC is the total electron content in the vertical direction (el/m^2), Z is the signal path zenithal angle in relation to the plane of the mean altitude of 350 km approximately, and f is the platform transmitter frequency (Hz).

At the ionospheric pierce point Z is given as

$$\sin Z = \frac{R_E \cos\gamma}{R_E + H}, \tag{3.2}$$

where R_E is the Earth's equatorial radius, H and γ are altitude and satellite elevation angle, respectively.

Substituting (3.2) in (3.1) and differentiating in function to time, we get

$$\dot{R}_I = -\frac{36.2\ \mathrm{VTEC}\ \cos\gamma\sin\gamma\dot{\gamma}}{f^2\left(1 - 0.9\cos^2\gamma\right)^{3/2}}, \tag{3.3}$$

where $\dot{\gamma}$ is satellite elevation angle rate.

Equation (3.1) models the ionospheric signal delay and (3.3) models the time variation of this delay. This can be applied to the ionospheric correction based on the Doppler shift measurements.

The delay due to the ionosphere is sensitive to the variable VTEC. The values used for this variable were obtained from IRI-2001 (International Reference Ionosphere) [6].

3.2. Tropospheric models. Three tropospheric models are considered. The first model is the Hopfield empiric model for the tropospheric delay in function of the temperature and pressure values measured on ground. It is given by [4]

$$T_r^s = T_{ZH} m_b(\gamma) + T_{ZW} m_w(\gamma), \tag{3.4}$$

where

$$T_{ZH} = 155.2 \times 10^{-7} \frac{P}{T} H_d \tag{3.5}$$

is the zenithal delay of the dry component,

$$T_{ZW} = 155.2 \times 10^{-7} \frac{4810\,e}{T^2} H_W \tag{3.6}$$

is the zenithal delay of the humid component, T is the temperature (in degrees K), P is the dry pressure (in hPa), e is the humid pressure (in hPa), and

$$\begin{gathered} H_d = 40136 + 148.72(T - 273.16), \\ H_W = 11000\,\text{m}. \end{gathered} \tag{3.7}$$

$m_b(\gamma)$ and $m_W(\gamma)$ are mapping functions that relate the dry and humid delay components with the elevation angle (γ) in degrees and are given by

$$\begin{gathered} m_b(\gamma) = \csc\left(\gamma^2 + 6.25\right)^{1/2}, \\ m_w(\gamma) = \csc\left(\gamma^2 + 2.25\right)^{1/2}. \end{gathered} \tag{3.8}$$

The second model is the Saastamoinen model [7–9]:

$$\begin{gathered} T_{ZH} = 0.002277DP, \\ T_{ZW} = 0.002277eD\left(\frac{1255}{T} + 0.05\right), \end{gathered} \tag{3.9}$$

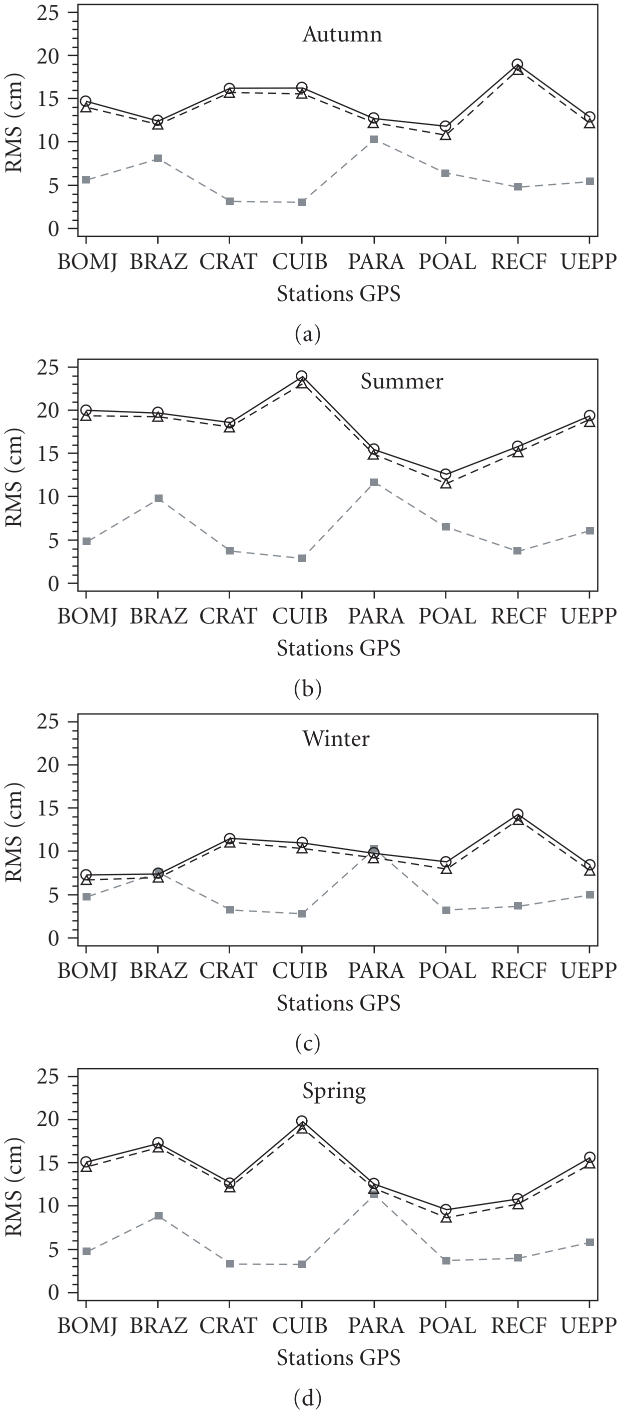
where $D = (1 + 0.0026\cos 2\varphi + 0.00028H)$, and φ and H (in km) are the satellite latitude and altitude, respectively.

The third model is a dynamic model that is being used at Center for Weather Forecasting and Climate Studies (CPTEC-INPE) to provide the zenithal tropospheric delay (Z_{TD}). The predictions of Z_{TD} values are obtained from estimation of temperature, surface atmospheric pressure and humidity generated by the numeric weather prediction (NWP) with observed initial conditions [10]. The dynamic model data are available on the Internet site http://satelite.cptec.inpe.br/htmldocs/ztd/zenital.htm.

4. Qualitative analysis of the ionospheric and tropospheric effects

This section presents a qualitative analysis of the tropospheric delay values (hydrostatic and wet components) using the CPTEC's dynamic model [10] and the ionospheric delay values obtained from the IRI's model [6].

Figure 4.1 shows the root mean square (RMS) errors of the zenithal tropospheric delay resulted from comparison between the dynamic modeling of [10], the Hopfield empiric



(a)

(b)

(c)

(d)

Figure 4.1. Zenithal tropospheric delay RMS values resulted from comparison of Hopfield (∘), Saastamoinem (△), dynamic (■) models, and GPS reference data (source: [10]).

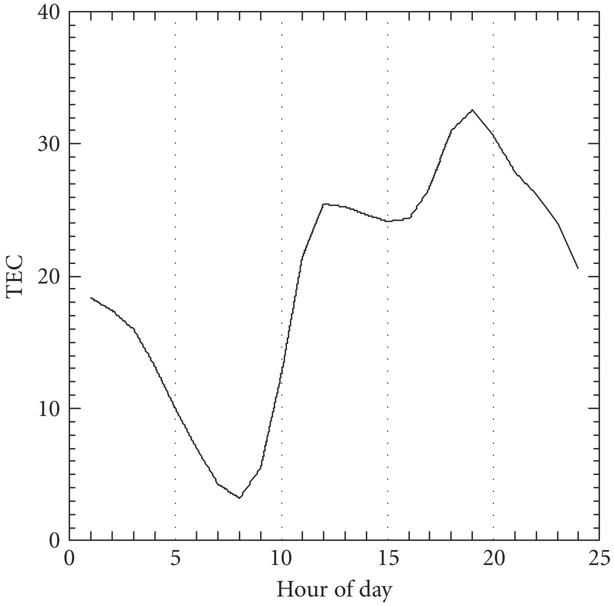


Figure 4.2. VTEC values for DCP no. 32590 location (IRI-2001).

model [4], the Saastamoinem model [7–9], and GPS reference data from Brazilian Continuous GPS Network (RBMC) collected during one year starting from March 2004. Observe that the maximum RMS difference between the zenithal values is 20 cm approximately, considering the summer season and the GPS ground station in Cuiabá. This difference happens because in the Hopfield and Saastamoinem models mean values are used for the temperature, hydrostatic, and wet components while in the dynamic modeling the temperature is a real measurement. Besides, the Hopfield and Saastamoinem models standard mean values are obtained in the subtropical areas such as Europe and North America. For those reasons, the evaluation of the tropospheric effects in this work considers the data from dynamic modeling because it is more suitable for tropical conditions of Brazil.

Figure 4.2 shows the VTEC values used in the numeric simulations (IRI, 2001). These values were obtained considering the location of the 32590 DCP on 2006, April 7th. Observe that in this figure the largest value occurs at 7 pm UTC. The location errors due to this effect for 11 am and 6 pm were of 24.6 m and 39.5 m, respectively, as presented in Table 4.1.

In [11], the ionospheric zenithal signal delay considering Rio de Janeiro city (Brazil) is 30 meters approximately, and using the model IRI is 10 meters approximately. This difference happens because TEC causes a decrease in the GPS signal, and in the region above Brazil it depends strongly on the ionospheric equatorial anomaly [11]. For tropical regions like in Brazil, ionospheric irregularities occurrence can affect drastically the GPS performance.

Table 4.1. Geographical location errors considering DCP no. 32590 as obtained with simulated conditions, simulated location error due to ionospheric effect, simulated location error due to tropospheric effect, and simulated location error considering both effects.

Date UTC Time	Min elevation angle (deg)	Max elevation angle (deg)	Simulated location error (km)	Simulated ionospheric effect error (km)	Simulated tropospheric effect error (km)	Simulated location error with both effects (km)
April, 6 2 pm	4	27	0.0063	0.1296	0.0304	0.1628
April, 7 11 am	15	42	0.0154	0.1455	0.0304	0.1910
April, 7 6 pm	9	33	0.0063	0.1874	0.0399	0.2267
April, 10 9 am	10	73	0.0503	0.0545	0.0789	0.1835

The IRI ionospheric model results were used in this work for the DCP location. Knowing that the model can also be inaccurate, we can conclude that the location error due to the ionospheric effect can be larger than the values presented here.

5. Results

The results and analysis of the data collection platform geographic location due to the ionospheric and the tropospheric effects are shown here, demonstrating the location accuracy achieved. We obtained data collected in Cuiabá Reception Station (Central Brazil) using the Brazilian satellite SCD-2 of low orbit with 25° inclination in relation to Equator and altitude of approximately 750 km. We considered a nearby Data Collecting Platform number 32 590 with known latitude of 15.55293°S and longitude of 56.06875°W. We gathered representative data sets of 32 590 DCP transmitter at three different days and times as shown in Tables 4.1 and 5.1.

Table 4.1 shows the results considering simulated SCD2 Doppler shift values, representing simulated conditions. Table 5.1 shows the results form another analysis considering data measurements gathered in actual conditions.

The minimum elevation (γ_{min}) and the maximum elevation (γ_{max}) angles of the transmitted beam from the platform to satellite are presented in the second and third columns. We used the geographical location algorithm [1] to generate the results. The geographical location errors without considering ionospheric and tropospheric effects, simulated conditions, (e) are represented in the fourth column. The last three columns show location error results considering ionospheric effect ($e(\dot{R}_I)$), tropospheric effect ($e(\dot{R}_T)$), and both effects ($e(\dot{R}_I, \dot{R}_T)$) in the simulated Doppler shift measurements.

It can be observed in Table 4.1 that the error in DCP location in the simulated Doppler measurements due to the ionospheric effect was larger than the error considering the

Table 5.1. Geographical location errors for the DCP no. 32590 as obtained today without any correction (actual conditions), actual location error with ionospheric correction, actual location error with tropospheric correction, and final location error considering both corrections.

Date UTC Time	Min elevation angle (deg)	Max elevation angle (deg)	Actual location error (km)	Actual location error with ionospheric correction (km)	Actual location error with tropospheric correction (km)	Actual location error with both corrections (km)
April, 6 2 pm	11	27	0.35	0.25	0.33	0.24
April, 7 11 am	15	42	1.06	0.90	1.01	0.87
April, 7 6 pm	9	33	1.75	1.98	1.86	2.10
April, 10 9 am	10	73	2.17	1.87	1.74	1.47

tropospheric effect as expected. The maximum error in the observed location was 187 m for the ionospheric effect and the minimum error was 30 m for tropospheric effect.

Now considering actual conditions with data obtained from flying satellites as shown in Table 5.1, we can observe that in most cases the location error decreases when we consider both corrections as shown in the last column as expected. For these test cases, the errors decreased more than 100 m.

In Table 5.1, several visual inspections of the results generated for the actual case were made. In the first pass (April 6, 2 pm) it was observed that when eliminating the measurement with elevation below 10 degrees, the result was consistent with the expectations. Unlikely the introduction of this measurement in the total of 6 measurements worsened the final result.

Already in the third pass (April 7, 6 pm) two measurements were taken with elevation smaller than 10 degrees, remaining only two good measurements in the total of 4 measurements. With this, the final result was inconsistent with the expected result.

6. Conclusions

The performance analysis of the geographic location of data collection platforms considering the ionospheric and the tropospheric effects is shown. Two different analyses were made: in the first analysis results were obtained considering simulated Doppler shift of the SCD2 satellite pass, as representative of ideal conditions, and the total incremental error in the observed location was less than 227 m; in the second analysis, with data files depicting actual conditions, we can observe that in most cases the location error decreases around hundred meters. The April 7, 6 pm case (Table 5.1) presented insufficient data and its result was not considered representative.

On the average, for these tested cases, we can conclude that the location errors decreased when including the ionospheric and tropospheric corrections.

In as much as the location accuracy being a function of several factors such as on board oscillator stability, DCP transmitter oscillator stability, satellite orbital elements accuracy, data collection processing equipment performance, number of reception stations, among others, the correction of the effects of signal propagation through the ionosphere and troposphere can be an important factor to be considered.

The results herein indicate that the correction of the ionospheric and tropospheric effects can reduce the geographical location errors and improve the performance of the geographical location software.

A proposed follow-on research is to analyze in detail the ionospheric and tropospheric effects considering several DCPs located in different regions, as well as the year season to verify the impact of seasonal effects in the performance of the system. Furthermore, other models such as IONEX from IGS, and other mapping functions are being considered to be included in the system.

Acknowledgments

The authors thank the following people that contributed to this work: Mr. Luiz Fernando Sapucci (INPE) for supplying the tropospheric dynamic model data; Dr. Eurico R. de Paula (INPE) for the usage of IRI model; and to CNPq (Grant no. 382746/2005-8) and to FAPESP (Grant no. 2005/04497-0) for the financial support.

References

[1] C. T. Sousa, “Geolocalização de Transmissores com Satélites Usando Desvio Doppler em Tempo Quase Real,” Ph.D. dissertation, Space Engineering and Technology, Space Mechanics and Control Division, INPE-Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, 2000.

[2] J. A. Klobuchar, “Ionospheric effects on GPS,” in *Global Positioning System: Theory and Applications*, B. W. Parkinson and J. J. Spilker Jr., Eds., vol. 1, pp. 485–515, American Institute of Aeronautics and Astronautics, Washington, DC, USA, 1996.

[3] J. F. G. Mônico, *Posicionamento pelo NAVSTAR-GPS: Descrição, Fundamentos e Aplicações*, Unesp, São Paulo, Brazil, 2000.

[4] G. Seeber, *Satellite Geodesy: Foundations, Methods, and Applications*, Walter de Gruyter, Berlin, Germany, 1993.

[5] K. Aksnes, P. H. Andersen, and E. Haugen, “A precise multipass method for satellite Doppler positioning,” *Celestial Mechanics*, vol. 44, no. 4, pp. 317–338, 1988.

[6] D. Bilitza, International Reference Ionospheric Model—IRI, 2000 e 2001, http://modelweb.gsfc.nasa.gov/models/iri.html.

[7] J. Saastamoinem, “Contribution to the theory of atmospheric refraction,” *Bulletin Geodésiqué*, vol. 105, pp. 279–298, 1972.

[8] J. Saastamoinem, “Contribution to the theory of atmospheric refraction,” *Bulletin Geodésiqué*, vol. 106, pp. 383–397, 1972.

[9] J. Saastamoinem, “Contribution to the theory of atmospheric refraction,” *Bulletin Geodésiqué*, vol. 107, pp. 13–34, 1972.

[10] L. F. Sapucci, L. A. T. Machado, and J. F. G. Mônico, "Predictions of Tropospheric Zenithal Delay for South America: Seasonal Variability and Quality Evaluation," *Revista Brasileira de Cartografia*. In press. 2007.

[11] E. R. de Paula, I. J. Kantor, and L. F. C. de Rezende, "Characteristics of the GPS signal scintillations during ionospheric irregularities and their effects over the GPS system," in *Proceedings of the 4th Brazilian Symposium on Inertial Engineering (SBEIN '04)*, São José dos Campos, Brazil, November 2004.

C. C. Celestino: Division of Space Systems (DSE), National Institute for Space Research (INPE),
12227-010 São José dos Campos, Brazil
Email address: claudia@dss.inpe.br

C. T. Sousa: Mathematics Department, DMA-FEG-UNESP, CP 205, 12500-000 Guaratinguetá, Brazil
Email address: cristina@dss.inpe.br

W. Yamaguti: Division of Space Systems (DSE), National Institute for Space Research (INPE),
12227-010 São José dos Campos, Brazil
Email address: yamaguti@dss.inpe.br

H. K. Kuga: Division of Space Mechanics and Control (DMC), National Institute for Space Research
(INPE), 12227-010 São José dos Campos, Brazil
Email address: hkk@dem.inpe.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 12741, 26 pages
doi:10.1155/2007/12741

Research Article

Incompressible Turbulent Flow Simulation Using the κ-ε Model and Upwind Schemes

V. G. Ferreira, A. C. Brandi, F. A. Kurokawa, P. Seleghim Jr.,
A. Castelo, and J. A. Cuminato

Received 30 November 2006; Accepted 25 March 2007

Recommended by José Manoel Balthazar

In the computation of turbulent flows via turbulence modeling, the treatment of the convective terms is a key issue. In the present work, we present a numerical technique for simulating two-dimensional incompressible turbulent flows. In particular, the performance of the high Reynolds κ-ε model and a new high-order upwind scheme (adaptative QUICKEST by Kaibara et al. (2005)) is assessed for 2D confined and free-surface incompressible turbulent flows. The model equations are solved with the fractional-step projection method in primitive variables. Solutions are obtained by using an adaptation of the front tracking GENSMAC (Tomé and McKee (1994)) methodology for calculating fluid flows at high Reynolds numbers. The calculations are performed by using the 2D version of the *Freeflow* simulation system (Castello et al. (2000)). A specific way of implementing wall functions is also tested and assessed. The numerical procedure is tested by solving three fluid flow problems, namely, turbulent flow over a backward-facing step, turbulent boundary layer over a flat plate under zero-pressure gradients, and a turbulent free jet impinging onto a flat surface. The numerical method is then applied to solve the flow of a horizontal jet penetrating a quiescent fluid from an entry port beneath the free surface.

1. Introduction

With the rapid advance of computer technology, numerical modeling has become an important tool in the understanding of fluid dynamics phenomena. One of the challenging tasks is the computation of incompressible turbulent flows (mainly those with free surfaces), which can, in principle, be carried out by the direct numerical integration of instantaneous Navier-Stokes equations. Unfortunately, due to the large computational effort involved, this technique has been restricted to flows at low Reynolds numbers. Practical calculations at the present time must therefore be based on the unsteady Reynolds averaged Navier-Stokes (URANS) equations, with the high Reynolds κ-ε turbulence model. However, the performance of this modeling decisively depends on the form that the nonlinear advective terms are approximated.

A wide variety of techniques for discretizing the nonlinear advective terms has been proposed over the last 20 years, given that the combination of NVD (normalized variable diagram) [1] and TVD (total variation diminishing) [2] is one of the most popular in the CFD community. For instance, [3] proposed the VONOS (variable-order non-oscillatory scheme), an NVD scheme which emerged, according to [4, 5], as an acceptable upwinding tool for simulation of free surface flows. Reference [6] proposed a third-order accurate and limited scheme named WACEB (weighted-average coefficient ensuring boundedness) TVD. Numerical results for scalar convection problems show that this scheme has the same ability of QUICK in reducing numerical diffusion without introducing spurious extrema (oscillations). However, this scheme still has convergence problems for non-Newtonian flows. As a remedy, [7] devised a high-resolution scheme called CUBISTA (convergent and universally bounded interpolation scheme for treatment of advection) TVD. The evaluation of the accuracy and convergence properties of the scheme was measured in two-dimensional cases by using linear and nonlinear problems and for Newtonian and non-Newtonian flows.

In the last years, a great effort has been made to develop high-order bounded advection schemes that combine the TVD and NVD formulations. Using this combination, recently, [8] derived an upwind scheme for unsteady flow fields (called adaptative QUICKEST), which did a very good job in solving laminar incompressible free surface flows (see [9] for details). The main motivation for the present work is to simulate incompressible turbulent free surface flows at high Reynolds numbers. By using the standard κ-ε turbulence model and the adaptative QUICKEST scheme, the present paper describes an effective 2D finite difference methodology for the numerical solution of this class of flows. The calculations are performed by the 2D version of the *Freeflow* simulation system of [10].

The paper is organized as follows. First, the model equations are set out (Section 2). The initial and boundary conditions are then presented (Section 3). The adaptative QUICKEST scheme is described in Section 4. The numerical technique is given in Section 5, while the finite difference discretization is described in Section 6. Numerical results are presented and discussed in Section 7. Conclusions are presented in Section 8.

2. Equation models

The flow regime of interest in this paper is modeled by the time-dependent, incompressible, constant property 2D Reynolds averaged Navier-Stokes equations, mass conservation equation, and κ-ε model in the primitive variable formulation. In conservative and nondimensional forms, these equations, omitting averaging symbols, can be written as

$$\frac{\partial u}{\partial t}+\frac{\partial(uu)}{\partial x}+\frac{\partial(uv)}{\partial y}=-\frac{\partial p_e}{\partial x}+\frac{1}{\text{Re}}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)+\frac{1}{\text{Fr}^2}g_x$$
$$+\frac{1}{\text{Re}}\left[2\frac{\partial}{\partial x}\left(\nu_t\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(\nu_t\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\right)\right], \tag{2.1}$$

$$\frac{\partial v}{\partial t}+\frac{\partial(vv)}{\partial y}+\frac{\partial(vu)}{\partial x}=-\frac{\partial p_e}{\partial y}+\frac{1}{\text{Re}}\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right)+\frac{1}{\text{Fr}^2}g_y$$
$$+\frac{1}{\text{Re}}\left[2\frac{\partial}{\partial y}\left(\nu_t\frac{\partial v}{\partial y}\right)+\frac{\partial}{\partial x}\left(\nu_t\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\right)\right], \tag{2.2}$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0, \tag{2.3}$$

$$\frac{\partial \kappa}{\partial t}+\frac{\partial(\kappa u)}{\partial x}+\frac{\partial(\kappa v)}{\partial y}=\frac{1}{\text{Re}}\left[\frac{\partial}{\partial x}\left(\left(\frac{1+\nu_t}{\sigma_\kappa}\right)\frac{\partial \kappa}{\partial x}\right)+\frac{\partial}{\partial y}\left(\left(\frac{1+\nu_t}{\sigma_\kappa}\right)\frac{\partial \kappa}{\partial y}\right)\right]+P-\varepsilon, \tag{2.4}$$

$$\frac{\partial \varepsilon}{\partial t}+\frac{\partial(\varepsilon u)}{\partial x}+\frac{\partial(\varepsilon v)}{\partial y}=\frac{1}{\text{Re}}\left[\frac{\partial}{\partial x}\left(\left(\frac{1+\nu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x}\right)+\frac{\partial}{\partial y}\left(\left(\frac{1+\nu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial y}\right)\right]+\frac{C_{1\varepsilon}P-C_{2\varepsilon}\varepsilon}{T_t}. \tag{2.5}$$

In the above equations, t is the time, $u = u(x,y,t)$ and $v = v(x,y,t)$ are, respectively, the components in the x and y directions of the local time-averaged velocity vector field of the fluid, $\kappa = \kappa(x,y,t)$ is the local time-averaged turbulent kinetic energy of the fluctuating motion, $\varepsilon = \varepsilon(x,y,t)$ is the turbulence dissipation rate of κ, $p_e = p + (2/3)(1/\text{Re})\kappa$ is the effective scalar pressure field divided by the density, and $g = (g_x, g_y)$ is the acceleration due to gravity. The nondimensional parameters $\text{Re} = UL/\nu$ and $\text{Fr} = U/\sqrt{L|g|}$ denote the associated Reynolds and Froude numbers, respectively, in which U is a characteristic velocity scale and L is a length scale of the flow. The nondimensional turbulent viscosity ν_t, turbulent shear stress production P, and turbulence time scale T_t are, respectively, defined as

$$\nu_t = C_\mu \kappa T_t, \tag{2.6}$$

$$P = \nu_t\left[2\left(\frac{\partial u}{\partial x}\right)^2+2\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)^2\right], \tag{2.7}$$

$$T_t = \kappa\varepsilon^{-1}, \tag{2.8}$$

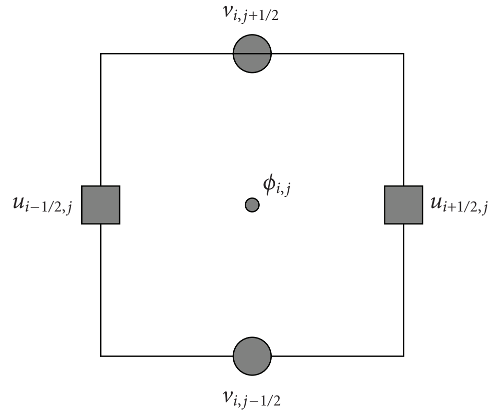


Figure 3.1. Computational cell showing where the variables are discretized. The variable ϕ can be p_e, κ, ε, or ν_t.

where the model constants, obtained from experimental results, are considered as $C_\mu = 0.09$, $C_{1\varepsilon} = 1.44$, $C_{2\varepsilon} = 1.92$, $\sigma_\kappa = 1.0$, and $\sigma_\varepsilon = 1.3$. Equations (2.1)–(2.8) have been non-dimensionalized with the following scalings:

$$\begin{gathered}\overline{u} = uU, \qquad \overline{v} = vU, \qquad \overline{x} = xL, \qquad \overline{y} = yL, \qquad \overline{t} = tLU^{-1}, \qquad \overline{p}_e = peU^2, \\ \overline{\kappa} = \kappa\nu UL^{-1}, \qquad \overline{\varepsilon} = \varepsilon\nu U^2L^{-2}, \qquad \overline{\nu}_t = \nu_t\nu, \qquad \overline{g}_x = g_x|\mathbf{g}|, \qquad \overline{g}_y = g_y|\mathbf{g}|,\end{gathered} \tag{2.9}$$

where variables with a bar refer to their corresponding dimensional variables.

3. Initial and boundary conditions

Equations (2.1)–(2.5) are coupled nonlinear differential equations and, together with the eddy viscosity model (2.6), are sufficient, in principle, to solve for the five unknowns u, v, p_e, κ, and ε when appropriate initial and boundary conditions are specified. In this work, a staggered grid is used where the effective pressure, the turbulent kinetic energy, and the dissipation rate are stored at the centre of a computational grid cell, while velocities are stored at the cell edges. A typical cell showing the physical locations at which these dependent variables are defined is illustrated in Figure 3.1. With this grid system, effective pressure boundary conditions are not needed. The boundary and initial conditions have been implemented as follows.

The initial conditions for the mean velocities and effective pressure are specified in the same way as in the laminar case (see [4] or [9]), that is, these variables are prescribed. It is difficult to specify initial conditions for the turbulent variables, since they must be in agreement with the physics of the problem. Thus, for the problems considered in this work, we prescribe the initial conditions for κ and ε, and hence ν_t, as functions of an upstream turbulent intensity $I = 8.0 \times 10^{-2}$. This specific intensity level is used because it is within the bounds of realistic values. In nondimensional form, the turbulent variables can be written as

$$\kappa = I\,\mathrm{Re}, \qquad \varepsilon = 0.33\times 10^{-2}\left(\mathrm{Re}^{-1}\kappa^{3}\right)^{1/2}. \tag{3.1}$$

Four types of boundary conditions have been implemented, namely: inflow, outflow, free surface, and rigid-wall boundaries. At the inflow, the velocities u and v are prescribed while the values of κ and ε are estimated in such a way that they are consistent with the initial conditions (3.1). At the outflow, the streamwise gradient for each variable is required to be equal to zero. At a free surface, we consider the fluid to be moving into (or out of) a passive atmosphere (zero-pressure) and, in the absence of surface tension forces, the normal and tangential components of the stress must be continuous across the free surface; hence on such a surface, we have (see, e.g., [11])

$$\mathbf{n}\cdot(\tau\cdot\mathbf{n}) = p_{\text{ext}}, \tag{3.2}$$

$$\mathbf{m}\cdot(\tau\cdot\mathbf{n}) = 0. \tag{3.3}$$

Here, $\mathbf{n}$ and $\mathbf{m}$ are, respectively, unit normal and tangential vectors to the surface, p_{ext} is the external (atmospheric) pressure (assumed zero in this paper), and $\tau = \tau(p_e, \nu_t, \mathbf{u})$ is the Cauchy stress-tensor given by

$$\tau = -p_e\mathbf{I} + \mathrm{Re}^{-1}\left(1+\nu_t\right)\left\{\nabla\mathbf{u} + (\nabla\mathbf{u})^{T}\right\}, \tag{3.4}$$

where $\mathbf{I}$ denotes the identity tensor. Equations (3.2) and (3.3) are discretized by accurate local finite difference approximations on the free surface, namely, from condition (3.2) one determines the effective pressure; and from (3.3) one obtains the velocities at the free surface. Due to the complexity of the dynamics of the turbulence near the free surface, the values of the turbulent variables κ and ε at the boundary are difficult to specify. For instance, it is not known how turbulence interacts with surface tension, and therefore, it is difficult to specify the distribution of κ on an irregular moving boundary. So, as a first approximation, we assume that the free surface is locally flat and the movement of the fluid does not cause any discontinuities at the boundary. In summary, the turbulent variables at the free surface are determined by imposing

$$\frac{\partial\kappa}{\partial n} = 0, \qquad \frac{\partial\varepsilon}{\partial n} = 0. \tag{3.5}$$

The derivatives in (3.5) are approximated by first-order (either forward or backward) finite difference schemes.

The κ-ε model, as formulated in (2.4)-(2.5), cannot be applied as the calculation approaches a rigid wall. This is because the turbulent time scale in (2.5) exhibits a singular nature near a wall, since the turbulent kinetic energy κ tends to zero there [12]. For this reason, the wall function is employed in the near-wall region. In this case, the fundamental equation for determining the fictitious velocities and turbulent variables near a rigid wall is the total momentum flux τ_ω given by [13]

$$\left(\frac{1}{\text{Re}}(1+\nu_t)\left|\frac{\partial u}{\partial n}\right|\right)\bigg|_{\text{wall}} \approx u^{*2} = \tau_\omega, \tag{3.6}$$

where u represents the mean velocity component tangential to the rigid wall, and u^* is the friction velocity. The values of κ and ε in the inertial sublayer are, respectively, prescribed by the well-known relations

$$\kappa = \text{Re}\frac{u^{*2}}{C_\mu^{1/2}}, \qquad \varepsilon = \text{Re}\frac{u^{*3}}{Ky}, \tag{3.7}$$

where $K = 0.41$. In the viscous region close to the wall, we use the strategy of [14], that is,

$$\kappa = \text{Re}\frac{u^{*2}}{C_\mu^{1/2}}\left(\frac{y^+}{y_c^+}\right)^2, \qquad \varepsilon = \sqrt{\frac{1}{\text{Re}}}\frac{\kappa^{3/2}}{l^*}, \tag{3.8}$$

where y^+ is defined as $y^+ = \text{Re}u^*y$, and l^* represents the length scale proposed by [15]. Neglecting the buffer layer of the turbulent boundary layer, the critical value of y^+ (denoted by y_c^+) in (3.8) separates the viscous sublayer from the inertial sublayer.

3.1. Wall boundary conditions. The strategy adopted here to describe the solution of the flow near a rigid wall is the wall function which describes the asymptotic behavior of the turbulent variables near the wall. The main advantages of the wall function approach are (a) the need to extend the calculation right down to the wall is avoided, a fact which saves computing time and storage; and (b) it is not necessary to account for the viscous effects in the turbulence model. In summary, the behavior of the mean velocity profiles in the viscous and inertial sublayers is given by (see, e.g., [16, 17] or [18])

$$u^+ - y^+ = 0, \tag{3.9}$$

$$\ln(Ey^+) - Ku^+ = 0, \tag{3.10}$$

where $u^+ = u/u^*$ and $E = \exp(KB)$; B is an empirical constant and is usually chosen to correspond to a hydrodynamically smooth wall. One of the central questions in the application of the wall functions (3.6)–(3.8) and (3.9)-(3.10) is the accurate determination of the friction velocity, and hence the wall shear stress. This is determined from relation (3.9) or (3.10), depending on the local Reynolds number y^+. The Newton-Raphson method was used to obtain u^* from (3.10), with $u^* = 11.60$ as initial condition. This initial value was obtained from the numerical solution of the system (3.9)-(3.10) (see also [19]). To begin with, we need to know the critical Reynolds number y_c^+ in (3.8). By neglecting the transition sub-layer, the friction velocity is estimated in the following specific way: with the tangential velocity u^* known in the first grid cell adjacent to the wall, u^* is updated according to the value of y^+ given by (3.9). If y^+ is less than y_c^+, we use (3.9); on the other hand, if it is not, we employ (3.10). The fictitious velocities are calculated by the central-difference approximation of (3.6) for a known wall shear stress.

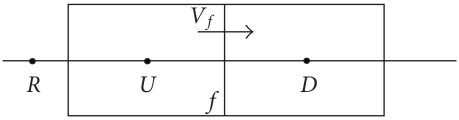


Figure 4.1. Neighboring nodes D, U, and R of the f face.

4. Treatment of nonlinear advection terms

For the flow regime considered in this work, the momentum and turbulence transport equations are dominated by convection, and it is well recognized that standard discretization (i.e., QUICK, central difference, or Lax-Wendroff) for the nonlinear terms leads to oscillatory solutions. In this paper, the discretization of all advective terms in transport equations (2.1)-(2.2) and (2.4)-(2.5) is performed by using the adaptative QUICKEST scheme of [8] (see also [9]). This high-order upwind technique is derived from the normalized variable of [20] and by enforcing the total variation diminishing property of [2, 21]. Consequently, it satisfies the CBC of [22]. The main idea in the derivation of this scheme was to combine accuracy and monotonicity, while ensuring flexibility (it depends on a free parameter). The adaptative QUICKEST scheme enjoyies the property that total variation of the variables does not increase with time, thus spurious numerical oscillations are not generated. The numerical solution can be second- or third-order accurate in the smooth parts of the solution, but only first-order near regions with large gradients. In summary, a general interfacial flow property $\widehat{\phi}_f$ is implemented in the current *Freeflow* code by the functional relationship

$$\widehat{\phi}_f = \begin{cases} (2-\theta)\widehat{\phi}_U, & 0 < \widehat{\phi}_U < a, \\ \widehat{\phi}_U + \dfrac{1}{2}(1-|\theta|)(1-\widehat{\phi}_U) - \dfrac{1}{6}(1-\theta^2)(1-2\widehat{\phi}_U), & a \le \widehat{\phi}_U \le b, \\ 1-\theta+\theta\widehat{\phi}_U, & b < \widehat{\phi}_U < 1, \\ \widehat{\phi}_U, & \text{elsewhere}, \end{cases} \tag{4.1}$$

where $\theta = V_f \cdot \delta t/\delta x$ is the convective Courant number, V_f is a convective velocity, and δx is the grid spacing, and $\widehat{\phi}_{()} = (\phi_{()} - \phi_R)/(\phi_D - \phi_R)$ is Leonard's normalized variable (see [1]). The subscripts D, U, and R referring to values at the downstream, upstream, and remote-upstream locations are defined according to the sign of V_f at f face (see Figure 4.1). The constants a and b in (4.1) are given by

$$a = \frac{2-3|\theta|+\theta^2}{7-6\theta-3|\theta|+2\theta^2}, \qquad b = \frac{-4+6\theta-3|\theta|+\theta^2}{-5+6\theta-3|\theta|+2\theta^2}. \tag{4.2}$$

The corresponding flux limiter for the adaptative QUICKEST scheme is as follows (see details in [9]):

$$\psi(r_f) = \begin{cases} 2(1-\theta)r_f, & 0 < r_f < a', \\ \dfrac{2}{3} - |\theta| + \dfrac{\theta^2}{3} + \left(\dfrac{1-\theta^2}{3}\right) r_f, & a' \le r_f \le b', \\ 2(1-\theta), & r_f > b', \\ 0, & r_f < 0, \end{cases} \tag{4.3}$$

where

$$a' = \frac{\theta^2 + 2 - 3|\theta|}{5 - 6\theta + \theta^2}, \qquad b' = \frac{-4 + 6\theta + \theta^2 - 3|\theta|}{\theta^2 - 1}. \tag{4.4}$$

The implementation of this scheme will be presented later (see Section 6).

5. Numerical solution procedure

The governing equations (2.1)–(2.5) are solved in a partly segregated manner using an extension of the GENSMAC methodology of [23] for turbulent flow fields. A detailed description of this technique is provided in [24]. Based on the predictor-corrector method (see, e.g., [25]), the numerical solution procedure is an explicit finite difference, first- or second-order accurate numerical method for calculation of free surface flows as well as confined flows.

For calculations, a uniform Cartesian staggered grid system is used, where the effective pressure, the turbulent kinetic energy, and the dissipation rate are stored at the center of a computational grid cell, while velocities are stored at the cell edges. For flows possessing free surface, this boundary generally moves, and therefore the domain of interest deforms with time. In this case, the front-tracking MAC (marker and cell) method [26] is adopted in *Freeflow* to determine the free surface location. In summary, the interface is represented discretely by connected Lagrangian markers to form a front which lies within and moves through an Eulerian mesh; as the front moves and deforms, interface points are added/deleted and reconnected as necessary (for details, see [10]). To advance the numerical solution in time, the projection method of [27] is employed (see also [25]).

6. Discretizations

The differential equations are discretized using the finite difference technique on a uniform staggered grid system. The temporal derivatives were discretized using the first-order forward difference (Euler's method), while the spatial derivatives were approximated by standard second-order central differences with the exception of the advection terms (denoted here by CONV$(\cdot)$), which are approximated by the adaptative QUICKEST scheme. The Poisson equation in discretized scheme (see [9]) is done by using the usual five-point Laplacian operator, and the corresponding symmetric-positive definite linear system is solved by the conjugate-gradient method. In summary, fluid flow equations (2.1)–(2.5) take the following discretized form.

(i) *Momentum-$\tilde{u}$*:

$$\begin{aligned}
&\tilde{u}^{n+1}_{i+1/2,j}\\
&\quad= u^{n}_{i+1/2,j} + \delta t \Bigg\{ \mathrm{CONV}(u)\big|_{i+1/2,j} - \frac{\tilde{p}_{e_{i+1,j}} - \tilde{p}_{e_{i,j}}}{\delta x}\\
&\qquad + \frac{1}{\mathrm{Re}\,\delta y}\Bigg[\left(\frac{u_{i+1/2,j+1} - u_{i+1/2,j}}{\delta y} - \frac{v_{i+1,j+1/2} - v_{i,j+1/2}}{\delta x}\right)\\
&\qquad\qquad - \left(\frac{u_{i+1/2,j} - u_{i+1/2,j-1}}{\delta y} - \frac{v_{i+1,j-1/2} - v_{i,j-1/2}}{\delta x}\right)\Bigg]\\
&\qquad + \frac{2}{\mathrm{Re}\,\delta x^2}\left(\nu_{t\,i+1,j}\left(u_{i+3/2,j} - u_{i+1/2,j}\right) - \nu_{t\,i,j}\left(u_{i+1/2,j} - u_{i-1/2,j}\right)\right)\\
&\qquad + \frac{1}{\mathrm{Re}\,\delta y}\Bigg[\nu_{t\,i+1/2,j+1/2}\left(\frac{u_{i+1/2,j+1} - u_{i+1/2,j}}{\delta y} + \frac{v_{i+1,j+1/2} - v_{i,j+1/2}}{\delta x}\right)\\
&\qquad - \nu_{t\,i+1/2,j-1/2}\left(\frac{u_{i+1/2,j} - u_{i+1/2,j-1}}{\delta y} + \frac{v_{i+1,j-1/2} - v_{i,j-1/2}}{\delta x}\right)\Bigg] + \frac{1}{\mathrm{Fr}^2} g_x \Bigg\}^{n},
\end{aligned} \tag{6.1}$$

where the superscript n denotes the time level, and

$$\mathrm{CONV}(u)\big|^{n}_{i+1/2,j} = -\left[\frac{\partial(uu)}{\partial x} + \frac{\partial(uv)}{\partial y}\right]^{n}_{i+1/2,j}. \tag{6.2}$$

(ii) *Momentum-$\tilde{v}$*:

$$\begin{aligned}
&\tilde{v}^{n+1}_{i,j+1/2}\\
&\quad= v^{n}_{i,j+1/2} + \delta t \Bigg\{ \mathrm{CONV}(v)\big|_{i,j+1/2} - \frac{\tilde{p}_{e_{i,j+1}} - \tilde{p}_{e_{i,j}}}{\delta y}\\
&\qquad - \frac{1}{\mathrm{Re}\,\delta x}\Bigg[\left(\frac{u_{i+1/2,j+1} - u_{i+1/2,j}}{\delta y} - \frac{v_{i+1,j+1/2} - v_{i,j+1/2}}{\delta x}\right)\\
&\qquad\qquad - \left(\frac{u_{i-1/2,j+1} - u_{i-1/2,j}}{\delta y} - \frac{v_{i,j+1/2} - v_{i-1,j+1/2}}{\delta x}\right)\Bigg]\\
&\qquad + \frac{2}{\mathrm{Re}\,\delta y^2}\left(\nu_{t\,i,j+1}\left(v_{i,j+3/2} - v_{i,j+1/2}\right) - \nu_{t\,i,j}\left(v_{i,j+1/2} - v_{i,j-1/2}\right)\right)\\
&\qquad + \frac{1}{\mathrm{Re}\,\delta x}\Bigg[\nu_{t\,i+1/2,j+1/2}\left(\frac{u_{i+1/2,j+1} - u_{i+1/2,j}}{\delta y} + \frac{v_{i+1,j+1/2} - v_{i,j+1/2}}{\delta x}\right)\\
&\qquad - \nu_{t\,i-1/2,j+1/2}\left(\frac{u_{i-1/2,j+1} - u_{i-1/2,j}}{\delta y} + \frac{v_{i,j+1/2} - v_{i-1,j+1/2}}{\delta x}\right)\Bigg] + \frac{1}{\mathrm{Fr}^2} g_y \Bigg\}^{n},
\end{aligned} \tag{6.3}$$

with

$$\mathrm{CONV}(v)\big|_{i,j+1/2}^{n} = -\left[\frac{\partial(vv)}{\partial y} + \frac{\partial(vu)}{\partial x}\right]_{i,j+1/2}^{n}. \tag{6.4}$$

(iii) *Mass conservation*:

$$\frac{u_{i+1/2,j} - u_{i-1/2,j}}{\delta x} + \frac{v_{i,j+1/2} - v_{i,j-1/2}}{\delta y} = 0. \tag{6.5}$$

(iv) *Poisson equation for* ψ:

$$\frac{\psi_{i+1,j} - 2\psi_{i,j} + \psi_{i-1,j}}{\delta x^2} + \frac{\psi_{i,j+1} - 2\psi_{i,j} + \psi_{i,j-1}}{\delta y^2} = \frac{\widetilde{u}_{i+1/2,j} - \widetilde{u}_{i-1/2,j}}{\delta x} + \frac{\widetilde{v}_{i,j+1/2} - \widetilde{v}_{i,j-1/2}}{\delta y}. \tag{6.6}$$

(v) κ-*equation*:

$$\begin{aligned}
\kappa_{i,j}^{n+1}& \\
={}& \kappa_{i,j}^{n} + \delta t\Bigg\{\mathrm{CONV}(\kappa)\big|_{i,j} + \frac{1}{\mathrm{Re}}\Bigg[\frac{1}{\delta x^2}\Bigg(\Bigg(1 + \frac{\nu_{ti+1,j} + \nu_{ti,j}}{2\sigma_\kappa}\Bigg)(\kappa_{i+1,j} - \kappa_{i,j}) \\
&- \Bigg(1 + \frac{\nu_{ti,j} + \nu_{ti-1,j}}{2\sigma_\kappa}\Bigg)(\kappa_{i,j} - \kappa_{i-1,j})\Bigg) + \frac{1}{\delta y^2}\Bigg(\Bigg(1 + \frac{\nu_{ti,j+1} + \nu_{ti,j}}{2\sigma_\kappa}\Bigg)(\kappa_{i,j+1} - \kappa_{i,j}) \\
&- \Bigg(1 + \frac{\nu_{ti,j} + \nu_{ti,j-1}}{2\sigma_\kappa}\Bigg)(\kappa_{i,j} - \kappa_{i,j-1})\Bigg)\Bigg] + P_{i,j} - \varepsilon_{i,j}\Bigg\}^{n},
\end{aligned} \tag{6.7}$$

where

$$\mathrm{CONV}(\kappa)\big|_{i,j}^{n} = -\left[\frac{\partial(\kappa u)}{\partial x} + \frac{\partial(\kappa v)}{\partial y}\right]_{i,j}^{n}. \tag{6.8}$$

(vi) ε-*equation*:

$$\begin{aligned}
\varepsilon_{i,j}^{n+1}& \\
={}& \varepsilon_{i,j}^{n} + \delta t\Bigg\{\mathrm{CONV}(\varepsilon)\big|_{i,j} + \frac{1}{\mathrm{Re}}\Bigg[\frac{1}{\delta x^2}\Bigg(\Bigg(1 + \frac{\nu_{t_{i+1,j}} + \nu_{t_{i,j}}}{2\sigma_\varepsilon}\Bigg)(\varepsilon_{i+1,j} - \varepsilon_{i,j}) \\
&- \Bigg(1 + \frac{\nu_{t_{i,j}} + \nu_{t_{i-1,j}}}{2\sigma_\varepsilon}\Bigg)(\varepsilon_{i,j} - \varepsilon_{i-1,j})\Bigg) + \frac{1}{\delta y^2}\Bigg(\Bigg(1 + \frac{\nu_{t_{i,j+1}} + \nu_{t_{i,j}}}{2\sigma_\varepsilon}\Bigg)(\varepsilon_{i,j+1} - \varepsilon_{i,j}) \\
&- \Bigg(1 + \Bigg(\frac{\nu_{t_{i,j}} + \nu_{t_{i,j-1}}}{2\sigma_\varepsilon}\Bigg)(\varepsilon_{i,j} - \varepsilon_{i,j-1})\Bigg)\Bigg)\Bigg] + \frac{C_{1\varepsilon}P_{i,j} - C_{2\varepsilon}\varepsilon_{i,j}}{T_{i,j}}\Bigg\}^{n},
\end{aligned} \tag{6.9}$$

where

$$\mathrm{CONV}(\varepsilon)\big|_{i,j}^{n} = -\left[\frac{\partial(\varepsilon u)}{\partial x} + \frac{\partial(\varepsilon v)}{\partial y}\right]_{i,j}^{n}. \tag{6.10}$$

The production of turbulence, the eddy viscosity, and the time scale are discretized, respectively, as follows:

$$\begin{aligned} P^n_{i,j} = \nu^n_{t_{i,j}} \Bigg\{ & \bigg[\frac{2}{\delta x^2}\left(u_{i+1/2,j} - u_{i-1/2,j}\right)^2 + \frac{2}{\delta y^2}\left(v_{i,j+1/2} - v_{i,j-1/2}\right)^2 \bigg] \\ & + \bigg[\frac{1}{4\delta y}\left(u_{i+1/2,j+1} + u_{i-1/2,j+1} - u_{i+1/2,j-1} - u_{i-1/2,j-1}\right) \\ & \quad + \frac{1}{4\delta x}\left(v_{i+1,j+1/2} + v_{i+1,j-1/2} - v_{i-1,j+1/2} - v_{i-1,j-1/2}\right) \bigg]^2 \Bigg\}^n, \\ \nu^n_{t_{i,j}} = C_\mu \frac{\left(\kappa^n_{i,j}\right)^2}{\varepsilon^n_{i,j}}, & \qquad T^n_{t_{i,j}} = \frac{\kappa^n_{i,j}}{\varepsilon^n_{i,j}}. \end{aligned} \tag{6.11}$$

For the nonlinear advection terms in the momentum equations (the advection terms of κ and ε equations follow a similar procedure), the application of the adaptative QUICKEST scheme is as follows. For simplicity, only the discretization of the nonlinear terms in u-component of the time-averaged Navier-Stokes equations will be presented. The discretization of the other nonlinear term is made in a similar way. In position $(i+1/2, j)$ of the mesh, this term can be approximated by the following conservative scheme:

$$\begin{aligned} \left(\frac{\partial(uu)}{\partial x} + \frac{\partial(uv)}{\partial y}\right)\bigg|_{i+1/2,j} \approx & \frac{\overline{u}_{i+1,j}u_{i+1,j} - \overline{u}_{i,j}u_{i,j}}{\delta x} \\ & + \frac{\overline{v}_{i+1/2,j+1/2}u_{i+1/2,j+1/2} - \overline{v}_{i+1/2,j-1/2}u_{i+1/2,j-1/2}}{\delta y}, \end{aligned} \tag{6.12}$$

where the velocities $\overline{u}_{i+1,j}$, $\overline{u}_{i,j}$, $\overline{v}_{i+1/2,j+1/2}$ and $\overline{v}_{i+1/2,j-1/2}$ are obtained by averaging. For instance, $\overline{v}_{i+1/2,j-1/2}$ is approximate by $\overline{v}_{i+1/2,j-1/2} \approx 0.5(v_{i,j-1/2} + v_{i+1,j-1/2})$. The velocities $u_{i,j}$ and $u_{i+1,j}$ are calculated (the other velocities follow similar procedures) by the following.

(i) When $\overline{u}_{i,j} > 0$ and $\hat{u}_{i-1/2,j} = (u_{i-(1/2),j} - u_{i-(3/2),j})/(u_{i+(1/2),j} - u_{i-(3/2),j})$, the value of $u_{i,j}$ is

$$u_{i,j} = \begin{cases} u_{i-1/2,j} & \text{if } \hat{u}_{i-1/2,j} \notin [0,1], \\ (2-\theta)u_{i-1/2,j} - (1-\theta)u_{i-3/2,j} & \text{if } 0 < \hat{u}_{i-1/2,j} < a, \\ \alpha_D u_{i+1/2,j} + \alpha_U u_{i-1/2,j} - \alpha_R u_{i-3/2,j} & \text{if } a \leq \hat{u}_{i-1/2,j} \leq b, \\ (1-\theta)u_{i+1/2,j} + \theta u_{i-1/2,j} & \text{if } b < \hat{u}_{i-1/2,j} < 1. \end{cases} \tag{6.13}$$

(ii) When $\overline{u}_{i,j} < 0$ and $\hat{u}_{i+1/2,j} = (u_{i+(1/2),j} - u_{i+(3/2),j})/(u_{i-(1/2),j} - u_{i+(3/2),j})$, the value of $u_{i,j}$ is

$$u_{i,j} = \begin{cases} u_{i+1/2,j} & \text{if } \hat{u}_{i+1/2,j} \notin [0,1], \\ (2-\theta)u_{i+1/2,j} - (1-\theta)u_{i+3/2,j} & \text{if } 0 < \hat{u}_{i+1/2,j} < a, \\ \alpha_D u_{i-1/2,j} + \alpha_U u_{i+1/2,j} - \alpha_R u_{i+3/2,j} & \text{if } a \leq \hat{u}_{i+1/2,j} \leq b, \\ (1-\theta)u_{i-1/2,j} + \theta u_{i+1/2,j} & \text{if } b < \hat{u}_{i+1/2,j} < 1. \end{cases} \tag{6.14}$$

(iii) When $\overline{u}_{i+1,j} > 0$ and $\hat{u}_{i+1/2,j} = (u_{i+(1/2),j} - u_{i-(1/2),j})/(u_{i+(3/2),j} - u_{i-(1/2),j})$, the value of $u_{i+1,j}$ is

$$u_{i+1,j} = \begin{cases} u_{i+1/2,j} & \text{if } \hat{u}_{i+1/2,j} \notin [0,1], \\ (2-\theta)u_{i+1/2,j} - (1-\theta)u_{i-1/2,j} & \text{if } 0 < \hat{u}_{i+1/2,j} < a, \\ \alpha_D u_{i+3/2,j} + \alpha_U u_{i+1/2,j} - \alpha_R u_{i-1/2,j} & \text{if } a \leq \hat{u}_{i+1/2,j} \leq b, \\ (1-\theta)u_{i+3/2,j} + \theta u_{i+1/2,j} & \text{if } b < \hat{u}_{i+1/2,j} < 1. \end{cases} \tag{6.15}$$

(iv) When $\overline{u}_{i+1,j} < 0$ and $\hat{u}_{i+3/2,j} = (u_{i+(3/2),j} - u_{i+(5/2),j})/(u_{i+(1/2),j} - u_{i+(5/2),j})$, the value of $u_{i+1,j}$ is

$$u_{i+1,j} = \begin{cases} u_{i+3/2,j} & \text{if } \hat{u}_{i+3/2,j} \notin [0,1], \\ (2-\theta)u_{i+3/2,j} - (1-\theta)u_{i+5/2,j} & \text{if } 0 < \hat{u}_{i+3/2,j} < a, \\ \alpha_D u_{i+1/2,j} + \alpha_U u_{i+3/2,j} - \alpha_R u_{i+5/2,j} & \text{if } a \leq \hat{u}_{i+3/2,j} \leq b, \\ (1-\theta)u_{i+1/2,j} + \theta u_{i+3/2,j} & \text{if } b < \hat{u}_{i+3/2,j} < 1, \end{cases} \tag{6.16}$$

where

$$\alpha_D = \frac{1}{6}(\theta^2 - |\theta| + 2), \qquad \alpha_U = \frac{1}{6}(-2\theta^2 + 3|\theta| + 5), \qquad \alpha_R = \frac{1}{6}(1 - \theta^2). \tag{6.17}$$

The Courant number is calculated in the code by analyzing the direction in which information propagates, that is, the sign of a previously calculated local normal averaged velocity $\overline{u}$ at a face of the control volume, and by computing the expression $\theta = \overline{u} \cdot \delta t/\delta x$ in each control volume.

7. Numerical tests

In order to validate the actual *Freeflow* code, incremented with the original κ-ε model and adaptative QUICKEST scheme, we report now the numerical results for a turbulent flow over a backward-facing step, the turbulent boundary layer over a flat surface, and a turbulent jet impinging onto a flat surface. Also, as an example of application, we present a turbulent planar jet penetrating into a pool.

Table 7.1. Estimates for the reattachment length.

Mesh	Adaptative QUICKEST	CUBISTA	WACEB	VONOS
Coarse	6.86	7.13	7.10	6.12
Medium	6.03	6.10	6.06	5.76
Fine	5.50	5.51	5.50	5.42

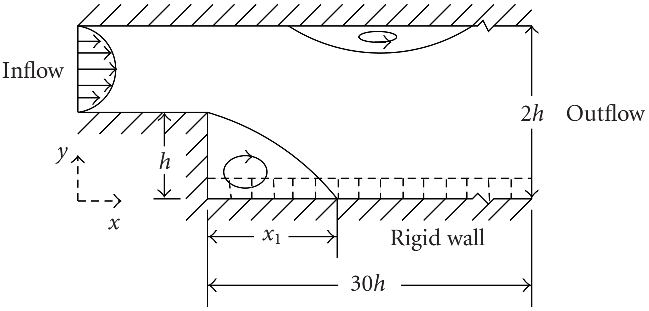


Figure 7.1. Configuration for turbulent flow over a backward-facing step problem.

7.1. Turbulent flow over a backward-facing step. The turbulent flow over a backward-facing step is a standard test case, often used for validation of turbulence models. This flow is computationally challenging, because both a primary and a secondary recirculation eddy vertex occur. The problem configuration is illustrated in Figure 7.1. By using the current *Freeflow* code with a fully developed parabolic velocity profile prescribed at the inlet section, we simulate this flow at Re $= 1.32 \times 10^5$. This is based on the maximum velocity $U_{\max} = 1.0\,\mathrm{ms}^{-1}$ at that section and the height of the step $h = 0.1$ m.

Computations were performed on three different meshes, namely, the coarse mesh (200×15 computational cells, $\delta x = \delta y = 0.02$ m); the medium mesh (400×30 computational cells, $\delta x = \delta y = 0.01$ m); and the fine mesh (800×60 computational cells, $\delta x = \delta y = 0.005$ m). Table 7.1 depicts values of the reattachment length x_1 on the three meshes, using four advection schemes, including the adaptative QUICKEST. By comparing these numerical results with experimental data of [28], which obtained value of $x_1 = 7.1$, one can see that our numerical results underpredict the experimental reattachment point by 20%–25%. The CUBISTA scheme for the coarse mesh provides a value greater than 7.1. The good results with WACEB on coarse meshes can be attributed to the value of the y^+. On the other hand, it can be seen that our numerical results are in good agreement with the numerical result of [29], which found that $x_1 = 6.0$. From this same table, it should also be observed that WACEB, CUBISTA, and adaptative QUICKEST schemes provide good results, while the VONOS scheme gives much less satisfactory results. We believe that most of this difference may be attributed to the fact that the WACEB, CUBISTA, and adaptative QUICKEST schemes are TVD, whereas the VONOS scheme is not.

Table 7.2. Other estimates for the reattachment length.

Scheme	HOAB	STOIC	SMART	CLAM	FOU
x_1	5.405	5.396	5.377	5.357	4.607

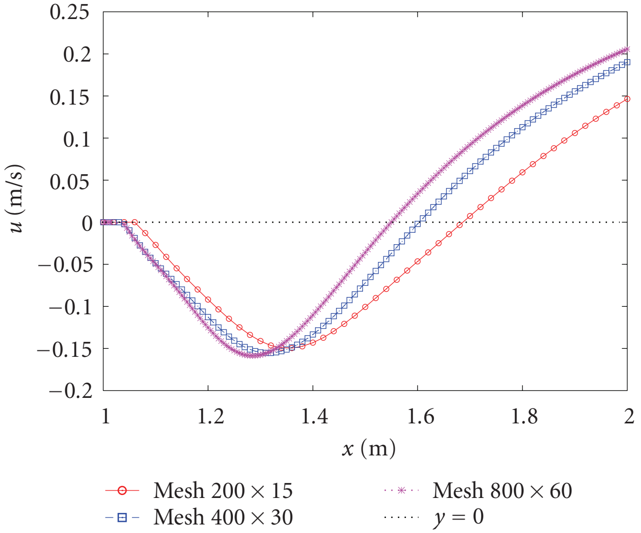


Figure 7.2. Comparison on three meshes of u velocity component using the adaptative QUICKEST scheme.

For simple comparison, Table 7.2 shows other estimates for the reattachment length obtained by HOAB, STOIC, SMART, CLAM, and FOU schemes (see [30]). From this table and Table 7.1, one can observe that the adaptative QUICKEST scheme provided a consistent reattachment length. In addition, a convergence test of the numerical solution obtained with the adaptative QUICKEST scheme on these three meshes was made. This is illustrated in Figure 7.2, which shows how the reattachment length was estimated (the change in the sign of the u-velocity profile adjacent to the lower bounding wall) in the code.

One can see from this figure that both the velocity profile and the reattachment length tend to converge to a solution near the numerical one in the fine mesh. For illustration, Figures 7.3 and 7.4 present the pressure contours and v-component of the velocity field, on the medium mesh, using adaptative QUICKEST scheme.

7.2. Turbulent flow past a flat plate. The interaction between the fluid and the boundary wall is of great importance in turbulent flows. Due to the strong velocity gradients occurring near the wall, a large amount of turbulence is generated. This turbulence plays a very important role in physical phenomena as reattachment of separated regions. In

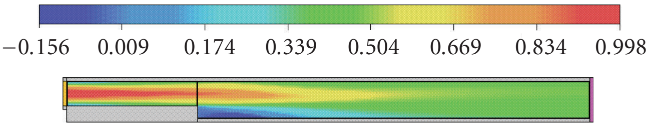


Figure 7.3. u-velocity component using the adaptative QUICKEST and Re $= 1.32 \times 10^5$.

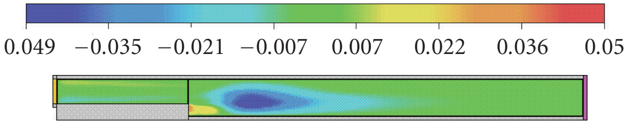
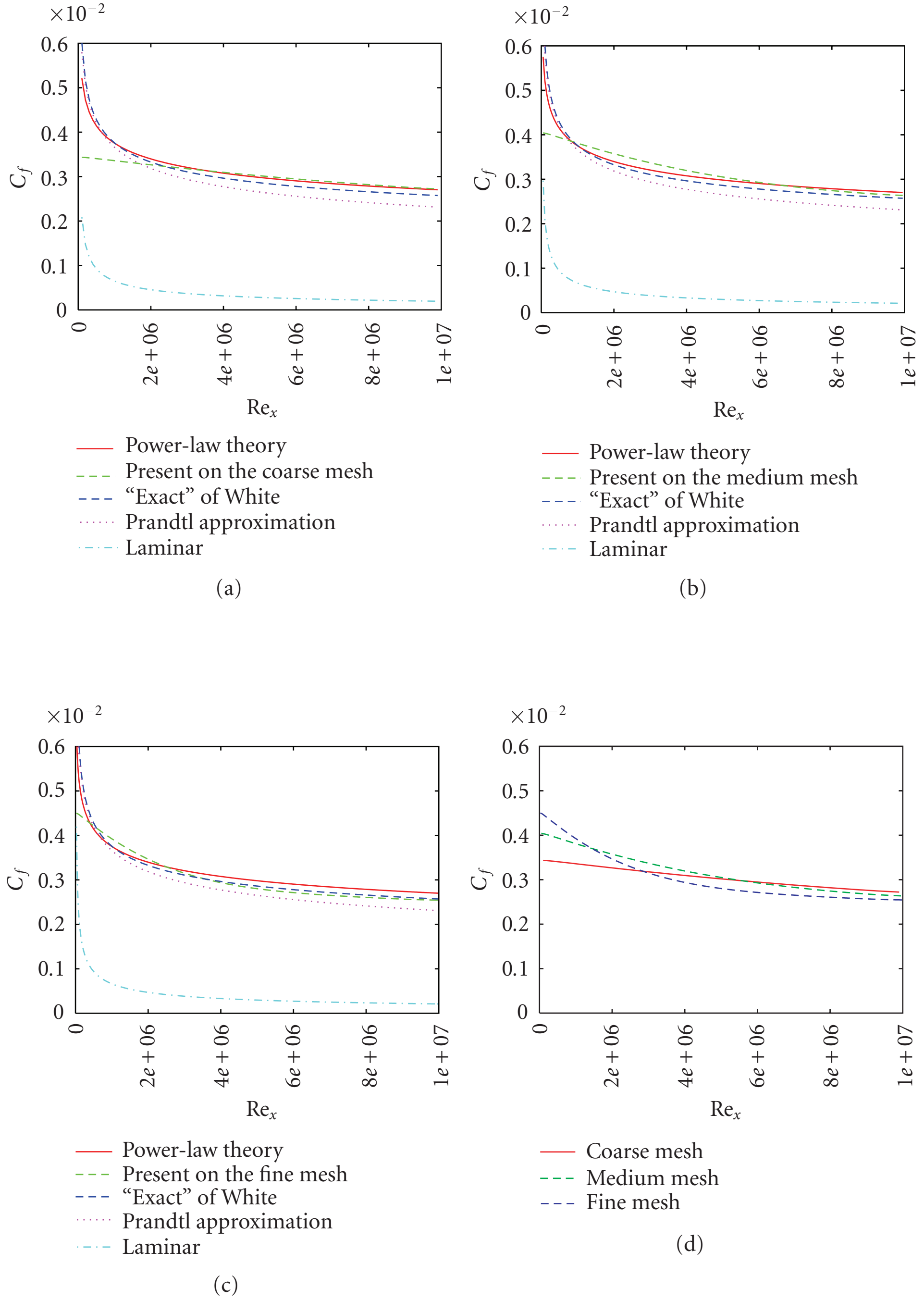


Figure 7.4. v-velocity component using the adaptative QUICKEST and Re $= 1.32 \times 10^5$.

this subsection, a two-dimensional turbulent boundary layer over a flat plate is simulated according to the classical zero-pressure gradient theory. This is justified by the fact that the plate is flat, and thus, the only contribution to the pressure gradient comes from the product between the dynamic viscosity and the second derivative of the longitudinal velocity with respect to the transversal coordinate which, in the present case, can be neglected. Physically, in a zero-pressure gradient turbulent boundary layer, the point of inflection is at the wall itself; there can be no flow separation [31].

This fluid flow problem has been extensively studied in the literature (see, e.g., [5]), and numerous formulae have been proposed to estimate the coefficient of skin friction (C_f). In order to solve this problem, a uniform free stream boundary condition is imposed at the inlet, and the Reynolds number, based on length and velocity scales of unity, is Re $= 2.0 \times 10^6$. Figure 7.5 compares the calculated dimensionless turbulent skin friction coefficient $C_f = 2\tau_w$ with the estimates given by Prandtl approximation, Power-law theory, and the "exact" profile of White (see [18]). These figures display C_f against the local Reynolds number $\mathrm{Re}_x = U_0 x/\nu$ at the (nondimensional) time $t = 6.0$ calculated for the following three different-sized meshes, namely, the coarse mesh (20×100 computational cells, $\delta x = \delta y = 0.05\,\mathrm{m}$); the medium mesh ($40 \times 200$ computational cells, $\delta x = \delta y = 0.025\,\mathrm{m}$); and the fine mesh ($80 \times 400$ computational cells, $\delta x = \delta y = 0.0125\,\mathrm{m}$). Additionally, the corresponding laminar result is also included for simple comparison. As shown in Figures 7.5(a), 7.5(b), and 7.5(c), the numerical estimates are generally satisfactory for Re_x beyond 1.0×10^6. It can also be observed from Figure 7.5(d) that when the coarse mesh was twice refined, there appears to be convergence of the numerical solution to a profile near the power-law theory and the "exact" White relation. On the other hand, for $\mathrm{Re}_x \leq 1.0 \times 10^6$, a systematic discrepancy existed and this may be due to the uniform meshes used and/or the initial velocity profile not being sufficiently turbulent at the entrance region.

7.3. Turbulent jet impinging onto a flat surface. A jet impinging normally onto a flat rigid surface is a good example of a free surface flow, but is difficult to simulate because



(a)

(b)

(c)

(d)

Figure 7.5. Comparison of the local skin friction on a flat plate for turbulent flow, showing several theoretical estimates and those obtained by the present finite difference scheme on three meshes: (a) coarse; (b) medium; (c) fine; and (d) shows comparison of the three numerical solutions.

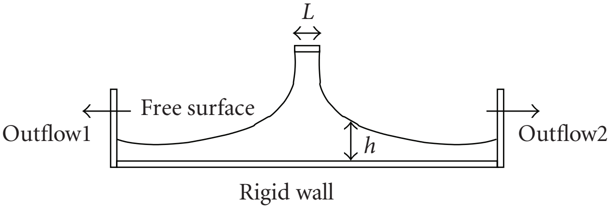


Figure 7.6. Configuration of a free jet impinging onto a rigid surface.

the free surface boundary conditions must be specified on an arbitrarily moving boundary. This free surface flow in turbulent regime is also chosen as a representative test case because there is (see [4, 32]) an approximated analytical solution for the total thickness of the fluid layer flowing on the surface (see the illustration in Figure 7.6). In summary, for a given volumetric flux Q through the inlet section of diameter $L = 2a$, the analytical solution is

$$h(x) = \begin{cases} \dfrac{81(7A)^{1/4}k}{800}\left(\dfrac{\nu}{Q}\right)^{1/4}(x+l) & \text{if } x \geq x_0, \\[2ex] a + \left(1 - \dfrac{A}{k}\right)\delta & \text{if } x < x_0, \end{cases} \tag{7.1}$$

where

$$\begin{aligned} \delta(x) &= \left(\frac{81}{320(9A-2)}\right)^{4/5} 7^{1/5}k\left(\frac{a\nu}{Q}\right)^{1/5} x^{4/5}, \\ x_0 &= \frac{320(9A-2)}{81 \times 7^{1/4}A^{5/4}}a\,\mathrm{Re}^{1/4}, \\ l &= \frac{160(1-2A)}{9 \times 7^{1/4}A^{5/4}}a\,\mathrm{Re}^{1/4}. \end{aligned} \tag{7.2}$$

In (7.1), $A = 0.239$ and $k = 0.260$. The problem configuration is illustrated in Figure 7.6.

By using three different meshes, namely, the coarse mesh (200×50 computational cells, $\delta x = \delta y = 0.001$ m); the medium mesh (400×100 computational cells, $\delta x = \delta y = 0.0005$ m); and the fine mesh (800×200 computational cells, $\delta x = \delta y = 0.00025$ m), the *Freeflow* code, equipped with the adaptative QUICKEST advection scheme and κ-ε model, run this moving free boundary problem at Reynolds number 5.0×10^4, which was based on the maximum velocity $U_{\max} = 1.0$ m/s and diameter of the inlet $L = 0.01$ m (or volumetric flux $Q = \nu$, Re $= 0.01\,\mathrm{m}^2/\mathrm{s}$). On these three meshes, a comparison is made between the free surface height (the total thickness of the layer), obtained from our numerical solutions and from the analytical viscous solution of Watson. This is displayed in Figure 7.7 and its enlargement in Figure 7.8. One can see from these figures that the

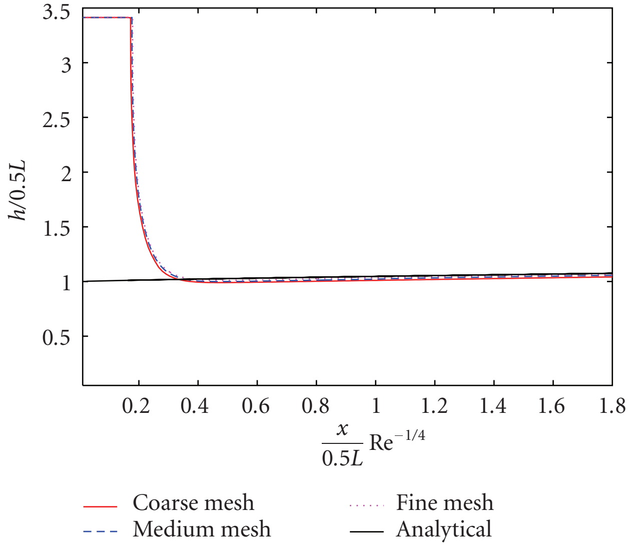


Figure 7.7. Comparison on three meshes between numerical solution and analytical solution of Watson.

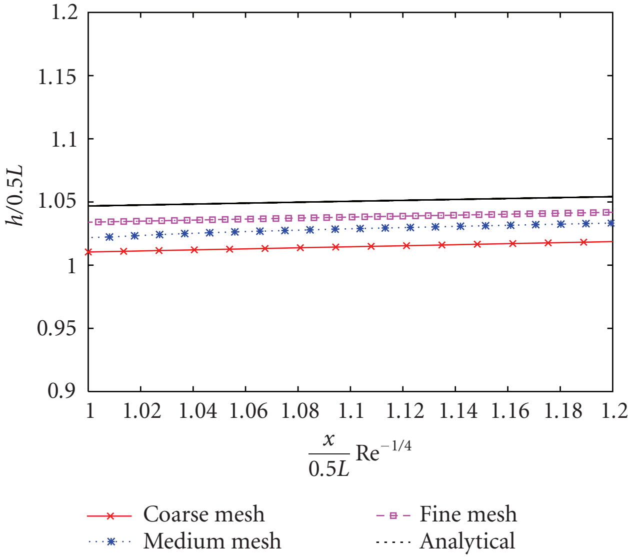


Figure 7.8. Enlargement of a section of Figure 7.7.

numerical results on these meshes are similar, showing, in some regions, a small difference when compared to Watson's solution. We believe that most of this difference may be attributed to insufficient grid points, being used near the rigid wall.

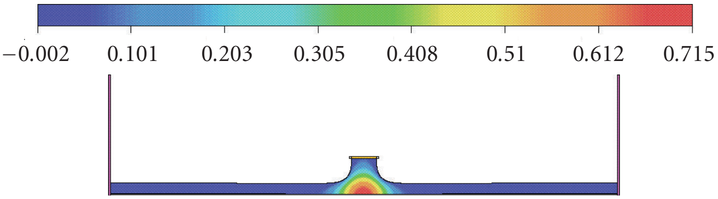


Figure 7.9. Pressure field of a turbulent free jet using adaptative QUICKEST scheme at Re = 5.0×10^4.

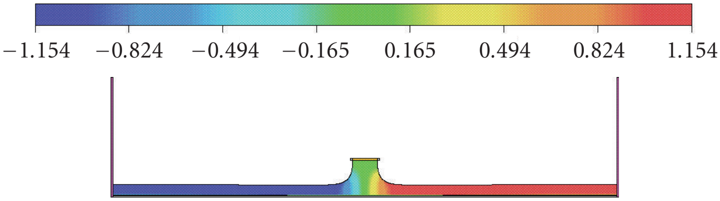


Figure 7.10. u-velocity component of a turbulent free jet using adaptative QUICKEST scheme at Re = 5.0×10^4.

For illustration, Figures 7.9–7.11 present at the time $t = 1.0$ the pressure and velocities fields on the medium mesh using the adaptative QUICKEST scheme.

7.4. Application: a horizontal jet penetrating a quiescent fluid. We conclude this work by presenting the numerical simulation of a horizontal jet penetrating a quiescent fluid from an entry port at depth $H = 6.0\,\text{m}$ beneath the free surface. The purpose here is to show that the actual *Freeflow* can simulate the largest eddies present in the flow and their nonlinear interaction with a free surface. This free surface flow problem has also been simulated by [5] using a classical upwind scheme. The geometrical configuration and parameters for this free surface fluid flow are shown in Figure 7.12. In this numerical simulation, the associated Reynolds and Froude numbers are $\text{Re} = DU_0/\nu = 5.0 \times 10^4$ and $\text{Fr} = U_0/\sqrt{gD} \approx 12.77$, respectively. The mesh used in this test case is 100×100 computational cells ($\delta x = \delta y = 0.010\,\text{m}$). The development of pressure and velocities distributions, together with the free surface elevation at various times, are presented in Figures 7.13 through 7.15. In this case, the interaction with the free surface occurs only at the later stages of the flow development. Initially, one can observe the growth of the instability of the boundary layers between the entering jet and the stagnant fluid, and subsequently, the formation of a pair of counter-rotating eddies. Later on, the first pair of eddies propagates towards the free surface.

The result obtained in this simulation should be interpreted as representing the 2D motion of one realization occurring at scales greater than the discretization scale. In other words, both the free surface position and the velocity fields computed here may be regarded as the deterministic motion at these larger scales. Indeed, this simulation may be

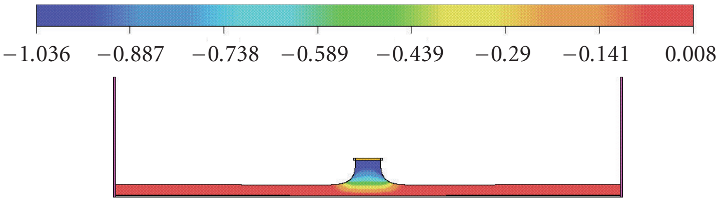


Figure 7.11. v-velocity component of a turbulent free jet using adaptative QUICKEST scheme at $\mathrm{Re} = 5.0 \times 10^4$.

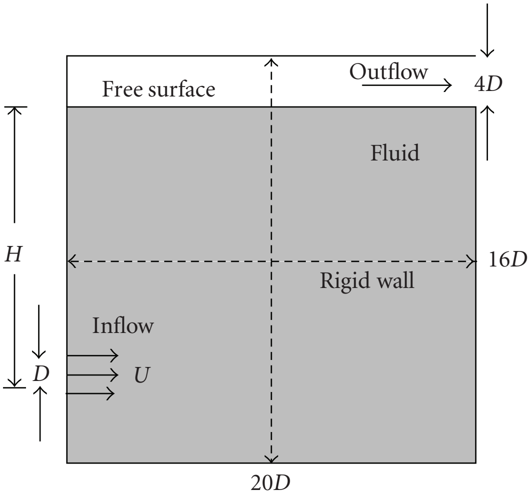


Figure 7.12. Geometry and parameters for flow of a penetrating planar jet in a pool: $U_0 = 2\,\mathrm{ms}^{-1}$ and $D = 0.05\,\mathrm{m}$.

thought of URANS or as VLES (very large scale simulations), as opposed to the 3D, and much more expensive, LES. Turbulent flow simulations using LES and DNS approaches have been performed by other authors, but are mostly restricted to very low (or negligible) Froude numbers.

8. Conclusions

In this work, a finite difference numerical technique for simulating 2D incompressible turbulent flows was described. The *Freeflow* simulation system coupled with the original high Reynolds κ-ε turbulence model and the high-order adaptative QUICKEST advection scheme has been applied to simulate three problems, namely, turbulent flow over a backward-facing step, the turbulent boundary layer over a flat surface (zero-pressure gradient case), and a turbulent free jet impinging onto a flat rigid wall. According to the computed results, the new version of the *Freeflow* code can, in fact, predict both confined and free surface turbulent flows with satisfactory accuracy. In order to illustrate the robustness and applicability of the code to compute the interactions between a free surface

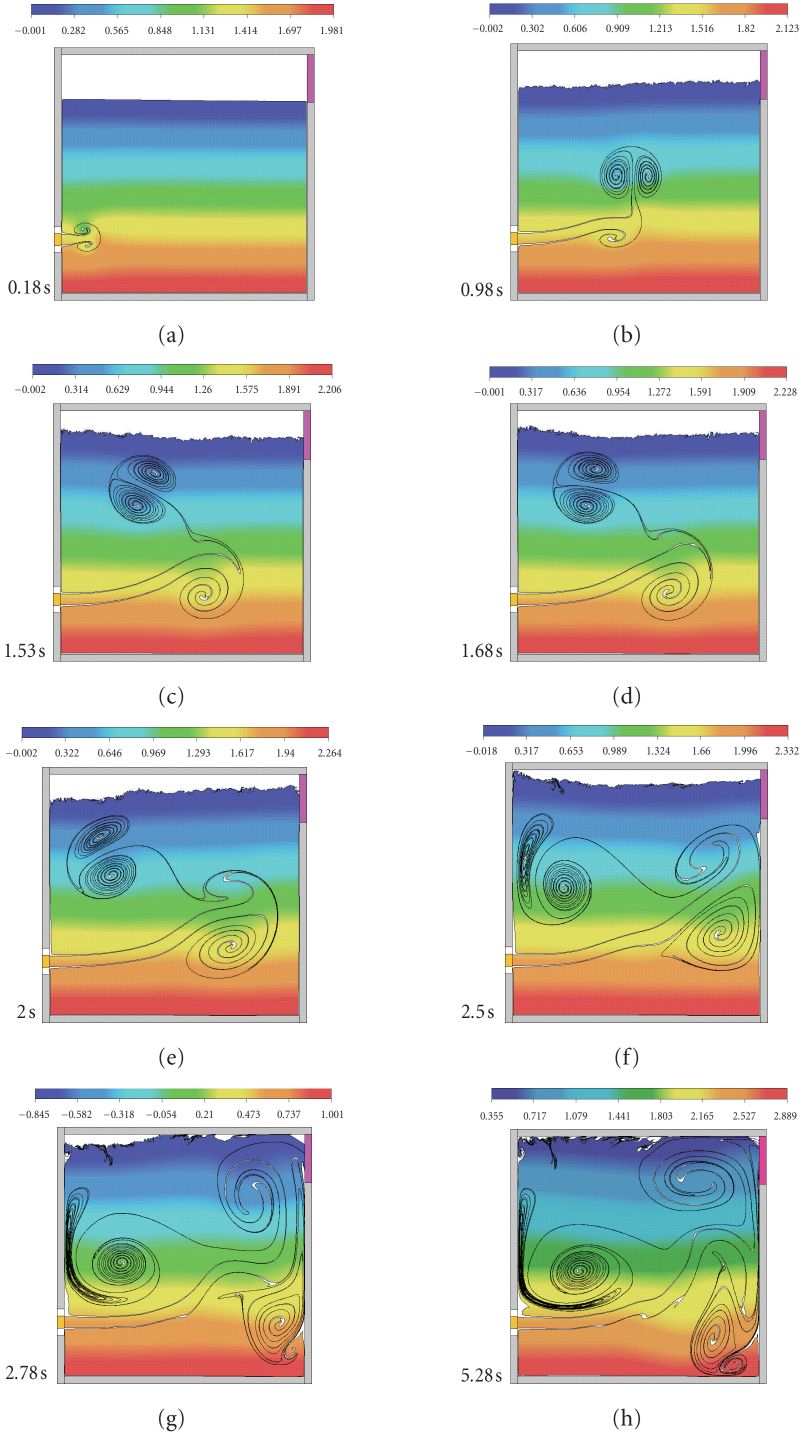


Figure 7.13. Evolution of the pressure contours of a jet in a fluid portion.

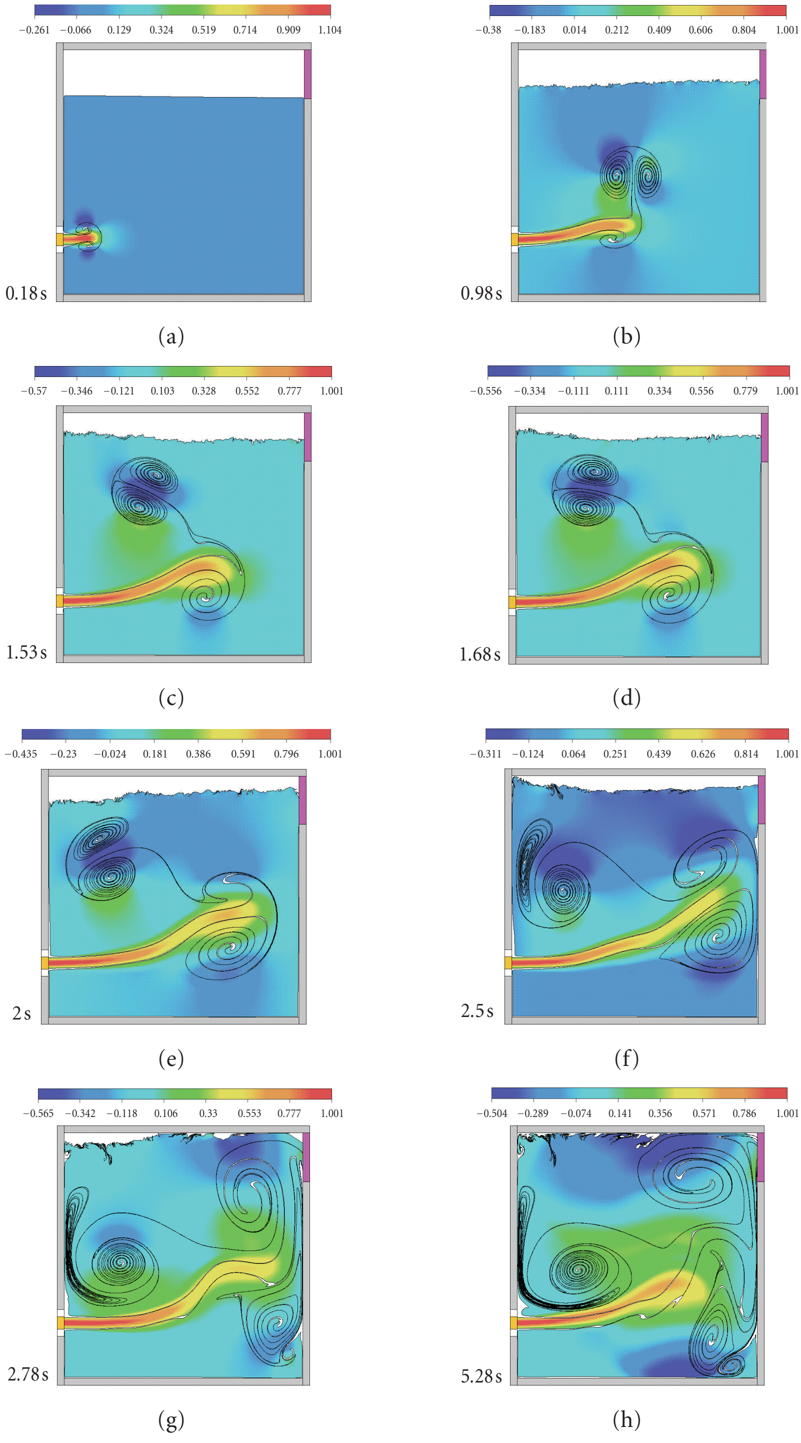


Figure 7.14. Evolution of the *u*-velocity contours of a jet in a fluid portion.

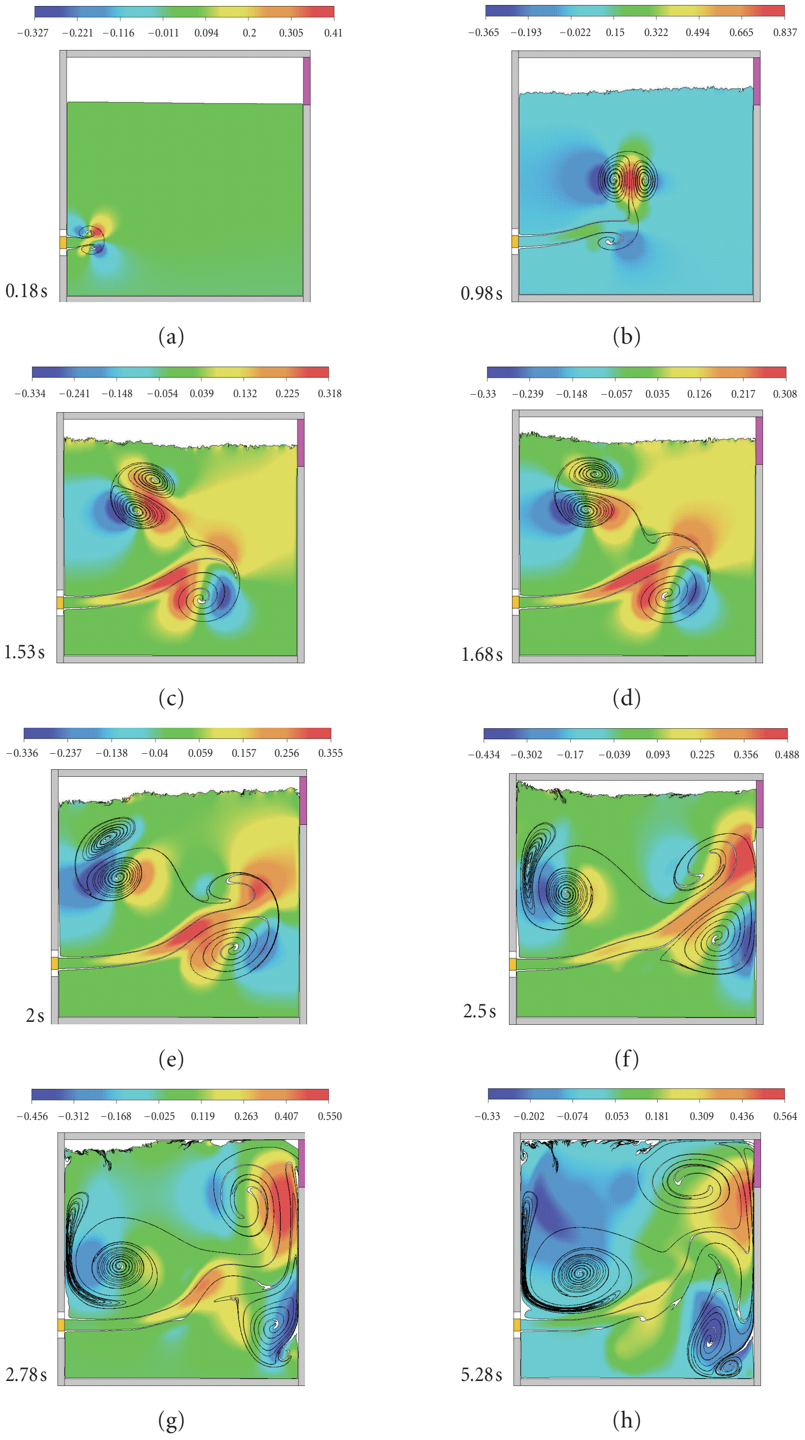


Figure 7.15. Evolution of the v-velocity contours of a jet in a fluid portion.

and vortical structures at high Froude numbers, the numerical method was applied to solve the flow of a horizontal jet penetrating a quiescent fluid from an entry port beneath the free surface.

Particularly, the best upwind schemes emerging from this study were WACEB, CUBISTA, and adaptative QUICKEST. The VONOS scheme presented poor results in the case of the internal flow over a backward-facing step (the exact fact can be seen in Table 7.1), and we believe that this is because it is not TVD; in another words, VONOS does not guarantee convergence. Although not shown here, the computed results using traditional high-order schemes (e.g., central difference, QUICK, and Lax-Wendroff) were corrupted by numerical instabilities. For the problems of a turbulent flow past a flat plate and a free jet impinging normally onto flat rigid wall, the adaptative QUICKEST scheme presented similar results to WACEB and CUBISTA schemes.

The price to pay when using the adaptative QUICKEST is that the CPU time is greater than or equal to the CPU time required by WACEB and CUBISTA. However, the adaptative QUICKEST is more flexible and more widely applicable than the others, since it is based on the local Courant number.

All the results present in this work are consistent with the previous numerical results of [5], indicating that the current *Freeflow* code is also able to predict turbulent confined and free surface fluid flows with adequate accuracy. And we believe that these results could be improved by incorporating more physics in the modeling. For this to be realized, the authors are considering adaptations for the renormalization group method (RNG) and the realizable κ-ε model.

Acknowledgments

Support for this research article was provided by agencies FAPESP, Grants 02/11774-2, 04/16064-9, 04/15185-7, 05/51458-0, and CNPq, Grants 304201/05-7 and 472030/06-0.

References

[1] B. P. Leonard, “Simple high-accuracy resolution program for convective modelling of discontinuities,” *International Journal for Numerical Methods in Fluids*, vol. 8, no. 10, pp. 1291–1318, 1988.

[2] A. Harten, “On a class of high resolution total-variation-stable finite-difference schemes,” *SIAM Journal on Numerical Analysis*, vol. 21, no. 1, pp. 1–23, 1984.

[3] A. Varonos and G. Bergeles, “Development and assessment of a variable-order non-oscillatory scheme for convection term discretization,” *International Journal for Numerical Methods in Fluids*, vol. 26, no. 1, pp. 1–16, 1998.

[4] V. G. Ferreira, M. F. Tomé, N. Mangiavacchi, et al., “High-order upwinding and the hydraulic jump,” *International Journal for Numerical Methods in Fluids*, vol. 39, no. 7, pp. 549–583, 2002.

[5] V. G. Ferreira, N. Mangiavacchi, M. F. Tomé, A. Castelo, J. A. Cuminato, and S. McKee, “Numerical simulation of turbulent free surface flow with two-equation $k-\epsilon$ eddy-viscosity models,” *International Journal for Numerical Methods in Fluids*, vol. 44, no. 4, pp. 347–375, 2004.

[6] B. Song, G. R. Liu, K. Y. Lam, and R. S. Amano, “On a higher-order bounded discretization scheme,” *International Journal for Numerical Methods in Fluids*, vol. 32, no. 7, pp. 881–897, 2000.

[7] M. A. Alves, P. J. Oliveira, and F. T. Pinho, "A convergent and universally bounded interpolation scheme for the treatment of advection," *International Journal for Numerical Methods in Fluids*, vol. 41, no. 1, pp. 47–75, 2003.

[8] M. K. Kaibara, V. G. Ferreira, H. A. Navarro, J. A. Cuminato, A. Castelo, and M. F. Tomé, "Upwinding schemes for convection dominated problems," in *Proceedings of the 18th International Congress of Mechanical Engineering (COBEM '05)*, Ouro Preto, MG, Brazil, November 2005.

[9] V. G. Ferreira, C. M. Oishi, F. A. Kurokawa, et al., "A combination of implicit and adaptative upwind tools for the numerical solution of incompressible free surface flows," *Communications in Numerical Methods in Engineering*, vol. 23, no. 6, pp. 419–445, 2007.

[10] A. Castello, M. F. Tomé, C. N. L. César, S. McKee, and J. A. Cuminato, "Freeflow: an integrated simulation system for three-dimensional free surface flows," *Journal of Computing and Visualization in Science*, vol. 2, no. 4, pp. 199–210, 2000.

[11] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, vol. 6 of *Course of Theoretical Physics*, Butterworth-Heinemann, Newton, Mass, USA, 1975.

[12] R. Peyret and T. D. Taylor, *Computational Methods for Fluid Flow*, Springer Series in Computational Physics, Springer, New York, NY, USA, 1983.

[13] D. C. Wilcox, *Turbulence Modeling for CFD*, DCW Industries, La Caada, Calif, USA, 1993.

[14] D. L. Sondak and R. H. Pletcher, "Application of wall functions to generalized nonorthogonal curvilinear coordinate systems," *AIAA journal*, vol. 33, no. 1, pp. 33–41, 1995.

[15] H. L. Norris and W. C. Reynolds, "Turbulent channel flow with a moving wavy boundary," Tech. Rep. TR TF-7, Departament of Mechanics Engineering, Stanford University, Palo Alto, Calif, USA, 1980.

[16] P. Bradshaw, Ed., *Turbulence*, vol. 12 of *Topics in Applied Physics*, Springer, Berlin, Germany, 2nd edition, 1978.

[17] D. C. Wilcox, "Reassessment of the scale-determining equation for advanced turbulence models," *AIAA journal*, vol. 26, no. 11, pp. 1299–1310, 1988.

[18] F. M. White, *Viscous Fluid Flow*, McGraw-Hill, New York, NY, USA, 1991.

[19] F. Menter and T. Esch, "Elements of industrial heat transfer predictions," in *Proceedings of the 16th International Congress of Mechanical Engineering (COBEM '01)*, Ouro Uberlndia, MG, Brazil, November 2001.

[20] B. P. Leonard, "The ULTIMATE conservative difference scheme applied to unsteady one-dimensional advection," *Computer Methods in Applied Mechanics and Engineering*, vol. 88, no. 1, pp. 17–74, 1991.

[21] P. K. Sweby, "High resolution schemes using flux limiters for hyperbolic conservation laws," *SIAM Journal on Numerical Analysis*, vol. 21, no. 5, pp. 995–1011, 1984.

[22] P. H. Gaskell and A. K. C. Lau, "Curvature-compensated convective transport: SMART, a new boundedness-preserving transport algorithm," *International Journal for Numerical Methods in Fluids*, vol. 8, no. 6, pp. 617–641, 1988.

[23] M. F. Tomé and S. McKee, "GENSMAC: a computational marker and cell method for free surface flows in general domains," *Journal of Computational Physics*, vol. 110, no. 1, pp. 171–186, 1994.

[24] V. G. Ferreira, *Análise e Implementação de Esquemas de Convecção e Modelos de Turbulência para Simulação de Escoamentos Incompressíveis Envolvendo Superfícies Livres*, Ph.D. thesis, Departament of Computer Science and Statistics, USP - University of São Paulo, São Carlos, Brazil, 2001.

[25] S. Armfield and R. Street, "An analysis and comparison of the time accuracy of fractional-step methods for the Navier-Stokes equations on staggered grids," *International Journal for Numerical Methods in Fluids*, vol. 38, no. 3, pp. 255–282, 2002.

[26] F. H. Harlow and J. E. Welch, "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface," *Physics of Fluids*, vol. 8, no. 12, pp. 2182–2189, 1965.

[27] A. Chorin, "A numerical method for solving incompressible viscous flow problems," *Journal Computational of Physics*, vol. 2, no. 1, pp. 12–26, 1967.

[28] J. Eaton and J. P. Johnston, "Turbulent flow reattachment: an experimental study of the flow and structure behind a backward-facing step," Tech. Rep. TR MD-39, Stanford University, Stanford, Calif, USA, 1980.

[29] S. Thangam and C. G. Speziale, "Turbulent flow past a backward-facing step:a critical evaluation of two-equation models," *AIAA Journal*, vol. 30, no. 5, pp. 1314–1320, 1992.

[30] A. C. Brandi, *Estratégias "Upwind" e Modelagem $k-\epsilon$ para Simulação Numérica de Escoamentos com Superfícies Livres em Altos Números de Reynolds*, Ph.D. thesis, Departament of Computer Science and Statistics, USP - University of São Paulo, São Carlos, Brazil, 2005.

[31] F. M. White, *Fluid Mechanics*, McGraw-Hill, New York, NY, USA, 1979.

[32] E. J. Watson, "The radial spread of a liquid jet over a horizontal plane," *Journal of Fluid Mechanics*, vol. 20, pp. 481–499, 1964.

V. G. Ferreira: Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP), Caixa Postal 668, CEP 13560-970 São Carlos, SP, Brazil
Email address: pvgf@icmc.usp.br

A. C. Brandi: Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP), Caixa Postal 668, CEP 13560-970 São Carlos, SP, Brazil
Email address: analice@sc.usp.br

F. A. Kurokawa: Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP), Caixa Postal 668, CEP 13560-970 São Carlos, SP, Brazil
Email address: kurokawa@icmc.usp.br

P. Seleghim Jr.: Departamento de Engenharia Mecânica, Escola de Engenharia de São Carlos, Universidade de São Paulo (USP), Caixa Postal 359, CEP 13566-590 São Carlos, SP, Brazil
Email address: seleghim@sc.usp.br

A. Castelo: Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP), Caixa Postal 668, CEP 13560-970 São Carlos, SP, Brazil
Email address: castelo@icmc.usp.br

J. A. Cuminato: Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP), Caixa Postal 668, CEP 13560-970 São Carlos, SP, Brazil
Email address: jacumina@icmc.usp.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 24627, 28 pages
doi:10.1155/2007/24627

Research Article

Dynamical Simulation and Statistical Analysis of Velocity Fluctuations of a Turbulent Flow behind a Cube

T. F. Oliveira, R. B. Miserda, and F. R. Cunha

Received 12 September 2006; Revised 18 January 2007; Accepted 6 March 2007

Recommended by José Manoel Balthazar

A statistical approach for the treatment of turbulence data generated by computer simulations is presented. A model for compressible flows at large Reynolds numbers and low Mach numbers is used for simulating a backward-facing step airflow. A scaling analysis has justified the commonly used assumption that the internal energy transport due to turbulent velocity fluctuations and the work done by the pressure field are the only relevant mechanisms needed to model subgrid-scale flows. From the numerical simulations, the temporal series of velocities are collected for ten different positions in the flow domain, and are statistically treated. The statistical approach is based on probability averages of the flow quantities evaluated over several realizations of the simulated flow. We look at how long of a time average is necessary to obtain well-converged statistical results. For this end, we evaluate the mean-square difference between the time average and an ensemble average as the measure of convergence. This is an interesting question since the validity of the ergodic hypothesis is implicitly assumed in every turbulent flow simulation and its analysis. The ergodicity deviations from the numerical simulations are compared with theoretical predictions given by scaling arguments. A very good agreement is observed. Results for velocity fluctuations, normalized autocorrelation functions, power spectra, probability density distributions, as well as skewness and flatness coefficients are also presented.

1. Introduction

In spite of considerable progress in computer technology, numerical methods, and turbulence modeling during the last several decades, reliable prediction of complex turbulent flows at high Reynolds number remains an elusive target. Turbulent flows and, in particular, wall-bounded flows exhibit wide ranges of spatial and temporal scales that increase with Reynolds number. Tsherefore, direct numerical simulation (DNS) is limited to relatively low Reynolds numbers, as the number of grid points required for DNS increases proportionally to $\mathrm{Re}^{9/4}$ [1]. Smaller timesteps result in an extra 3/4 power for the total cost scaling as Re^3. Due to the limitations of DNS, great expectations have been placed on large eddy simulation (LES) [2]. Large eddy simulation is an important technique in the study of turbulent flows. In LES, the governing equations are spatially averaged allowing the large-scale motion to be solved. On the other hand, from this averaging process, the so-called subgrid terms arise, requiring constitutive models to their evaluation [3]. LES requires less computational effort than direct numerical simulations, which attempts to solve all scales present in the turbulent flow [4]. Other important characteristic is the unsteady feature of LES. This implies that a statistical treatment is needed in order to permit an accurate characterization of the simulated turbulent flow. In addition, several theoretical studies on small-scale two-dimensional nonlocal turbulence, where the interactions of small scales with the large vortices dominate in the small-scale dynamics, have been developed in the current literature [5]. Bouris and Bergeles [6] have found that the two-dimensional large eddy simulation using a fine grid resolution, especially in the near wall region, gives a good representation of the quasi-two-dimensional mechanisms of the flow since they are directly simulated instead of being modeled as with statistical turbulence models. In addition, the two-dimensional LES performed by them has proven to be much better than any of the Reynolds-averaged Navier-Stokes (RANS) models when the major two-dimensional mechanisms of the flow and the statistical turbulence quantities are examined.

In a general case, a formal statistical treatment is based on probability averages evaluated over an ensemble of several realizations of the same process, which defines a stochastic set. For an ergodic process, the probability average can be replaced by a temporal average, and the statistical analysis is more feasible. Nevertheless, when the turbulence is dominated by large and coherent structures, typically strongly correlated, the ergodic hypothesis cannot be assumed and only a probability or statistical average (i.e., ensemble averages) should be used to describe correctly the statistical quantities of the flow [7, 8]. In an LES context, the total time of simulation needs to be long enough to ensure the ergodicity of the process and to get converging statistics.

The main goal of this paper is to perform a statistical treatment of turbulent velocity signals resulting from numerical simulations, in particular, large eddy numerical simulations (LESs). The large eddy simulations are performed for the limit of high Reynolds number (Re) and low Mach number (Ma) compressible flows. The scalings show the relative importance of the subgrid terms when the flow obeys the $\mathrm{Re} \gg 1$ and $\mathrm{Ma} \ll 1$ limits. A scaling analysis is also developed in order to estimate the deviation ε between the time average and probability average associated with the ergodicity hypothesis. In addition, from the turbulent flow over a backward-facing step simulated a long time, behavior analysis is carried out in order to quantify the integral scales for ten different positions on the flow domain. Based on this correlation time, a stochastic set is built and a statistical analysis is performed. The velocity time series of the flow are analyzed statistically using the formal probability average approach and confronted with the statistics given by the conventional time average analysis. The deviation between the two approaches is characterized by the ergodic parameter ε. We look at how long of a time average is necessary to obtain well-converged statistical results, and evaluate the mean-square difference between the time average and an ensemble average as the measure of convergence. Certainly this is an interesting question since the validity of the ergodic hypothesis is implicitly assumed in every turbulent flow simulation and its analysis. Turbulent intensities, skewness and flatness factors are also examined. All statistical quantities investigated are calculated using the probability average approach and the associated error bars are always evaluated. More recently, an extension of the ideas and of the method explored in this paper has been investigated experimentally for a three-dimensional flow [9].

2. Flow governing equations

2.1. Average equations. Let us consider a generic flow property that can be a function of space and time $\phi(\mathbf{x},t)$. The spatial average $\overline{\phi}(\mathbf{x},t)$ is defined as

$$\overline{\phi}(\mathbf{x},t) = \int_{\Omega} \phi(\mathbf{r},t) G(\mathbf{x}-\mathbf{r}) d\mathbf{r}, \tag{2.1}$$

where $\mathbf{x}$ is the position vector, $\mathbf{r}$ is the displacement vector regarding $\mathbf{x}$, and Ω denotes the volume of $\mathbf{r}$-space over which the integral is taken (i.e., space average). The function $G(\mathbf{x}-\mathbf{r})$ is a filter function $G : \mathbb{R}^3 \to [0,1]$ and satisfies

$$\lim_{|\mathbf{x}-\mathbf{r}|\to\infty} G(\mathbf{x}-\mathbf{r}) = 0, \qquad \int_{\Omega} G(\mathbf{x}-\mathbf{r}) d\mathbf{r} = 1. \tag{2.2}$$

This averaging process still regards the linearity and the commutability with the spatial and temporal derivatives

$$\overline{\phi+\psi} = \overline{\phi} + \overline{\psi}, \qquad \overline{\frac{\partial \phi}{\partial s}} = \frac{\partial \overline{\phi}}{\partial s}, \tag{2.3}$$

where $s = \mathbf{x}, t$. The properties (2.3) are derived from the continuity of ϕ and the properties of the filter function presented in (2.2) [10]. Now, a density-weighted average process is more appropriat for compressible models. This process corresponds to the well-known Favre filtered [11], defined as

$$\widetilde{\phi} = \frac{\overline{\rho\phi}}{\overline{\rho}}, \tag{2.4}$$

where ρ is the density of the fluid. Note that according to (2.4), $\overline{\rho\phi} = \overline{\rho}\widetilde{\phi}$. This identity is largely applied for averaging the governing equations shown below. The averaged mass and momentum balance equations written using index notation in the three-dimensional

Euclidian space are given, respectively, by

$$\frac{\partial \bar{\rho}}{\partial t} + \frac{\partial}{\partial x_i}(\bar{\rho}\tilde{u}_i) = 0, \tag{2.5}$$

$$\frac{\partial}{\partial t}(\bar{\rho}\tilde{u}_i) + \frac{\partial}{\partial x_j}(\bar{\rho}\widetilde{u_iu_j}) = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j}(2\mu\bar{S}_{ij}), \tag{2.6}$$

where

$$\bar{S}_{ij} = \left(\bar{D}_{ij} - \frac{1}{3}\frac{\partial \bar{u}_k}{\partial x_k}\delta_{ij}\right), \qquad \bar{D}_{ij} = \frac{1}{2}\left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i}\right). \tag{2.7}$$

Here, u_i are the components of the velocity vector $\mathbf{u}$, p is the pressure, μ is the dynamical viscosity coefficient, D_{ij} are the components of the strain-rate tensor $\mathbf{D}$, and the symbol δ_{ij} (the Kroenecker delta) denotes the components of the identity tensor. Now, since the velocity covariance is defined as $\sigma_u = \widetilde{u_iu_j} - \tilde{u}_i\widetilde{u_j}$, the averaged product of velocities that appears in the second term on the left-hand side of (2.6) can be written as $\widetilde{u_iu_j} = \tilde{u}_i\tilde{u}_j + \sigma_u$. So, the averaged momentum equation written in the indicial notation becomes

$$\frac{\partial}{\partial t}(\bar{\rho}\tilde{u}_i) + \frac{\partial}{\partial x_j}(\bar{\rho}\tilde{u}_i\widetilde{u_j}) = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j}(2\mu\bar{S}_{ij}) - \frac{\partial}{\partial x_j}(\bar{\rho}\sigma_u). \tag{2.8}$$

Defining the tensor $\Sigma_{ij} \equiv -\bar{\rho}\sigma_u$, a modified Cauchy equation can be written as follows:

$$\bar{\rho}\frac{D\tilde{\mathbf{u}}}{Dt} = \nabla \cdot \bar{\mathbf{T}}. \tag{2.9}$$

Note that we have used the Favre filtered velocity for the material derivative, $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$, namely $\overline{\rho D\mathbf{u}/Dt} = \bar{\rho}D\tilde{\mathbf{u}}/Dt$, with the constitutive equation for the stress tensor given by

$$\bar{\mathbf{T}} = -\bar{p}\mathbf{I} + 2\mu\left[\bar{\mathbf{D}} - \frac{1}{3}(\nabla \cdot \bar{\mathbf{u}})\mathbf{I}\right] + \mathbf{\Sigma}, \tag{2.10}$$

where $\mathbf{I}$ is the identity tensor. In this formulation, $\mathbf{\Sigma}$ represents the momentum transport by velocities fluctuations in the subgrid scales. The same averaging process is applied to the energy equation leading to

$$\begin{aligned}\frac{\partial}{\partial t}(\bar{\rho}\tilde{e}_T) + \frac{\partial}{\partial x_j}(\bar{\rho}\tilde{u}_j\tilde{e}_T) = {} & -\frac{\partial}{\partial x_j}(\bar{p}\tilde{u}_j) + \frac{\partial}{\partial x_j}(2\mu\tilde{S}_{ij}\tilde{u}_i) - \frac{\partial \bar{q}_j}{\partial x_j} - \frac{\partial}{\partial x_j}\underbrace{\left[\bar{\rho}(\widetilde{u_je_T} - \tilde{u}_j\tilde{e}_T)\right]}_{I} \\ & - \frac{\partial}{\partial x_j}\underbrace{(\overline{pu_j} - \bar{p}\tilde{u}_j)}_{II} + \frac{\partial}{\partial x_j}2\mu\underbrace{(\overline{S_{ij}u_i} - \tilde{S}_{ij}\tilde{u}_i)}_{III},\end{aligned} \tag{2.11}$$

where q_i are the components of the heat flux vector $\mathbf{q}$, given by the Fourier constitutive equation $\mathbf{q} = -k_D\boldsymbol{\nabla}\Im$. Here, $\Im$ denotes temperature, k_D is the thermal conductivity, and e_T is the total energy, namely, $e_T = e + \mathbf{u} \cdot \mathbf{u}/2$, with e being the internal energy. The terms on the right-hand side of (2.11) identified by II and III represent the work done by the shear stress in the subgrid scale, whereas term I is the convective transport of total energy. Term I can be decomposed into two contributions as expressed below:

$$\overline{\rho}\big(\widetilde{u_j e_T} - \widetilde{u}_j\widetilde{e}_T\big) = \underbrace{\overline{\rho}\big(\widetilde{u_j e} - \widetilde{u}_j\widetilde{e}\big)}_{IV} + \underbrace{\frac{\overline{\rho}}{2}\big(\widetilde{u_j u_k u_k} - \widetilde{u}_j\widetilde{u_k u_k}\big)}_{V}. \tag{2.12}$$

Using the equation of state for perfect gases, terms II and IV are directly related by the following expression:

$$\overline{pu_j} - \overline{p}\widetilde{u}_j = (\gamma - 1)\big(\overline{\rho e u_j} - \overline{\rho e}\widetilde{u}_j\big) = (\gamma - 1)\overline{\rho}\big(\widetilde{eu_j} - \widetilde{e}\widetilde{u}_j\big). \tag{2.13}$$

In (2.13), $\gamma = c_p/c_v$, where c_p and c_v are the specific heat at constant pressure and volume, respectively. Adding terms II and IV, the vector Q_j that represents the transport of internal energy in the subgrid scales is defined, namely,

$$Q_j = \gamma\overline{\rho}\big(\widetilde{eu_j} - \widetilde{e}\widetilde{u}_j\big). \tag{2.14}$$

For instance, the terms III, V and the vector Q_j, resulting from the averaging process of energy equation, need to be modeled in a general case.

2.2. Scaling analysis. Before discussing a model for the subgrid terms, some scaling analysis will be performed in order to evaluate the relative importance of each subgrid contribution. The scales of the large eddies are set by the geometry and the speed of the stirring mechanism, while cut-off scales of the small eddies are determined by the action of viscosity. Here, one concentrates on the small scales for a flow with large eddies of given velocity, length, and time scales U_ℓ, ℓ, T_ℓ. The important Kolmogorov microscale for the smallest eddies is based on a further assumption that the smallest eddies depend only on the rate at which energy is put into the large eddies, that is, on one particular combination of U_ℓ, ℓ. The friction only acts on the smallest scale and the energy is supplied only at the large scale. The rate of dissipation $\epsilon = 2\nu\overline{\mathbf{D}' : \mathbf{D}'}$ is measured per unit of mass, and can be related to the macroscales by assuming that a significant fraction of the kinetic energy per unit of mass $k = (1/2)\overline{\mathbf{u}' \cdot \mathbf{u}'}$ in the large eddies is dissipated in the turnover time of the large eddies, that is, per unit time,

$$\rho\epsilon = \frac{\rho U_\ell^2}{T_\ell}, \quad \text{therefore, } \epsilon = \frac{U_\ell^3}{\ell}. \tag{2.15}$$

Now, the dimensions of this dissipations per unit mass ϵ are L^2T^{-3}, while the dimensions of the kinematic viscosity ν are L^2T^{-1}. Hence, by a simple dimensional analysis, we obtain the velocity, length, time, and strain-rate scalings of the Kolmogorov microscale: $U_k = (\nu\epsilon)^{1/4}$, $\ell_k = (\nu^3/\epsilon)^{1/4}$, $T_k = (\nu/\epsilon)^{1/2}$, and $S_k = (\epsilon/\nu)^{1/2}$ [12]. Introducing the Reynolds number of the large-scale eddies $\mathrm{Re} = U_\ell\ell/\nu$, one obtains

$$\frac{U_k}{U_\ell} = \mathrm{Re}^{-1/4}, \qquad \ell_k = \ell\,\mathrm{Re}^{-3/4}, \qquad T_k = T_\ell\,\mathrm{Re}^{-1/2}, \qquad S_k = S_\ell\,\mathrm{Re}^{1/2}. \tag{2.16}$$

The Kolmogorov microscale for the smallest eddies depends on the velocity and length scales of the large eddies in the combination $\epsilon = U_\ell/\ell$. Note that the turnover time of the smallest eddies T_k is shorter than the turnover time of the large eddies T_ℓ by the factor $\mathrm{Re}^{-1/2}$. Hence, mixing takes place faster and more efficiently on small scales than on large scales. Large-scale mixing, however, is described by the Taylor diffusivity $D_T = U_\ell \ell$. So, for a container of height H, the time for eddy diffusion is then $H^2/D_T = T_\ell H^2/\ell^2$ [12].

A second microscale, the Taylor microscale, uses a different combination to yield a slightly large scale. The Taylor microscale λ_T can be thought of as the boundary layer thickness on the edge of a large eddy, that is,

$$\lambda_T = (\nu t)^{1/2} \tag{2.17}$$

with t being the turnover time of the large eddies $T_\ell = \ell/U_\ell$. Hence, using the Reynolds number of the large eddies we can show that $\lambda_T = (\nu\ell/U_\ell)^{1/2} = \ell\,\mathrm{Re}^{-1/2}$.

Now, we turn to the scaling of the turbulence energy equation (2.11). First, note that the vector Q_j can be expressed in terms of the temperature in the subgrid scale, namely,

$$Q_j = \gamma\overline{\rho}\left(\widetilde{eu_j} - \widetilde{u}_j\widetilde{e}\right) = \overline{\rho}c_p\left(\widetilde{\mathfrak{I}u_j} - \widetilde{\mathfrak{I}}\widetilde{u}_j\right). \tag{2.18}$$

This diffusion mechanism is promoted by the velocity fluctuation transport in this scale. In such case, a typical scale of these velocity fluctuations is given by $U_\ell = \sqrt{\mathbf{u}' \cdot \mathbf{u}'}$, where $\mathbf{u}' = \mathbf{u} - \widetilde{\mathbf{u}}$. The temperature fluctuations scale like U_ℓ^2. Then, a typical scale for the vector Q_j is given by

$$Q_j \sim \frac{\overline{\rho}}{(\gamma - 1)\mathfrak{I}} c^2 U_\ell^3, \tag{2.19}$$

where R is the gas constant given by Carnot's relation $R = c_p - c_v$.

The important velocity gradient occurs at the smallest scales as mentioned above. Therefore, $\widetilde{S}_{ij} \sim (\epsilon/\nu)^{1/2} \sim (U_k/\ell_k) \sim \mathrm{Re}^{1/2}\, U_\ell/\ell$, and consequently

$$\mu\left(\overline{S_{ij}u_i} - \widetilde{S}_{ij}\widetilde{u}_i\right) \sim \frac{\overline{\rho}\nu U_\ell^2}{\ell}\mathrm{Re}^{1/2}. \tag{2.20}$$

It should be important to note that using the Taylor microscale to evaluate $\widetilde{S}_{ij} \sim U_\ell/\lambda_T$ leads to the same result given in (2.20) since $\lambda_T = \ell\,\mathrm{Re}^{-1/2}$ as mentioned before, and again $\widetilde{S}_{ij} \sim \mathrm{Re}^{1/2}\, U_\ell/\ell$.

The scaling of the term V in terms of the large eddies is given by

$$\overline{\rho}\left(\widetilde{u_j u_k u_k} - \widetilde{u}_j\widetilde{u_k u_k}\right) \sim \overline{\rho}(U_\ell)^3. \tag{2.21}$$

Now, it is possible to determine the scaling for the ratio of the terms III/Q_j and V/Q_j, respectively,

$$\left|\frac{(\overline{S_{ij}u_i} - \widetilde{S}_{ij}\widetilde{u}_i)}{Q_j}\right| \sim \mathrm{Re}^{-1/2},$$
$$\left|\frac{\overline{\rho}(\widetilde{u_j u_k u_k} - \widetilde{u}_j \widetilde{u_k u_k})}{Q_j}\right| \sim \left(\frac{\mathrm{Ma}}{\mathrm{Re}}\right)^2, \tag{2.22}$$

where $\mathrm{Re} = \ell U_\ell/\nu$ is the Reynolds number, $\mathrm{Ma} = U_\ell/c$ is the Mach number, and c is the speed of the sound. The resulting scaling indicates that for high Reynolds and low Mach numbers typically for the values of these parameters investigated in the present paper, the work done by shear stress and the kinetic energy transport done by subgrid eddies are very small in comparison to the transport of internal energy Q_j. The scalings are supported by Knight et al. [13], who have evaluated subgrid terms directly by their numerical simulation. After this dimensional analysis, we can say that the only two terms to be modeled for the limit $\mathrm{Re} \gg 1$ and $\mathrm{Ma} \ll 1$ are the subgrid stress tensor Σ_{ij} and the subgrid internal energy transport vector Q_j.

2.3. Constitutive relations for the remaining subgrid terms. The constitutive equation used to describe the subgrid stress tensor is the well-known Smagorinsky model [14],

$$\mathbf{\Sigma} = 2\mu_t \widetilde{\mathbf{D}}, \tag{2.23}$$

where the nonlocal turbulent viscosity μ_t is calculated under conditions of inertial equilibrium subrange of turbulence [12], namely,

$$\mu_t = \overline{\rho}(C_S \Delta\lambda)^2 \widetilde{\gamma}. \tag{2.24}$$

Here $\widetilde{\gamma}$ is the average shear rate defined as $\widetilde{\gamma} = (2\widetilde{\mathbf{D}} : \widetilde{\mathbf{D}})^{1/2} \sim \epsilon/k$. The filter width $\Delta\lambda \sim k^{3/2}/\epsilon$ is set equal to $2\delta_g$, where δ_g is the grid spacing. It is an indication that the smallest eddies are represented by two grid points. Note that $\mu_t \sim k^2/\epsilon$. The factor C_S is known as Smagorinsky's constant. Several values have been proposed for this constant ranging from 0.1 to 0.2 [15, 16]. In the present work, one has used $C_S = 0.20$, as suggested by Deardorff [16]. Thus, the model for the subgrid stress tensor takes the form

$$\mathbf{\Sigma} = 2\overline{\rho}(C_S \Delta\lambda)^2 \widetilde{\gamma}\widetilde{\mathbf{D}}. \tag{2.25}$$

The subgrid internal energy transport tensor Q_j is related to the diffusion of temperature in the subgrid scales due to velocities fluctuations and may be modeled as being a diffusive heat transport given by a nonlocal Fourier law in the form

$$Q_j = -k_t \frac{\partial \widetilde{\mathfrak{I}}}{\partial x_j}, \tag{2.26}$$

with the nonlocal turbulent heat conductivity k_t written in terms of a turbulent Prandtl number Pr_t [17],

$$k_t = \frac{c_p}{\mathrm{Pr}_t}\mu_t. \tag{2.27}$$

For the edge of a turbulent boundary layer, $\mathrm{Pr}_t = 0.6$ [18] and this value has been used in the simulations. The set of governing equations is made nondimensionalized by using a characteristic length and velocity ℓ and U_ℓ, respectively, and the properties of the nondisturbed flow. From this point through all over the work, we will omit any superscript notation and assume that all properties are dimensionless averaged quantities. The set of dimensionless governing equations simulated consists of the continuity, written such as in (2.5), and the momentum and energy-averaged equations given in dimensionless terms by

$$\begin{aligned}\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_j u_i) &= -\frac{\partial}{\partial x_i} + \frac{1}{\mathrm{Re}}\frac{\partial}{\partial x_j}\left[2(\mu+\mu_t)S_{ij}\right],\\ \frac{\partial}{\partial t}(\rho e_T) + \frac{\partial}{\partial x_i}(\rho e_T u_i) &= -\frac{\partial}{\partial x_i}(p u_i) + \frac{1}{\mathrm{Re}}\frac{\partial}{\partial x_i}(2\mu S_{ij})\\ &\quad + \frac{1}{(\gamma-1)\,\mathrm{Pr}\,\mathrm{Ma}^2\,\mathrm{Re}}\frac{\partial}{\partial x_i}\left[(k+k_t)\frac{\partial \Im}{\partial x_i}\right],\end{aligned} \tag{2.28}$$

where Pr is the Prandtl number, $\mathrm{Pr} = c_p\mu/k_D$.

3. Statistical analysis

As mentioned before, the main goal of this work is to treat statistically turbulent velocity signals either from numerical simulations or experimental observations. An important question addressed here is to look at how long of a time average is necessary to obtain well-converged statistical results. Indeed, we have looked at the difference between the time average and an ensemble average as the measure of this convergence.

The flow is considered as a stochastic process given by the family of functions $u = u(t,\alpha)$, where $\alpha = 1,\dots,N$ are the realizations of the process according to Figure 3.1, and in the present context, $u = u(t,\alpha)$ denotes the velocity of the flow. By stationary we mean that the form of the probability distribution functions does not depend on a shift of the time origin. More precisely, we say that a random process is stationary when the probability distribution of the stochastic processes $u(t,\alpha)$ and $u(t_o + t,\alpha)$ are the same for any t_o. For a stationary random process then, we may, in principle at least, determine the various probability distributions from the observations of $u(t)$ for one realization of the system over a long period of time T. This time being much longer than the integral scale Θ (i.e., velocity fluctuation correlation time). This long-time record can be cut up into pieces of length T_λ (where T_λ is much longer than any periodicities occurring in the process), and these pieces may be treated as observations of different realizations of the system in an ensemble of similarly prepared systems. We will restrict the discussion to fluctuating quantities that are statistically steady, so that their mean values and variance are not function of time. Only under this condition does the idea of a time average make sense. In a

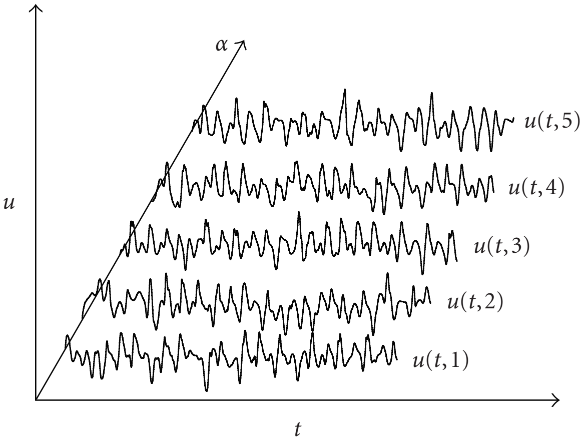


Figure 3.1. Ensemble of the stochastic process $u(t,\alpha)$. Each temporal series α_i denotes one realization (or experiment) in a given point of the flow domain. The plot illustrates realizations of the flow.

typical laboratory situation, the fluctuating velocity $u(t,\alpha)$ should be the streamwise velocity component measured in a wind tunnel behind a cube. In particular, the relative amount of time that $u(t,\alpha)$ spends at various levels is measured.

The underlying assumption here is the so-called ergodic hypothesis which states that for a stationary random process, a large number of observations made on a single system at N arbitrary instants of time have the same statistical properties as observing N arbitrarily chosen systems at the same time from an ensemble of similar systems. In dealing with general stochastic process, there are two types of mean values that can be evaluated. One is the probability average obtained by a number sufficiently larger (N) of observations at some fixed time t, denoting this average by $\langle u(t)\rangle$, and the other is the time average made for a function $u(t)$, denoting this average by $\overline{u}$. The requirement that a time average should converge to a mean value, that is, that the error should become smaller as the integration time T increases, and that the mean value found this way should always be the same, is the ergodicity [19]. This point is discussed in more details in Section 3.2.

In the case of an ergodic stationary random process, both averages yield the same result, provided that the function u is finite and continuous in mean square [20]. The time average $\overline{u}$ over a sufficiently long realization α_o (i.e., for time much longer than the velocity fluctuation correlation time) of the flow is defined as [21]

$$\overline{u} = \lim_{T\to\infty} \frac{1}{T}\int_{t_o}^{t_o+T} u(t)dt. \tag{3.1}$$

The use of time averages corresponds to the typical laboratory situation, in which measurements are taken at fixed locations in a statistically steady, but often inhomogeneous, flow field. For a time average to make sense, the integral (3.1) has to be independent of t_o. In other words, the mean flow has to be stationary $\partial\overline{u}/\partial t = 0$, and consequently the mean value of the velocity fluctuations, $u' = u(t) - \overline{u}$, itself is zero by definition. Here, the averaging time T needed to measure mean values depends on the accuracy desired as discussed in Section 3.2.

Now, whether each realization has the same probability to occur, a statistical (or probability) average is defined as being [22]

$$\langle u(t)\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{\alpha=1}^{N} u(t,\alpha). \tag{3.2}$$

A fluctuation about the probability average is defined as being $u'(t,\alpha) = u(t,\alpha) - \langle u(t)\rangle$. The variance (or the turbulent kinetic energy $(1/2)\langle u'(t)^2\rangle$) is calculated from the probability average

$$\langle u'^2(t)\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{\alpha=1}^{N}\left[u(\alpha,t) - \langle u(t)\rangle\right]^2 \tag{3.3}$$

as being the probability average of the square of the velocity fluctuation. While the time average of the square of the velocity fluctuations is given by

$$\overline{u'^2} = \lim_{T\to\infty}\frac{1}{T}\int_{t_o}^{t_o+T} u'^2(t)dt. \tag{3.4}$$

3.1. Correlation function and the spectral density. The random processes that do occur often in flow applications are those where $u'(t)$ and $u'(t')$ will be correlated at least for small values of $\tau = |t' - t|$. There are two more functions associated with a continuous stationary random process that is central to a statistical description. These two functions are the velocity fluctuation autocorrelation function and the spectral density. The normalized velocity fluctuation autocorrelation function of a continuous stationary random process is defined as [23]

$$R(t,t') = \frac{\langle u'(t)u'(t')\rangle}{\langle u'(t)^2\rangle}. \tag{3.5}$$

Now, since a shift in the origin of time does not affect any of the statistical properties of a stationary random process, the probability density functions simplify from $f(u',t)$ and $g(u'_t,t;u'_{t'},t')$ to $f(u')$ and $g(u'_t,u'_{t'},t-t')$, respectively. Here, $f(u',t)du'$ is called the first probability distribution and $g(u'_t,t;u'_{t'},t')du'_t du'_{t'}$, the second probability distribution, is the joint probability of finding $u'(\alpha,t)$ between u'_t and $u'_t + du'_t$ at time t and between $u'_{t'}$ and $u'_{t'} + du'_{t'}$ at time t'. In this particular case, when both the probability average and the normalized autocorrelation function do not vary with a shift in the origin of time, we simply write $\langle u\rangle$, $\langle u'^2\rangle$, and $R(\tau)$ and the process is said to be statistically stationary, with the time shift $\tau = t' - t$.

The other central function for the statistical analysis here is the spectral density of u'. It is well known that the variance of the process $\langle u'^2\rangle$ corresponds to the average power dissipated in the interval $(-T,T)$. Then

$$\langle u'^2\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} E(\omega)d\omega, \tag{3.6}$$

where $E(\omega)$ is the power spectrum of $u'(t)$. Thus, $E(\omega)d\omega/2\pi$ is the average power dissipated with frequencies between ω and $\omega + d\omega$. The velocity fluctuation autocorrelation function and the spectral density are connected by the Wiener-Khintchine theorem which states that [7]

$$\begin{aligned} C_\tau(\tau) = \langle u'(t+\tau)u'(t)\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty} E(\omega)e^{i\omega\tau}d\omega, \\ E(\omega) &= \int_{-\infty}^{\infty} C_\tau(\tau)e^{-i\omega\tau}d\tau. \end{aligned} \tag{3.7}$$

Note that the autocorrelation function is related with the autocorrelation coefficient by the variance, that is, $C_\tau(\tau) = \langle u'^2\rangle R(\tau)$. Thus, according to the above theorem, the correlation function and the spectral density are simply Fourier transforms of each other.

3.2. Ergodicity. As stated before in this work, we look at how long of a time average is necessary to obtain well-converged statistical results. For this end, we need to evaluate the mean-square difference between the time average and an ensemble average as the measure of convergence. This is an interesting question since the validity of the ergodic hypothesis is implicitly assumed in every turbulent flow simulation and its analysis. Thus, using the definition of correlation function given in (3.5), one obtains [7, 24]

$$\sigma^2(T) = \overline{(\overline{u} - \langle u(t)\rangle)^2} = \frac{\overline{u'^2}}{T^2}\iint_0^T R(t'-t)dt\,dt' \frac{2\langle u'^2\rangle}{T}\int_0^T \left(1-\frac{\tau}{T}\right)R(\tau)d\tau. \tag{3.8}$$

Equation (3.8) is an important result that relates the correlation function with the variance σ^2. Note that if $T \to \infty$ leads to $\sigma^2 \to 0$, this implies the ergodicity condition. Thus, the mean value of the fluctuating quantity can be determined by a time average with accuracy defined by the size of the integral time scale Θ. In fact, this is the requirement that a time average should converge to a mean value with the error becoming smaller as the integration time T increases. An ergodic variable not only becomes uncorrelated with itself at large time step, but it also becomes statistically independent of itself. An equation like (3.8) may be also used to evaluate the mean-square error of the difference between the average value of $u(t)$ in the laboratory (evolving finite integration time) and the true mean value (requiring integration over an infinitely long time). Usually, in a process in which the correlation function decays rapidly for a relatively short time τ, the ergodic condition is verified. In particular, for a turbulence in which the correlation function has the same exponential decay of the one corresponding to a random walk process, we have

$$R(\tau) \sim e^{-\tau/\Theta}, \tag{3.9}$$

where Θ is a correlation time associated to an interval in which the events are weakly correlated. For large time intervals compared to Θ, the flow $u(t)$ becomes statistically independent of itself, so that Θ is a measure for the time interval over which $u(t)$ remembers its past history.

In a Markovian diffusion process, the variance in the displacement from the starting position increases linearly in time. The coefficient of the linear growth is defined to be the diffusivity of the random walk $\Gamma = (1/2)d/dt(\langle x^2(t)\rangle)$. Again, the symbol $\langle\cdot\rangle$ denotes an average over several experiments.

Now, proceeding formally, Taylor's calculation of the eddy diffusivity in turbulence is given by

$$\Gamma = \frac{d}{dt}\left(\frac{1}{2}\langle x^2(t)\rangle\right) = \langle x(t)\dot{x}(t)\rangle = \int_0^t \langle \dot{x}(t')\dot{x}(t)\rangle dt'. \tag{3.10}$$

The diffusivity attains its constant value only after several correlation times. Thus, $\Gamma = \langle {u'}^2\rangle\Theta$, with $\langle {u'}^2\rangle$ being the mean square of the velocity fluctuations and Θ being the integral time scale,

$$\Theta = \int_0^\infty \frac{\langle \dot{x}(\tau)\dot{x}(0)\rangle d\tau}{\langle {u'}^2\rangle} = \int_0^\infty R(\tau)d\tau. \tag{3.11}$$

By the present analysis, we can show that when the above integral fails to converge, due to the slow decay of the velocity fluctuations autocorrelation, the diffusion becomes anomalous with $\langle x^2(t)\rangle \sim t^n$, with $n \neq 1$ [25, 26]. This behavior is characteristic in large-scale region of the turbulent flow investigated where these scales are strongly correlated, and the turbulence does not loose its memory. It is instructive to remind the reader that the velocity correlation needs to be computed as seen by a particle moving with the fluid, and hence Θ is called the Lagrangian integral-correlation time. The length scale ℓ introduced in Section 2 as the size of the eddies can now be defined as $\ell = u'\Theta$, that is, the distance one would move at $(\langle u'^2\rangle)^{1/2}$ (i.e., the root mean square of the velocity fluctuation or simply the RMS) during the correlation time Θ.

Now, the magnitude of the error associated with the nonergodicity (i.e., the convergence of time average) of the process is defined as being

$$\varepsilon = \frac{\sigma}{\langle u\rangle}. \tag{3.12}$$

Considering an exponential decay for the correlation function, the integral in (3.8) may be performed to give an estimation of ε, namely,

$$\varepsilon^2 \sim \frac{2\langle u'^2\rangle\Theta}{T\langle u\rangle^2} = \frac{2I^2\Theta}{T}, \tag{3.13}$$

where $I = \sqrt{\langle u'^2\rangle}/\langle u\rangle$ is the turbulence intensity of the flow that is a measure of the relative importance between the velocity fluctuation and the mean flow. It is clear from (3.13) that the convergence of the time average to a mean value can be determined to any accuracy desired if the integral scale Θ is finite. In particular, (3.13) gives a good estimation of the long time T needed to verify the ergodicity of the flow. The time average should also be used as a correct approach to describe the process from a statistical point of view. In this work, the result expressed in (3.13) is tested by direct evaluations of the variance σ^2.

As mentioned before, an important quantity to quantify flow memory is the Taylor time scale [23]. Using a Taylor series to expand $u'(t+\tau)$ in a neighborhood of t and supposing the process to be stationary, the correlation function based on an ensemble average may be written as

$$R(\tau)=\frac{\langle u'(t)u'(t+\tau)\rangle}{\langle u'^2\rangle}=1-\frac{\tau^2}{2!}\frac{1}{\langle u'^2\rangle}\left\langle\left(\frac{\partial u'}{\partial t}\right)^2\right\rangle+O(\tau^3), \tag{3.14}$$

or

$$R(\tau)\sim 1-\frac{\tau^2}{\lambda_T^2},\quad \text{where } \lambda_T=\left[\frac{2\langle u'^2\rangle}{\langle(\partial u'/\partial t)^2\rangle}\right]^{1/2}. \tag{3.15}$$

From (2.17) and (3.15), it is clear that λ_T is a short time scale of the correlation process. So, using a second-degree polynomial function in order to fit the correlation function for short times, λ_T may be estimated. Typically, the Taylor scale is larger than a dissipative time scale, but is not related to the integral scale observed in the macroscopic flow, that is, $\ell_k^2/\nu \ll \lambda_T \ll \ell/U_\ell$, where ℓ_k is the Kolmogorov dissipative length scale as defined in Section 2. Effectively, the Taylor scale is a memory characteristic time of the flow. If t is the present time, we can say that the flow has a strong dependence on the events that occur in the interval $(t-\lambda_T, t)$.

4. Numerical simulations

We now briefly summarize the sequence of steps that is necessary to perform our numerical simulations. Large eddy simulations admittedly require denser grids and more computer time than Reynolds averaged approaches, but in certain cases, such as the one under consideration here, it seems that the Reynolds averaged approaches fail to predict important statistical aspects of the flow. The purpose of this section is to provide an accurate simulation of the turbulent flow past a backward-facing step using a large eddy simulation in two dimensions. We have therefore obtained information on the statistics of the flow from the velocity time series generated from the simulations according to the procedure described in Section 3. We will show that certain region of the flow corresponding to the formation of coherent large-scale structure cannot be described by the well-converging time average approach.

The flow investigated was numerically simulated under a two-dimensional large eddy fashion. We use a finite-volume method on a two-dimensional Eulerian grid to solve hydrodynamic and energy equations of the flow in Cartesian coordinate for a structured mesh with colocalized variables. The governing equations are solved simultaneously. The Reynolds and Mach numbers based on the step height were Re = 38000 and Ma = 0.03, respectively. Figure 4.1 shows the flow domain and the location of the analyzed points in the flow and typical streamlines of the mean turbulent field. The streamwise and spanwise lengths of the computational domain were $20H$ and $2.5H$, respectively. The filtered governing equations described in Section 2 were discretized by using the explicit MacCormack method written for a finite volume formulation [27]. The numerical method uses a standard explicit predictor-corrector Euler algorithm to carry the temporal march. The set of governing equations was discretized using forward first-order differences in the predictor steps and backward first order differences in the corrector steps. In both

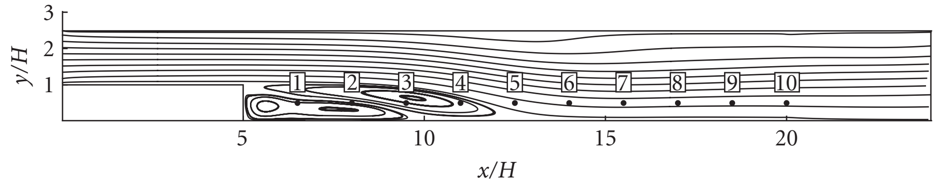


Figure 4.1. Flow domain and the indication of the velocity probes' position inside (probes 1–4) and outside (probes 5–10) the recirculating bubble. The probes are equally spaced with $\Delta(x/H) = 1.5$ from $x/H = 6.5$ (probe 1) to $x/H = 20$ (probe 10). The height of all probes is $y/H = 0.5$, with $H = 5.08$ cm and $U_\infty = 11.63$ m/s.

steps, the strain rate and the temperature gradient components were evaluated at the finite volume faces by central differences. We summarize next the basic procedure of the discretization process.

4.1. Discrete approximation of the balance equations. The general framework for the filtered balance equations described in Section 2 can be written in the form of the following vector equation:

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{\Pi} = 0, \tag{4.1}$$

where $\mathbf{\Pi} = \mathbf{E}\hat{\mathbf{e}}_1 + \mathbf{F}\hat{\mathbf{e}}_2$, and the vectors $\mathbf{U}$, $\mathbf{E}$, and $\mathbf{F}$ are defined in the same way as given by Anderson et al. [28],

$$\mathbf{U} = \begin{bmatrix} \rho \\ \rho u_1 \\ \rho u_2 \\ \rho e_T \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} \rho u_1 \\ \rho u_1^2 + p - C_1(\mu + \mu_t)S_{11} \\ \rho u_1 u_2 - C_1(\mu + \mu_t)S_{12} \\ \rho u_1 e_T + p u_1 - C_1(\mu S_{11})u_1 - C_1(\mu S_{12})u_2 C_2(k + k_t)\partial\Im/\partial x_1 \end{bmatrix}, \tag{4.2}$$

$$\mathbf{F} = \begin{bmatrix} \rho u_2 \\ \rho u_2 u_1 - C_1(\mu + \mu_t)S_{21} \\ \rho u_2^2 + p - C_1(\mu + \mu_t)S_{22} \\ \rho u_2 e_T + p u_2 - C_1(\mu S_{21})u - C_1(\mu S_{22})u_2 - C_2(k + k_t)\partial\Im/\partial x_2 \end{bmatrix}.$$

Here the subscripts 1 and 2 in the above vectors denote the components x_1 and x_2 of a flow quantity, $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ are the unit basis vectors in directions 1 and 2, respectively, and the parameters $C_1 = 2\,\mathrm{Re}^{-1}$ and $C_2 = [(\gamma - 1)\,\mathrm{Pr}\,\mathrm{Ma}^2\,\mathrm{Re}]^{-1}$.

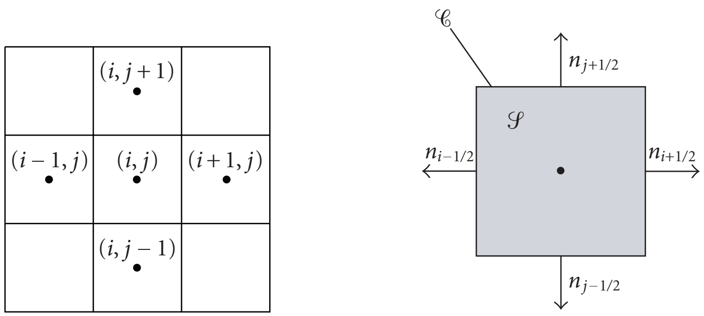


Figure 4.2. A two-dimensional sketch of the computational structured quadrangular grid used in the numerical simulations. The notation used is shown in the figure.

Now, consider that the flow domain Ω is subdivided in $N_{\mathscr{S}}$ small regions called control volumes, so that $\Omega = \bigcup_{i=1}^{N_{\mathscr{S}}} \mathscr{S}_i$. The volume average of the quantity $\mathbf{U}$ over a single control volume $\mathscr{S}$ is defined as being $\hat{\mathbf{U}} = (1/\mathscr{S}) \int_{\mathscr{S}} \mathbf{U}\, dS$. Now, integrating (4.1) over $\mathscr{S}$ and using the divergence theorem with the volume average definition, it is found that

$$\frac{\partial \hat{\mathbf{U}}}{\partial t} = -\frac{1}{\mathscr{S}} \int_{\mathscr{C}} \mathbf{\Pi} \cdot \mathbf{n}\, d\mathscr{C}, \tag{4.3}$$

where $\mathscr{C}$ is the contour of the elementary region $\mathscr{S}$ and $\mathbf{n}$ is a unit vector directed outward from the enclosed control volume. The flow domain Ω was discretized by using quadrangular control volumes in a structural mesh as illustrated in Figure 4.2. The surface integral (4.3) is evaluated by line integrals over the edges of the control volumes. The value of $\mathbf{\Pi}$ on an elementary volume edge is taken as being the volume average of $\mathbf{\Pi}$ over one of the control volumes bounded by this edge. In particular, it is considered that $\hat{\mathbf{\Pi}}$ is always constant with respect to the integral over $\mathscr{C}$, so that $\int_{\mathscr{C}} \mathbf{\Pi} \cdot \mathbf{n}\, d\mathscr{C} \approx \sum_{\beta=1}^{4} \hat{\mathbf{\Pi}}_{\beta} \cdot \mathscr{S}_{\beta}$, where $\mathscr{S}_{\beta}$ is a vector normal to the edge β with absolute value equal to the length of this edge. Under this condition, (4.3) reduces to the following approximation:

$$\frac{\partial \hat{\mathbf{U}}}{\partial t} \approx -\frac{1}{\mathscr{S}} \sum_{\beta=1}^{4} \hat{\mathbf{\Pi}}_{\beta} \cdot \mathscr{S}_{\beta}. \tag{4.4}$$

Since $\hat{\mathbf{\Pi}}$ is a function of $\hat{\mathbf{U}}$, (4.4) can be solved by an Euler method. In the MacCormack method, a predictor-corrector algorithm corresponding to a second-order Runge-Kutta procedure is applied. In the predictor step, the vector $\hat{\mathbf{\Pi}}$ was taken as being equal to the vector of the control volume downstream to the edge, whereas in the corrector step, $\hat{\mathbf{\Pi}}$ was associated to the volume upstream to the same edge. For instance, following the notation in Figure 4.2, consider the discrete quantity evaluated on the edge $i + 1/2$. At the predictor step, $\mathbf{\Pi}$ is calculated in terms of the quantities of the volume $(i+1, j)$. Subsequently, for the same edge, at the corrector step, $\mathbf{\Pi}$ is calculated in terms of the quantities of the volume (i, j). The components of the velocity gradient tensor and the temperature gradient on the faces of the control volume were evaluated by central difference. For instance, on the edge $i+1/2$, the term $\partial u_1/\partial x_1$ is approximated by

$$\frac{\partial u_1}{\partial x_1} \approx \frac{u_1(i+1,j)-u_1(i,j)}{x_1(i+1,j)-x_1(i,j)}, \tag{4.5}$$

where $x_1(i,j)$ and $x_1(i+1,j)$ denote the center coordinates of the control volumes (i,j) and $(i+1,j)$, respectively.

The compressible formulation of the flow allows the use of a state equation for the pressure. Consequently, additional velocity-pressure coupling algorithm was not required. This numerical procedure leads to a second-order precision discretization in both space and time derivatives. The accuracy and robustness of the numerical method are defined in the context of the classical MacCormack method described in details by [27]. No extra upwinding feature was implemented and the method is stable if the CFL number (i.e., Courant-Friedricks-Lewey number) is less than unity for all grid volumes, where CFL $= \Delta x/(u+c_s)$ with u and c_s being the largest velocity norm and the largest sound velocity at the volume boundary, respectively, and Δx is the local grid spacing [28]. In the present simulation, CFL ≈ 0.7, which has provided a stable condition for all simulations. The initial transient, corresponding to 608 flows through (i.e., 10^6 iterations), that is, 2.3 seconds of physical time, was neglected. The typical convective time scale of the flow $H/U_\infty \approx 0.1$ second and the longest simulation time was approximately 21 seconds. This wide interval was necessary because the velocity signal was fragmented into smaller temporal series when defining the stochastic set.

An equally spaced Cartesian grid was used to discretize the flow domain and to resolve the velocity, pressure, and temperature fields of the flow. The spatial resolution employed was 36 volumes/cube edge. This resolution was sufficient to resolve the turbulent length scales of interest. Additionally, a $100H$ stretched grid region was generated with 188 volumes in the streamwise direction, downstream of the regular grid. This region was necessary to dissipate the turbulent structures and provide a smooth condition at the outgoing section of the domain. The computations were carried out for a typical rectangular grid with (90×1088) control volumes. A uniform velocity profile at the inlet section was imposed with no turbulence intensity (i.e., laminar flow). No slip boundary conditions were employed in the spanwise and normal directions. The initial condition is formed by stagnated fluid with constant pressure and temperature. The dimensionless time step used was approximately $\Delta t = 5\times 10^{-3}$. The simulation time was $T = 48800H/U_\infty$. The velocity samples at probe positions were stored in intervals of $25\Delta t$. In postprocessing algorithm, the velocity time series have been used to compute the various statistical results to be presented next. The time separation between two sequential data fields was large enough compared to the integral time scale of the turbulent fluctuation Θ for the data fields to be considered are nearly independent realizations of the flow. The flow statistics have been obtained by or ensemble averaging over all stored data fields. The simulations all were carried out on a PC of 2.0 GHz processor and 1.0 GB of physical memory. The total CPU time required to perform the 2D LES simulation was about 27 days.

Figure 4.3 shows a typical evolution of the instantaneous vorticity field around the cube given by our two-dimensional LES. A complex turbulent wake downstream the cube is seen, with large-scale vortices of different intensities interacting along the wake. In this

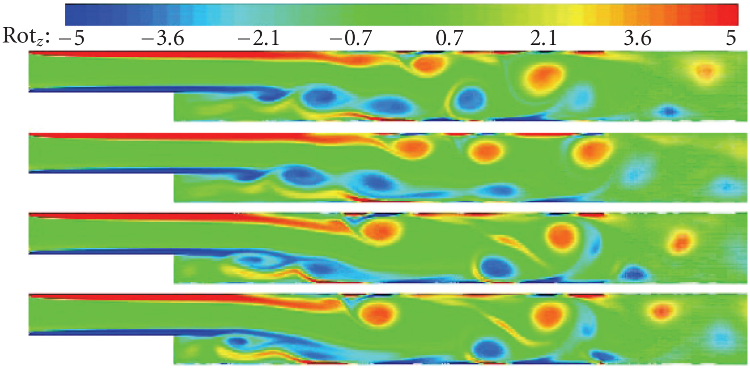


Figure 4.3. A typical time evolution of the instantaneous vorticity field obtained in the present work by using two-dimensional LES. From top to bottom, the associated dimensionless times are $tU_\infty/H =$ 448.3, 474.7, 501.1, and 527.5. The color scale indicates the vorticity intensity, increasing from left (blue) to right (red). The label Rot$_z$ means the component "z" of the vorticity or rotational of the velocity field (perpendicular to the plane of the flow).

region, the velocity signals were collected in order to build the stochastic set to be analyzed. In addition, we can see the flow separation and that a reattachment occurs more suddenly. The converged solutions were checked against the experimental results for the average reattachment point position given by Eaton and Johnston [29]. This is a typical parameter that has been commonly used to validate numerical simulations. The test simulation was carried out on the same conditions of the experimental setup, including flow domain geometry, thermodynamic properties of the fluid, and the imposed flow. Figure 4.4 depicts the dimensionless component of the average velocity in x-direction as a function of the position x/H. A very good agreement is seen between the reattachment length $x_R/H = 12.94$ predicted by the present numerical simulations and $x_R/H = 12.95$ measured experimentally [29]. The error was less than 1%. Halving the grid size produced a change that was not greater than 1% in this computed quantity.

The first step for the statistical characterization of the flow was to define the stochastic set in the probes positions. In order to build a stochastic process from the numerical simulations, a large temporal series was dropped into smaller temporal series corresponding to the realizations of the process. The resulting temporal series were then independent events of the turbulent flow, since the time scales involved were long enough for a complete decay of the correlation function. So, the velocity fluctuations become statistically independent with respect to events that have occurred in their past history. This procedure was equivalent to start a new simulation from a different initial condition which is uncorrelated with the past one. The dimensionless integral time scale $\Theta U_\infty/H$ may be determined by direct integration of the velocity fluctuation normalized autocorrelation function (3.11). In this work, however, we have estimated this parameter by direct inspection of the velocity fluctuation autocorrelation function as being approximately two times the time for the full decay of the autocorrelation function. This of course overestimates the integral scale value of the correlation time, but it corresponds to a suitable time

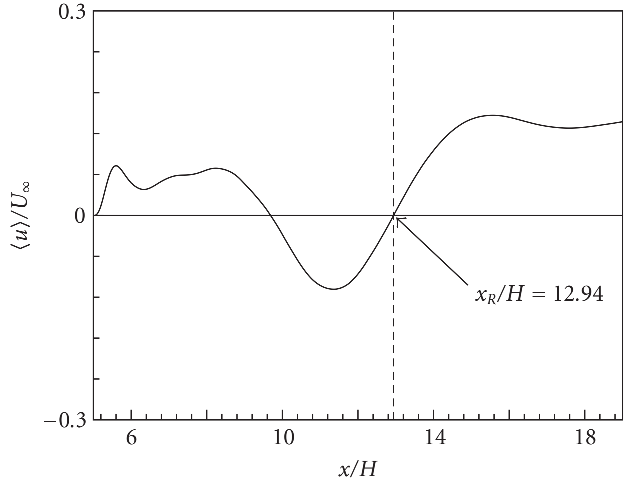


Figure 4.4. Dimensionless average velocity in the x-direction as a function of the dimensionless position x/H (solid line). The vertical dashed line indicates the reattachment length ($x_R/H = 12.5$) measured experimentally [29], whereas the arrow in the plot indicates the reattachment length ($x_R/H = 12.4$) given by the present work for CFL = 0.7 and $C_S = 0.27$.

scale which guaranteed the statistical independence of the set of realizations constructed from the original velocity signal.

5. Results and discussions

Figures 5.1(a) and 5.1(b) show the normalized velocity fluctuation autocorrelation function as a function of the nondimensional time for probes in two different dimensionless positions $x = 4H$ and $x = 8H$. The standard errors are calculated in terms of the standard deviation of the normalized autocorrelation function values which result from the statistics over the set of realizations, at each fixed time. The same procedure was applied to evaluate all values of error bars presented in this work. We can see a remarkable difference between the integral time scales for the probes 4 and 8 located inside and outside the recirculating bubble, respectively, as illustrated in Figure 4.1. While for the probe 4 (see Figure 5.1(a)), the dimensionless integral time scale was estimated to be $\Theta U_\infty/H \approx 500$, in probe 8 the corresponding time scale was $\Theta U_\infty/H \approx 10$ as displayed in Figure 5.1(b). The temporal series in the probe positions were divided into smaller time intervals corresponding to different independent experiments. Table 5.1 shows the length of these intervals $\Theta U_\infty/H$ and the number of realizations associated with each probe. By this procedure, the statistics of the flow were performed in terms of the probability averages. Figures 5.2(a), 5.2(b), 5.3(a), and 5.3(b) give the nondimensional average velocity in the x-direction (streamwise) and its associated root mean square statistics for the probes 4 and 8, respectively. We can see temporal oscillations in the values of the mean velocity fluctuation and velocity fluctuations that are more intense in the probe 4. These oscillations are a direct consequence of the difference between the temporal and probability averages, which is related to a nonergodic behavior of the turbulence in a region of

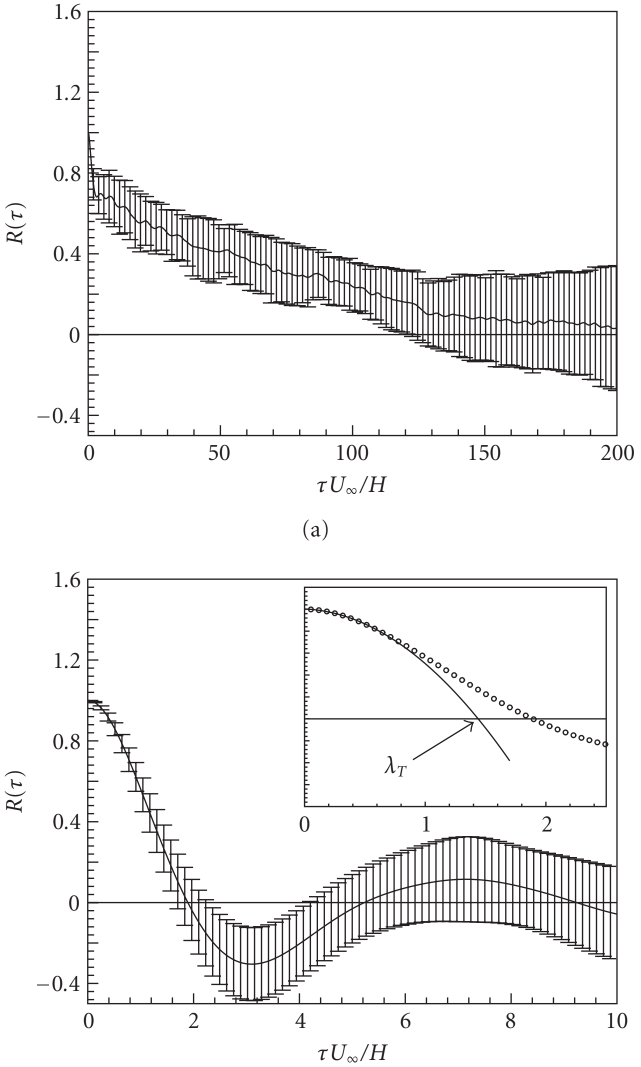


Figure 5.1. The normalized velocity fluctuation autocorrelation function (streamwise component) plotted as a function of the dimensionless time. The numerical results were obtained from LES. (a) Probe 4; (b) probe 8. The insert in the plot shown in (b) gives an estimation of the Taylor microscale λ_T by using the parabolic fit (3.15). Attempt to different time scales used in the plots. The error bars are also shown in both plots.

large-scale structures located in the transition shear layer region of the flow. Actually, the probes 4 and 5 are located in a place which appears to be more critical for the ergodicity than probes 1, 2, 3 that are exactly inside the recirculating region. Inside this region, the fluid parcels seem to describe approximately a rigid body rotation since the important

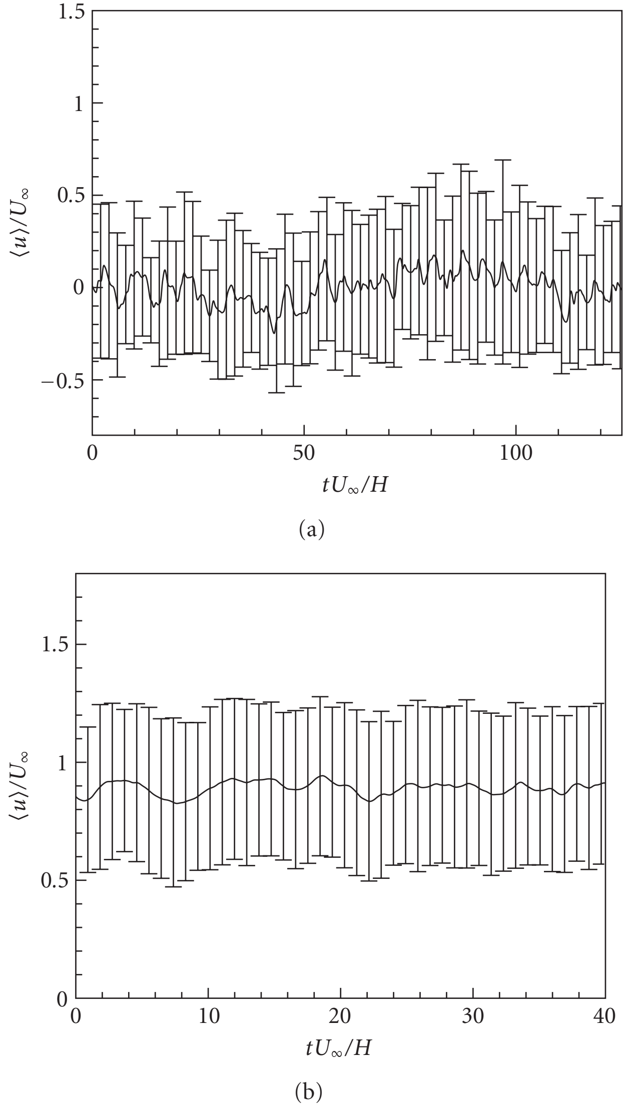
Table 5.1. The values of the integral time scale and number of realizations for each probe.

Probe	$\Theta U_\infty / H$	Number of realizations
1, 2, 3	100	40
4, 5	500	9
6	200	20
7, 8, 9, 10	10	100

mechanism known as vortex stretching is absent in two-dimensional flows. The probes 4 and 5 (mainly probe 4) on the other hand are located in a transition shear layer region where the fluid particles are subject to stronger velocity gradients, and the complex interactions of smaller scales with intermediate scales cannot be neglected. The coherence of the turbulent eddies in the shear layer is much increased if no three-dimensional instabilities are present (as it is the present case). The intensity of the coherent vortices grows via an inverse energy cascade, and eventually they start producing significant feedback on turbulence. Despite many achievements in numerical simulations of two-dimensional turbulence, the underlying physical mechanism of turbulent vortex interactions still remains unclear.

The mean-square difference between the probability average and the time average gives a direct measurement of the error ϵ, defined in (3.12). This error can also be estimated by the simple relation proposed in (3.13). Table 5.2 shows the values of the predicted and the ergodic deviation errors for each probe in the turbulent flow. The purpose of relation (3.13) is to give an estimation of the order of magnitude of the error introduced when a probability average is replaced by a temporal average. The results presented in Table 5.2 show a good agreement between the scaling based on an exponential decay of the autocorrelation function and the ergodic deviation error calculated numerically by using (3.12) with the autocorrelation sample computed from the numerical simulations. It is seen that for those points far outside from the recirculating bubble the turbulence behaves close to a random walk. On the other hand, the results also indicates that for those points inside and around the recirculating bubble, the ergodic deviation can be very high such as that found for the probe 4. So, the time average does not produce a meaningful statistics.

For the case of the probes 4 and 5, an exponential decay does not fit the real decay behavior of the normalized correlation function. It indicates that the turbulence in this region may have a quite different behavior of a typical random walk process. The dispersion process of momentum transport by velocity fluctuations seems to characterize an anomalous diffusion in the way described in Section 3. In that case, the integral in (3.8) must be evaluated numerically. The probe 4 shows a strong nonergodic property, suggesting that in this region a time average approach fails in describing the local turbulence. It is possible to infer that the probes in the neighboring or inserted into the recirculation bubble shown in Figure 4.1 have presented a significant deviation from the ergodicity. Consequently, the flow in this regions persists strongly correlated for a long time as shown in Table 5.2. It means that large turbulent structures dominate the flow in the recirculation bubble. In contrast in probes 7 to 10, the turbulence is characterized by



(a)

(b)

Figure 5.2. Dimensionless mean streamwise velocity as a function of the dimensionless time. (a) Numerical results of LES for probe 4, and (b) probe 8. Both plots show the error bars.

structures of smaller scales with short memory intervals and behavior closer to random fluctuations and a normal diffusion. In this case, a time average could be used to describe precisely the flow. From the plots in Figures 5.1(a) and 5.1(b), it is possible to evaluate the correlation degree of the flow process. In probes 4 and 5, the correlation functions decay very slowly with respect to the other ones. Their shapes are also different and an exponential or parabolic fit seems to be not appropriated anymore. For all other probes, an exponential fit was used to determine the finite time-scale integral as a measure of

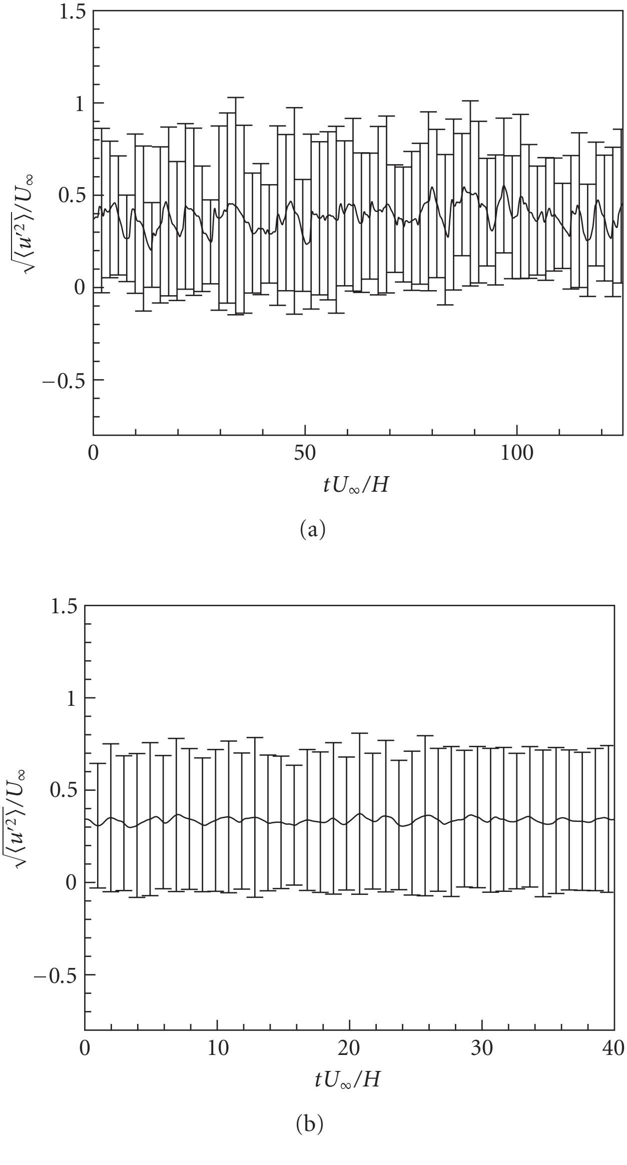


Figure 5.3. Dimensionless root mean square (RMS) as a function of the dimensionless time obtained from the LES (a) at probe 4 and (b) probe 8. The error bars are also shown in both plots.

the turbulence memory, corresponding to the time over which the velocity fluctuation is correlated with itself (i.e., the velocity fluctuation correlation time).

In addition, the two-dimensional power spectra relative to the probes 4 and 8 are presented in Figures 5.4(a) and 5.4(b), respectively. The details of the spectra calculations were mentioned in Section 3.1. The error bars are also considered in the plots. In these

Table 5.2. Comparison between the ergodic deviation error ϵ predicted by relation (3.13) and its value directly evaluating from the large eddy numerical simulation data by calculating the mean-square difference between the time and probability averages.

Probe	x/H	Computed error (%)	Predicted error (%)
1	6.5	16	20
2	8	12	20
3	9.5	14	20
4	11	4975	3400
5	12.5	24	13
6	14	10	10
7	15.5	3	3
8	17	3	3
9	18.5	4	3
10	20	3	3

plots, the abscissa is the logarithm of the nondimensional frequency, whereas the ordinate is defined so that the area beneath a logarithmic plot of $E(\omega)$ is proportional to the mean square of the fluctuating signal. Both spectra show that turbulence energy at small scales is decreased, while it is increased at large scales. In particular, one can see that the spectra of present two-dimensional simulations seem to change shape for moderate nondimensional frequencies ranging form 0.1 to 1. In the case of probe 4 located in the interface of the recirculating bubble as shown in Figure 4.1, we see that the energy cascade is characterized by a -3 spectral exponent which is different from the famous $\omega^{-5/3}$ valid for three-dimensional small scales of a local turbulence. The decay turbulence with $\omega^{-5/3}$ is observed only at higher frequency (smaller scales) as a result of the subgrid model used in numerical simulations which has been based on the Kolmogorov inertial equilibrium subrange of turbulence described in Section 2.3. The interval of the $\omega^{-5/3}$ spectrum is shown in both inserts of the mentioned figures. The presence of a spectrum interval with an ω^{-3} decay may be attributed to the mechanism of the inverse cascade which arises in two-dimensional turbulence inside the recirculating bubble. The resulting spectrum of probe 8 located outside the recirculating bubble (Figure 5.4(b)) exhibits virtually the same characteristics shown in the plot in Figure 5.4(a). However, the decay turbulence given by this spectrum is closer to ω^{-4} than to ω^{-3}. So, outside the recirculating region it is found that the turbulence decay corresponding to intermediate scales with dimensionless frequencies ranging form 1 to 10 gets steeper than -3. Most of these results seem to be in qualitatively agreement with the predictions given by the dimensional analysis of two-dimensional turbulent flow presented by Nazarenko and Laval [5].

The probability density functions associated with the turbulent velocity fluctuations in probes 4 and 8 are shown in Figures 5.5(a) and 5.5(b), respectively. The non-Gaussian deviation of the distribution function is clearly noticeable in probe 4, whereas in probe 8, the behavior of the probability density function is closer to a normal distribution. In particular, the behavior of the statistical distribution is quantified by the skewness and flatness

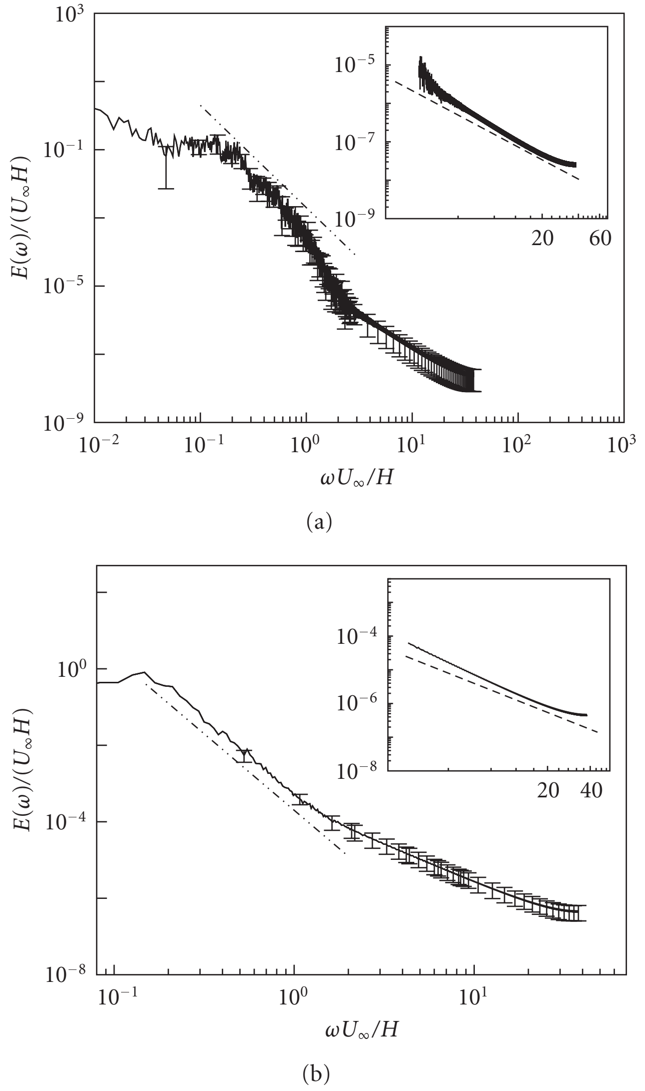
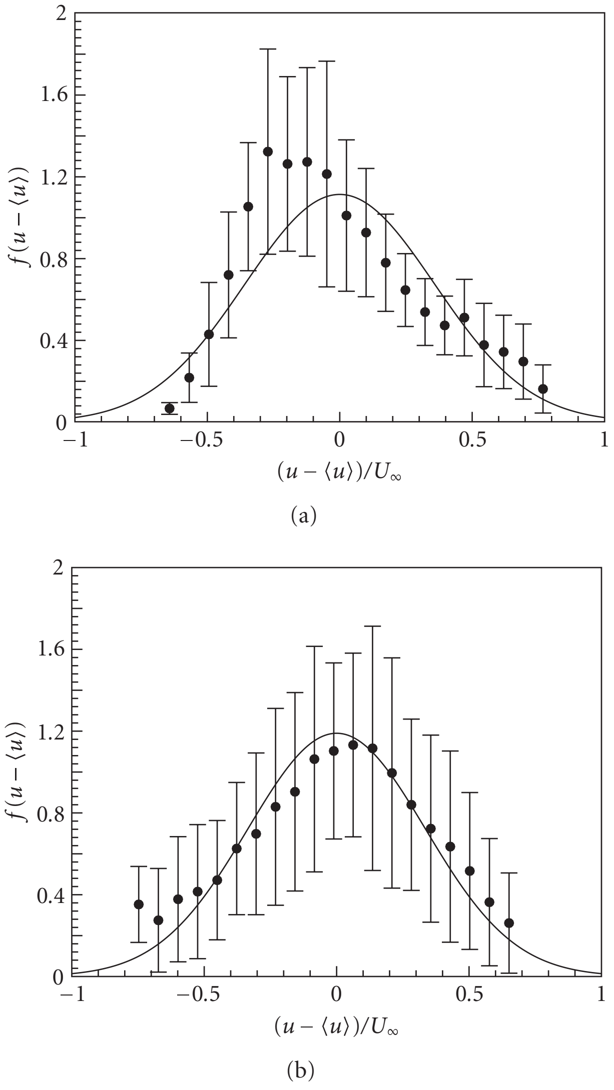


Figure 5.4. Dimensionless power spectra for the streamwise direction as a function of the dimensionless frequency for (a) probe 4 and (b) probe 8. The results were obtained from LES. Dashed line in (a) represents a decayment of the spectra with $(\omega U_\infty/H)^{-3}$, whereas the dashed line in (b) represents a spectrum with $(\omega U_\infty/H)^{-4}$. The inserts in the plots (a) and (b) show the Kolmogorov frequency decay $(\omega U_\infty/H)^{-5/3}$ for the smaller scale. The error bars are shown in both plots of the figure.

factors, defined as $\varphi = \langle u'(t)^3\rangle/\xi^3$ and $\kappa = \langle u'(t)^4\rangle/\xi^4$, where $\xi^2 = \langle u'(t)^2\rangle$, respectively. These factors and the turbulence intensities for each probe are listed in Table 5.3. It is well known that a normal distribution has $\varphi = 0$ and $\kappa = 3$. It is possible to infer that all processes display some non-Gaussian behavior. The turbulent intensity at probe 4 has the order of $2 \times 10^4\%$, whereas between others, the greater value of this parameter has



(a)

(b)

Figure 5.5. Probability density function. (a) Probe 4 and (b) probe 8. The solid line shows the standard Gaussian process. The error bars are shown in both plots.

the order of $10^2\%$ (see, e.g., probe 1, Table 5.3). This characteristic is an indication of the strong nonergodic property of the velocity fluctuations in probe 4. The interval of confidence represented by the error bars shown in the plots and by the associated errors to the quantities in Table 5.3 is relatively large. It suggests that for a complete characterization of these parameters, more realizations are required, and consequently it would demand additional computational effort in order to simulate much larger time intervals.

Table 5.3. The values of the turbulence intensities, skewness and flatness factors for each probe.

Probe	x/H	Turbulence intensity (%)	φ	κ
1	6.5	113 ± 38	-0.3 ± 0.6	2.7 ± 0.8
2	8	77 ± 26	-0.2 ± 0.5	2.7 ± 0.5
3	9.5	82 ± 16	0.2 ± 0.3	2.7 ± 0.5
4	11	$(2.1 \pm 0.4) \times 10^4$	0.5 ± 0.3	2.6 ± 0.6
5	12.5	88 ± 16	0.0 ± 0.3	2.3 ± 0.3
6	14	43 ± 6	-0.3 ± 0.3	2.5 ± 0.2
7	15.5	37 ± 4	-0.5 ± 0.4	2.9 ± 0.7
8	17	37 ± 4	-0.2 ± 0.5	2.9 ± 0.5
9	18.5	39 ± 3	0 ± 0.3	2.8 ± 0.4
10	20	41 ± 3	0 ± 0.2	2.7 ± 0.4

6. Concluding remark

In this paper, a rigorous statistical approach for the treatment of turbulent velocity fluctuations has been presented. We have looked at how long of a time average is necessary to obtain well-converged statistical results. For this end, we evaluate the mean-square difference between the time average and an ensemble average as the measure of convergence. From the numerical simulations, ten different points in the flow domain have been statistically treated using a probabilistic approach. The realizations of the statistical ensemble were defined by the cut up of a long-time velocity record into pieces of length much longer than the characteristic correlation time of the velocity fluctuations. Based on the velocity fluctuations correlation time, a statistical analysis of long time has been performed. The ergodicity assumption of the turbulence was investigated. The deviation ε of this criterium was evaluated and compared with theoretical predictions given by scaling arguments. The results have suggested that the deviation due to ergodicity assumption may be used as a criterion in order to predict the upper boundary simulation time required to have a convergence of the statistical characterization of the flow. We have found from the two-dimensional large eddy simulations that inside and outside a recirculating region, the turbulence decaying corresponding to intermediate scales (with dimensionless frequencies ranging form 1 to 10) gets steeper than the classical scaling $\omega^{-5/3}$ spectrum. We found ω^{-3} and ω^{-4} for a point inside and another outside the recirculating bubble, respectively. The decaying turbulence like $\omega^{-5/3}$ was seen only at the smaller scales attributed to the subgrid model used which was based on the Kolmogorov inertial equilibrium subrange of the turbulence. Probability functions, skewness and flatness coefficients have shown a deviation from the Gaussian behavior at all investigated positions. We have seen that the break of flow ergodicity property of the flow is directly related with the large-scale structures of the turbulence. In this flow regime, the integral time scale required to define the most commonly used time average is almost always unpredictable.

As future works, it would be important to think about developing more robust three-dimensional LES to test the ergodicity of turbulent flows in the presence of vortex-stretching and multistructure interactions. Two-dimensional LES calculations are clearly inferior to three-dimensional ones since certain important features of three-dimensional turbulence (vortex stretching) are not resolved. Moreover, two-dimensional large-scale structures are subject to three-dimensional instability which results in counter-rotation vortices. Behind this, there is a question of how the nonlocal interactions of large-scale vortices with intermediate and small-scale structures in three-dimensional turbulence affect the local ergodicity of the flow. Clearly, this fundamental topic requires further attention.

Acknowledgments

The authors gratefully acknowledge the support of CAPES, Brazil, CNPq, Brazil, and FINATEC, Universidade de Brasília, during the development of this work. The authors thank the referees for very insightful and stimulating comments. F. R. Cunha is the author of Correspondance.

References

[1] L. Persson, C. Fureby, and N. Svanstedt, “On homogenization-based methods for large-eddy simulation,” *Journal of Fluids Engineering*, vol. 124, no. 4, pp. 892–903, 2002.

[2] M. R. Visbal and D. P. Rizzetta, “Large-eddy simulation on curvilinear grids using compact differencing and filtering schemes,” *Journal of Fluids Engineering*, vol. 124, no. 4, pp. 836–847, 2002.

[3] P. Sagaut, *Large Eddy Simulation for Incompressible Flows: An Introduction*, Springer, New York, NY, USA, 1st edition, 1988.

[4] A. W. Vreman, “Direct and large eddy simulation of the compressible turbulent mixing layer,” Ph.D. dissertation, University of Twente, Enschede, The Netherlands, 1995.

[5] S. Nazarenko and J.-P. Laval, “Non-local two-dimensional turbulence and Batchelor’s regime for passive scalars,” *Journal of Fluid Mechanics*, vol. 408, pp. 301–321, 2000.

[6] D. Bouris and G. Bergeles, “2D LES of vortex shedding from a square cylinder,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 80, no. 1-2, pp. 31–46, 1999.

[7] D. A. McQuarrie, *Statistical Mechanics*, Harper & Row Publishers, New York, NY, USA, 1976.

[8] C. E. Ifeachor and B. W. Jervis, *Digital Signal Processing*, Addison-Wesley, New York, NY, USA, 2nd edition, 1993.

[9] T. F. Oliveira, F. R. Cunha, and R. F. M. Bobenrieth, “A stochastic analysis of a nonlinear flow response,” *Probabilistic Engineering Mechanics*, vol. 21, no. 4, pp. 377–383, 2006.

[10] Y. D. Sobral and F. R. Cunha, “A linear stability analysis of a homogeneous fluidized bed,” *Tendências em Matemática Aplicada e Computacional*, vol. 3, no. 2, pp. 197–206, 2002.

[11] A. Favre, “Turbulence: space-time statistical properties in supersonic flows,” *Physics of Fluids*, vol. 26, no. 10, pp. 2851–2863, 1983.

[12] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, Oxford, UK, 1987.

[13] D. Knight, G. Zhou, N. Okong’o, and V. Shukla, “Compressible Large Eddy Simulation Using Unstructured Grids,” in *AIAA 36th Aerospace Sciences Meeting & Exhibit, AIAA Paper 98-0535*, Reno, Nev, USA, January 1998.

[14] J. Smagorinsky, “General circulation experiment with the primitive equations, I. The basic experiment,” *Monthly Weather Review*, vol. 91, no. 3, pp. 99–164, 1963.

[15] D. K. Lilly, "Helicity," in *Lectures Notes on Turbulence*, J. Herring and J. Williams, Eds., pp. 171–218, World Scientific, River Edge, NJ, USA, 1987.

[16] J. W. Deardorff, "A numerical study of three-dimensional turbulent channel flow at large Reynolds numbers," *Journal of Fluid Mechanics*, vol. 41, no. 2, pp. 453–480, 1970.

[17] M. Lesieur, *Turbulence in Fluids*, vol. 1 of *Fluid Mechanics and Its Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2nd edition, 1990.

[18] L. Fulachier and R. Dumas, "Spectral analogy between temperature and velocity fluctuations in a turbulent boundary layer," *Journal of Fluid Mechanics*, vol. 77, no. 2, pp. 257–277, 1976.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 1965.

[20] G. K. Batchelor, *The Theory of Homogeneous Turbulence*, Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press, Cambridge, UK, 1953.

[21] G. Comte-Bellot and S. Corrsin, "Simple Eulerian time correlation of full- and narrow-band velocity signals in grid-generated, 'isotropic' turbulence," *Journal of Fluid Mechanics*, vol. 47, pp. 273–337, 1971.

[22] F. R. Cunha, G. C. Abade, A. J. Sousa, and E. J. Hinch, "Modeling and direct simulation of velocity fluctuations and particle-velocity correlations in sedimentation," *Journal of Fluids Engineering*, vol. 124, no. 4, pp. 957–968, 2002.

[23] H. Tennekes and J. L. Lumley, *A First Course in Turbulence*, The MIT Press, Cambridge, Mass, USA, 1990.

[24] G. K. Batchelor, "Small-scale variation of convected quantities like temperature in turbulent fluid—part 1: general discussion and the case of small conductivity," *Journal of Fluid Mechanics*, vol. 5, no. 1, pp. 113–133, 1959.

[25] P. H. Roberts, "Analytical theory of turbulent diffusion," *Journal of Fluid Mechanics*, vol. 11, no. 2, pp. 257–283, 1961.

[26] D. L. Koch and J. F. Brady, "Anomalous diffusion due to long-range velocity fluctuations in the absence of a mean flow," *Physics of Fluids A: Fluid Dynamics*, vol. 1, no. 1, pp. 47–51, 1989.

[27] C. Hirsch, *Numerical Computation of Internal and External Flows*, vol. 2, John Wiley & Sons, New York, NY, USA, 1st edition, 1990.

[28] D. A. Anderson, J. C. Tannehill, and R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer*, Series in Computational Methods in Mechanics and Thermal Sciences, Hemisphere, Washington, DC, USA, 1984.

[29] J. K. Eaton and J. P. Johnston, "Turbulent flow reattachment: an experimental study of the flow and structure behind a backward facinhg-step," Tech. Rep. Report MD-39, Stanford University, Stanford, Calif, USA, 1980.

T. F. Oliveira: Fluid Mechanics of Complex Flows Lab, Department of Mechanical Engineering,
University of Brasília, Campus Universitário Darcy Ribeiro, 70910-900 Brasília, DF, Brazil
Email address: taygoara@unb.br

R. B. Miserda: Fluid Mechanics of Complex Flows Lab, Department of Mechanical Engineering,
University of Brasília, Campus Universitário Darcy Ribeiro, 70910-900 Brasília, DF, Brazil
Email address: rfbm@unb.br

F. R. Cunha: Fluid Mechanics of Complex Flows Lab, Department of Mechanical Engineering,
University of Brasília, Campus Universitário Darcy Ribeiro, 70910-900 Brasília, DF, Brazil
Email address: frcunha@unb.br

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 65636, 14 pages
doi:10.1155/2007/65636

Research Article

A Stochastic Model for the HIV/AIDS Dynamic Evolution

Giuseppe Di Biase, Guglielmo D'Amico, Arturo Di Girolamo, Jacques Janssen, Stefano Iacobelli, Nicola Tinari, and Raimondo Manca

Received 28 September 2006; Revised 28 February 2007; Accepted 3 June 2007

Recommended by José Manoel Balthazar

This paper analyses the HIV/AIDS dynamic evolution as defined by CD4 levels, from a macroscopic point of view, by means of homogeneous semi-Markov stochastic processes. A large number of results have been obtained including the following conditional probabilities: an infected patient will be in state j after a time t given that she/he entered at time 0 (starting time) in state i; that she/he will survive up to a time t, given the starting state; that she/he will continue to remain in the starting state up to time t; that she/he reach stage j of the disease in the next transition, if the previous state was i and no state change occurred up to time t. The immunological states considered are based on CD4 counts and our data refer to patients selected from a series of 766 HIV-positive intravenous drug users.

1. Introduction

In this paper the homogeneous semi-Markov reliability stochastic model is proposed as a useful tool for predicting the evolution of the human immunodeficiency virus (HIV) infection and the probability of an infected patient's survival. This model, when compared to the most common epidemiologic data analyses, has the following advantages:

(i) not only is the randomness in the different states in which the infection can evolve into considered, but also the randomness of the time elapsed in each state;

(ii) all the states are interrelated, therefore any improvements are also considered;

(iii) a large number of disease states can be considered;

(iv) fewer and less rigid working hypotheses are needed;

(v) only raw data obtained from observations are needed, with no strong assumptions about any standard probability functions regarding the random variables analysed;

(vi) the conclusions are simply based on a list of all computed probabilities derived directly from raw data.

Semi-Markov processes were defined in the fifties independently of each other by Levy [1] and Smith [2]. A detailed theoretical analysis of semi-Markov processes was produced in Howard [3, 4]. Since then, they have been applied in a number of scientific fields including: engineering applications (systems reliability) [3–6], finance [7], insurance, actuarial and demographic sciences [6, 8, 9]. Semi-Markov models have also been employed in the field of biomedicine, for example, in applications to prevent, screen, and design cancer prevention trials, in Davidov [10], and Davidov and Zelen [11], respectively.

Moreover, many papers relating to HIV infection, have been written such as Lagakos et al. [12], Satten and Sternberg [13], Sternberg and Satten [14] and Sweeting et al. [15]. Foucher et al. [16] also considered various patients based on their ages by means of a parametric approach. Joly and Commenges [17] reduced the instability of nonparametric models but introduced some strong assumptions in order to estimate a posteriori intensity functions by penalizing the log-likelihood. Apart from [16], in all the papers quoted, the model solvability is related to the possibility that a patient might move through the states following the same direction. Our data has shown that there are no negligible probabilities of recovering from the disease, and so, in our dynamic analysis the unidirectionality hypothesis for the transitions among the states was not considered.

As regards the statistical analysis of semi-Markov processes, the fundamental references are Gill [18], Andersen et al. [19], Ouhbi and Limnios [20] and, more the recent, Limnios and Ouhbi [21] and Dabrowska and Ho [22].

Physicians consider that the HIV fully satisfies few and weak working hypotheses needed. Data refer to subjects selected from a series of 766 HIV-positive intravenous drug users screened at different Italian clinics in the period from October 1988 to December 1996. The cohort characteristics were described in [23]. The computation is done by means of *Mathematica* software designed and written by some of the authors.

2. Homogeneous semi-Markov process

In this part, the homogeneous semi-Markov process (HSMP) will be defined and the notation will be as given in [24].

In the SMP environment, two random variables run simultaneously:

$$X_n : \Omega \longrightarrow S, \quad T_n : \Omega \longrightarrow \mathbb{R}, \quad n \in \mathbb{N}, \tag{2.1}$$

X_n with state space $S = \{S_1, \dots, S_m\}$ represents the state at the nth transition. In the health care environment, the elements of S represent all the possible stages in which the disease may show level of seriousness. T_n, with state space equal to $\mathbb{R}$, represents the time of the nth transition. In this way, we cannot only consider the randomness of the states but also the randomness of the time elapsed in each state. The process (X_n, T_n) is assumed to be a homogeneous Markovian renewal process, see [25].

The kernel $\mathbf{Q} = [Q_{ij}(t)]$ associated with the process is defined as follows:

$$\begin{aligned} Q_{ij}(t) &= P[X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_0,\dots,X_{n-1};\ X_n = i;\ T_0,\dots,T_n] \\ &= P[X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i]. \end{aligned} \tag{2.2}$$

Thus, (Pyke [26])

$$p_{ij} = \lim_{t\to\infty} Q_{ij}(t); \quad i,j \in S,\ t \in \mathbb{R}, \tag{2.3}$$

where $\mathbf{P} = [p_{ij}]$ is the transition matrix of the embedded Markov chain in the process. Furthermore, it is necessary to introduce the probability that the process will leave state i in a time t as

$$H_i(t) = P[T_{n+1} - T_n \leq t \mid X_n = i]. \tag{2.4}$$

Obviously,

$$H_i(t) = \sum_{j=1}^{m} Q_{ij}(t). \tag{2.5}$$

It is now possible to define the distribution function of the waiting time in each state i, given that the state successively occupied is known,

$$G_{ij}(t) = P[T_{n+1} - T_n \leq t \mid X_n = i,\ X_{n+1} = j]. \tag{2.6}$$

Obviously, the related probabilities can be obtained by means of the following formula:

$$G_{ij}(t) = \begin{cases} \dfrac{Q_{ij}(t)}{p_{ij}} & \text{if } p_{ij} \neq 0 \\ 1 & \text{if } p_{ij} = 0. \end{cases} \tag{2.7}$$

The main difference between a continuous time Markov process and a semi-Markov process lies in the distribution functions $G_{ij}(t)$. In a Markov environment this function must be a negative exponential function. On the other hand, in the semi-Markov case, the distribution functions $G_{ij}(t)$ can be of any type. This means that the transition intensity can be decreasing or increasing.

If we apply the semi-Markov model in the health care environment, we can consider, by means of the $G_{ij}(t)$, the problem given by the duration of the time spent inside one of the possible disease states.

Now the HSMP $Z = (Z(t),\ t \in \mathbb{R})$ can be defined. It represents, for each waiting time, the state occupied by the process

$$Z(t) = X_{N(t)}, \quad \text{where } N(t) = \max\{n : T_n \leq t\}. \tag{2.8}$$

The transition probabilities are defined in the following way:

$$\phi_{ij}(t) = P[Z(t) = j \mid Z(0) = i]. \tag{2.9}$$

They are obtained by solving the following evolution equations:

$$\phi_{ij}(t) = \delta_{ij}(1 - H_i(t)) + \sum_{\beta=1}^{m} \int_0^t \dot{Q}_{ij}(\vartheta)\phi_{ij}(t-\vartheta)d\vartheta, \tag{2.10}$$

where δ_{ij} represents the Kronecker symbol.

The first addendum of formula (2.10) gives the probability that the system does not undergo transitions up to time t given that it was in state i at an initial time 0. In predicting the HIV/AIDS evolution model, it represents the probability that the infected patient does not shift to any new stage in a time t. In the second addendum, $\dot{Q}_{ij}(\vartheta)$ is the derivative at a time ϑ of $Q_{i\beta}(\vartheta)$ and it represents the probability that the system remained in a state i up to the time ϑ and that it shifted to state β exactly at a time ϑ. After the transition, the system will shift to state j following one of all the possible trajectories from state β to state j within a time $t-\vartheta$. In our application, it means that up to a time ϑ an infected subject remains in the state i. At the time ϑ, the patient moves into a new stage β and then reaches state j following one of the possible trajectories in some time $t-\vartheta$.

2.1. A description of HSMP numerical solution. In a previous paper, Corradi et al. [27] proved that it is easy to find the numerical solution of (2.10) by means of quadrature method. Moreover, they proved that the numerical solution of the process converges to the discrete time HSMP (DTHSMP).

Furthermore, in the same paper, it was proved that the DTHSMP process tends to be continuous if the discretization interval tends to 0. The discretization of formula (2.10) leads to the following infinite countable linear system:

$$\phi_{ij}^h(kh) = d_{ij}^h(kh) + \sum_{l=1}^{m}\sum_{\tau=1}^{k} v_{il}^h(\tau h)\phi_{lj}^h((k-\tau)h), \tag{2.11}$$

where h represents the discretization step

$$d_{ij}^h(kh) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 - H_i^h(kh) & \text{if } i = j, \end{cases}$$
$$v_{ij}^h(kh) = \begin{cases} 0 & \text{if } k = 0, \\ Q_{ij}^h(kh) - Q_{ij}^h((k-1)h) & \text{if } k > 0. \end{cases} \tag{2.12}$$

For more information on discretization see [28]. Relation (2.11) can be written in the following matrix form:

$$\mathbf{\Phi}^h(kh) - \sum_{\tau=1}^{k} \mathbf{V}^h(\tau h)\mathbf{\Phi}^h((k-\tau)h) = \mathbf{\Phi}^h(kh). \tag{2.13}$$

If $h = 1$, we have:

$$\phi_{ij}(k) = d_{ij}(k) + \sum_{l=1}^{m} \sum_{\tau=1}^{k} v_{il}(\tau)\phi_{lj}(k-\tau). \tag{2.14}$$

The following theorems have been proved in [27].

Theorem 2.1. *Equation (2.14) has a unique solution.*

Theorem 2.2. *The matrix $\mathbf{\Phi}^h(kh)$ is stochastic.*

Equation (2.14) is the evolution equation of the DTHSMP.

From all these results it follows that the solution of an SMP can be obtained by means of the DTSMP. Furthermore, we are interested in solving the problem in a finite time span. The solution can be found by means of a simple recursive method.

As a first step, (2.13) for $t = 0$ gives

$$\mathbf{D}^h(0) = \mathbf{\Phi}^h(0) = \mathbf{I}. \tag{2.15}$$

Knowing $\mathbf{\Phi}^h(0)$, it is possible to compute $\mathbf{\Phi}^h(h)$. Knowing these two matrices, it is possible to compute $\mathbf{\Phi}^h(2h)$ and so on.

3. Homogeneous semi-Markov reliability model

There are several semi-Markov models in reliability theory, see for example, Osaki [29] and more recently Limnios and Oprisan [5].

Let us consider a reliability system S that may be at any given time t in one of the states of $I = \{1,\dots,m\}$. The stochastic process of the successive states of S is $Z = \{Z(t),\ t \geq 0\}$. The state set is partitioned into sets U and D in the following way:

$$I = U \cup D, \quad U, D \neq \varnothing \quad \text{such that } U \cap D = \varnothing. \tag{3.1}$$

The subset U contains all “good” states in which the system is working and the subset D contains all “bad” states in which the system is not working properly or has failed.

The typical indicators used in reliability theory are the following:

(i) *the reliability function R* giving the probability that the system was always working from time 0 to a time t:

$$R(t) = P[Z(u) \in U : \forall u \in (0,t]]; \tag{3.2}$$

(ii) *the point-wise availability function A* giving the probability that the system is working at a time t whatever happens in $(0,t]$:

$$A(t) = P[Z(t) \in U]; \tag{3.3}$$

(iii) *the maintainability function M* giving the probability that the system will leave the set D within the time t being in D at time 0:

$$M(t) = 1 - P[Z(u) \in D, \forall u \in (0,t]]. \tag{3.4}$$

It has been shown in [5] that these three probabilities can be computed in the following way if the process is a homogeneous semi-Markov process with kernel $\mathbf{Q}$.

(i) *The point-wise availability function* A_i given that $Z(0) = i$:

$$A_i(t) = \sum_{j \in U} \phi_{ij}(t). \tag{3.5}$$

(ii) *the reliability function* R_i given that $Z(0) = i$.

To compute these probabilities, all the states of the subset D must be changed into absorbing states. $R_i(t)$ is given by solving the evolution equation of HSMP with the embedded Markov chain with $p_{ij} = \delta_{ij}$ if $i \in D$. The resulting formula is

$$R_i(t) = \sum_{j \in U} \widetilde{\phi}_{ij}(t), \tag{3.6}$$

where $\widetilde{\phi}_{ij}$ is the solution of (2.10) with all the states in D that are absorbing.

(iii) *The maintainability function* M_i given that $Z(0) = i$.

In this case, all the states of the subset U must be changed into absorbing states. $M_i(t)$ is given by solving the evolution equation of HSMP with the embedded Markov chain with $p_{ij} = \delta_{ij}$ if $i \in U$. The resulting formula is

$$M_i(t) = \sum_{j \in U} \widehat{\phi}_{ij}(t), \tag{3.7}$$

where $\widehat{\phi}_{ij}(t)$ is the solution of (2.10) with all the states in U that are absorbing.

4. Application of the model to the HIV/AIDS dynamic evolution

The acquired immunodeficiency syndrome (AIDS) is caused by the human immunodeficiency virus (HIV), a virus belonging to the lentivirus subgroup of retroviruses [30, 31]. The hallmark of the HIV infection is the progressive depletion of a class of lymphocytes named CD4+ or helper lymphocytes which play a pivotal regulatory role in the immune response to infections and tumours. The immune suppression resulting from the CD4+ decline leads to high susceptibility to opportunistic infections and possibly unusual tumours. Without appropriate antiretroviral treatment, AIDS is almost uniformly lethal [30, 31].

The natural history of HIV infection is characterized by a phase of latency that can last for several years, and evolves through consecutive steps [32] defined on the basis of CD4+ lymphocyte count and constitutional symptoms [33] with full blown AIDS representing the final stage of the disease [34]. The time spent in each stage of the disease is not predictable on the basis of clinical and immunological parameters.

HIV is transmitted primarily by sexual contact, syringe sharing amongst intravenous drug users, blood and blood products not properly screened. From an epidemiological point of view, the disease has spread worldwide. It is currently estimated that the total number cases of HIV infections is some 39.5 million, with a peak in the sub-Saharan African continent, and East Asian countries [35].

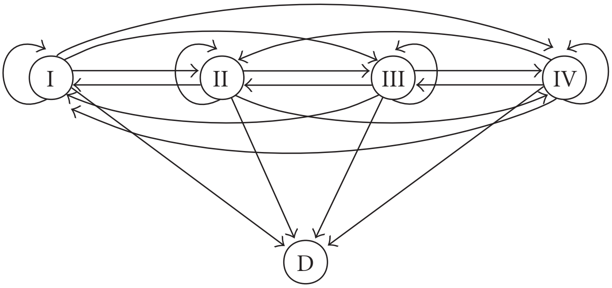


FIGURE 4.1. The model of the immunological stages a HIV/AIDS infected patient can go into.

Physicians believe that the fundamental hypothesis needed in order to apply the model in HIV/AIDS environment is satisfied. Indeed, as quoted in [36] the relation (2.6) is nearer to reality, that is, in the absence of treatment, the future of the patient depends only on the present state but not on all previous history.

Followup took $T = 87$ months (from October 1989 to December 1996). The retrospective study concerned a cohort of $K = 766$ HIV-positive intravenous drug users. Database fields were completed by means of a number of biological and clinical parameters obtained from 2488 medical examinations. In order to predict the HIV/AIDS evolution, we employed the following immunological states related to CD4+ count plus an absorbing state (the death of the patient): state I ($\text{CD4} > 500 \times 10^6$ cells/L), state II ($350 < \text{CD4} \leq 500$), state III ($200 < \text{CD4} \leq 350$), state IV ($\text{CD4} \leq 200$), and state D (absorbing state). We assume, therefore, that the HIV/AIDS infection shifts between five different degrees of seriousness. This choice was justified by the CDC immunological classification [33], and taking into account the recommendations of the DHHS (Department of Human and Health Services) for the initiation of antiretroviral therapy [37].

All that led to the following set of states:

$$S = \{\text{I}, \text{II}, \text{III}, \text{IV}, \text{D}\}. \tag{4.1}$$

Figure 4.1 represents the graph model. It shows all the immunological states an HIV/AIDS infected patient can undergo. All the states, apart from D, are interrelated, and also improvements are taken into account. It is also possible that an examination will show that a patient's state has not changed.

The first four states are working states (good states) and the last one is the only bad state. This is represented in the following two subsets:

$$U = \{\text{I}, \text{II}, \text{III}, \text{IV}\}, \qquad D = \{\text{D}\}. \tag{4.2}$$

In this case, the maintainability function M does not make sense because the default state D is absorbing and once an infected patient had entered this state it was no longer possible to leave it.

Furthermore, the fact that the only bad state is an absorbing state implies that the availability function A corresponds to the reliability function R.

Table 4.1. Transition frequencies matrix of the followed-up cohort and estimates of the transition matrix.

States	I	II	III	IV	D
I	381	135	42	19	6
II	115	252	129	51	8
III	26	108	319	144	31
IV	11	19	64	144	31

Another important result that can be obtained by means of the semi-Markov approach is the distribution function of the subject's death conditioned to the state held at time 0.

In the health care environment, the reliability model is substantially simplified. In fact, to obtain all the results that are relevant to our study it suffices to solve the system (2.11) numerically only once since $\widetilde{\phi} = \widehat{\phi}_{ij}(t) = \phi_{ij}(t)$.

In order to obtain the claimed results, we need to estimate the semi-Markov kernel $\mathbf{Q} = [Q_{ij}(t)]$ from our data set.

Firstly, we introduce the following symbols:

(i) K is the number of independent trajectories in our data set;

(ii) X_n^r is the state at nth transition of the rth subject;

(iii) T_n^r is the time in which the rth subject makes the nth transition;

(iv) $N^r = N^r(T) = \sup\{n \in \mathbb{N} : T_n^r \leq T\}$ is the total number of transitions held by the rth subject;

(v) $N_i^r = N_i^r(T) = \sum_{k=1}^{N^r} \mathbf{1}_{\{X_{k-1}^r = i\}}$ is the number of visits of the rth subject to the state i;

(vi) $N_i = N_i(T) = \sum_{r=1}^{K} N_i^r$ is the total number of visits of all subjects to the state i.

Then we consider the empirical kernel estimator defined in [21] by

$$\widehat{Q}_{ij}(t,K) = \frac{1}{N_i}\sum_{r=1}^{K}\sum_{l=1}^{N^r}\mathbf{1}_{\{X_{l-1}^r = i, X_l^r = j, T_l^r - T_{l-1}^r \leq t\}}. \tag{4.3}$$

In [21] it was proved that the empirical kernel estimator is uniformly strongly consistent and, properly centralized and normalized, it converges to the normal random variable.

Owing to lack of space, we do not show the kernel estimates, but we can make them available upon request. We report, in Table 4.1, the frequencies of the transitions between the states and, consequently, in Table 4.2, the estimates of the embedded Markov chain.

Obviously the obtained estimates $\widehat{Q}_{ij}(t,K)$ are used as input to estimate all the relevant variables listed in Section 5.

5. Numerical results

After solving the evolution equations of the semi-Markov model with kernel $\widehat{\mathbf{Q}}$, an extensive amount of information useful to a phisician can be obtained, including the following.

(1) $\widehat{\phi}_{ij}(t)$, that represents, for each t, for each $j \in \{\text{I, II, III, IV, D}\}$, and for each $i \in \{\text{I, II, III, IV}\}$ the probabilities of being in a state j after a time t given that she/he entered at time 0 (starting time) in the state i. In Figure 5.1, there is a graphical representation of

TABLE 4.2. Estimates of the transition matrix of the embedded Markov chain.

States	I	II	III	IV	D
I	0.654	0.232	0.072	0.033	0.010
II	0.207	0.454	0.232	0.092	0.014
III	0.041	0.172	0.508	0.229	0.049
IV	0.015	0.026	0.089	0.681	0.188
D	0	0	0	0	1

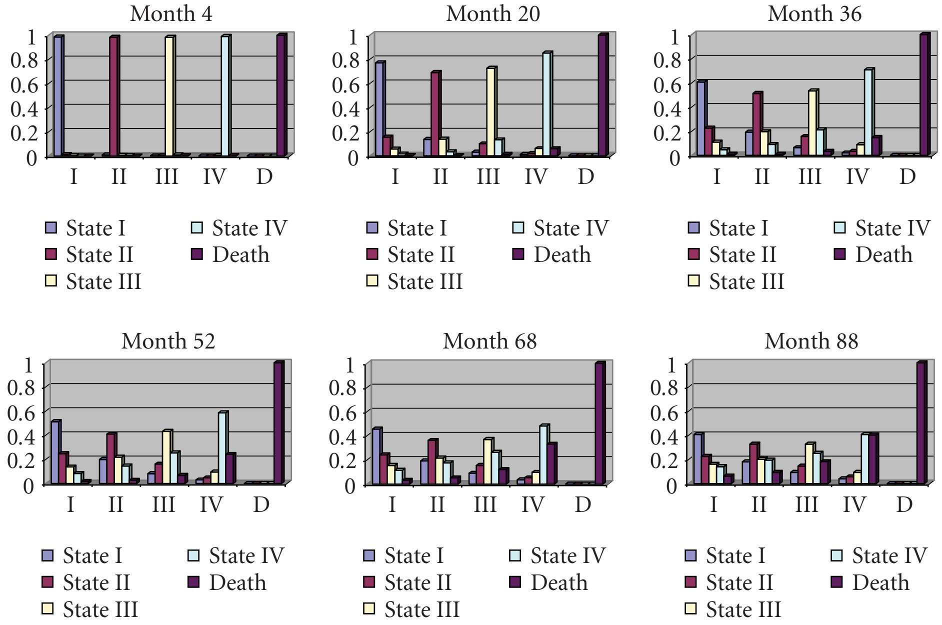


FIGURE 5.1. Conditional probabilities of being in state j after a month t given the starting state i. The starting states are in the axis categories.

such conditional probabilities. For the sake of brevity, only the values corresponding to lapses of sixteen months and up to month 88 are reported. They are all, however, available on request. It seems superfluous to underline the medical relevance of such computed probabilities. For example, if an HIV infected patient is in the third stage of the disease, with 21% risk, after 52 months he will be in the fourth stage (see Figure 5.1, Month 52).

(2) $\hat{R}_i(t) = \hat{A}_i(t) = \sum_{j \in U} \hat{\phi}_{ij}(t)$, that represents the conditional probabilities, given the starting state, that an infected patient will survive up to a time t. $\hat{R}_i(t)$ gives a physician vital information. In Figure 5.2, four curves, which depend on the starting state of the subject, have been computed. For example, if we look at the lowest curve we can read $\hat{R}_{i=\text{IV}}(42) = 0.8$ and we may conclude that, with a probability equal to 0.8, an infected patient that was in state IV will not die after 42 months.

(3) $1 - \hat{H}_i(t)$ represents the conditional probabilities of staying in the starting state until month t. In Figure 5.3 these conditional probabilities have been computed depending

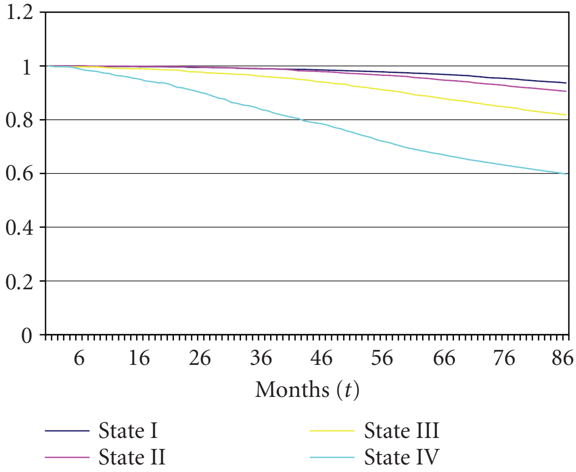


FIGURE 5.2. Survival conditional probabilities up to month t given the starting state.

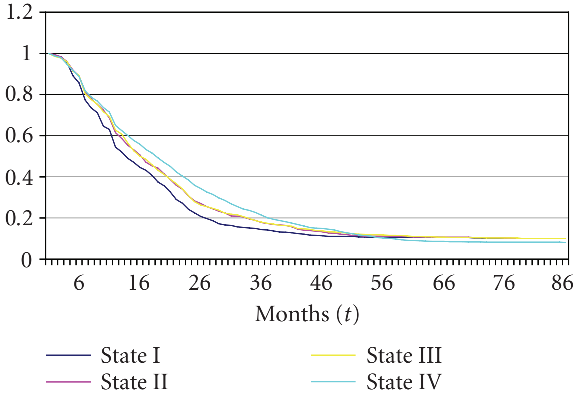


FIGURE 5.3. Stay on conditional probability in the starting state at least for a time t.

on the starting state. For example, if an HIV-infected patient comes under study at the fourth stage of the disease, with 40% risk, after 24 months he will still be in the fourth stage.

Before giving another result of current interest for epidemiologic purposes that can be obtained in an SMP environment, the concept of the first transition after time t must be introduced. More precisely, it is supposed that a subject at time 0 was in state i and it is known that with probability $(1 - H_i(t))$ he does not shift from state i. Under these hypotheses, it is possible to know the probability of the next transition is to state j. This probability will be denoted by $\varphi_{ij}(t)$. In terms of formulas it means the following:

$$\varphi_{ij}(t) = P[X_{n+1} = j \mid X_n = i,\ T_{n+1} - T_n > t]. \tag{5.1}$$

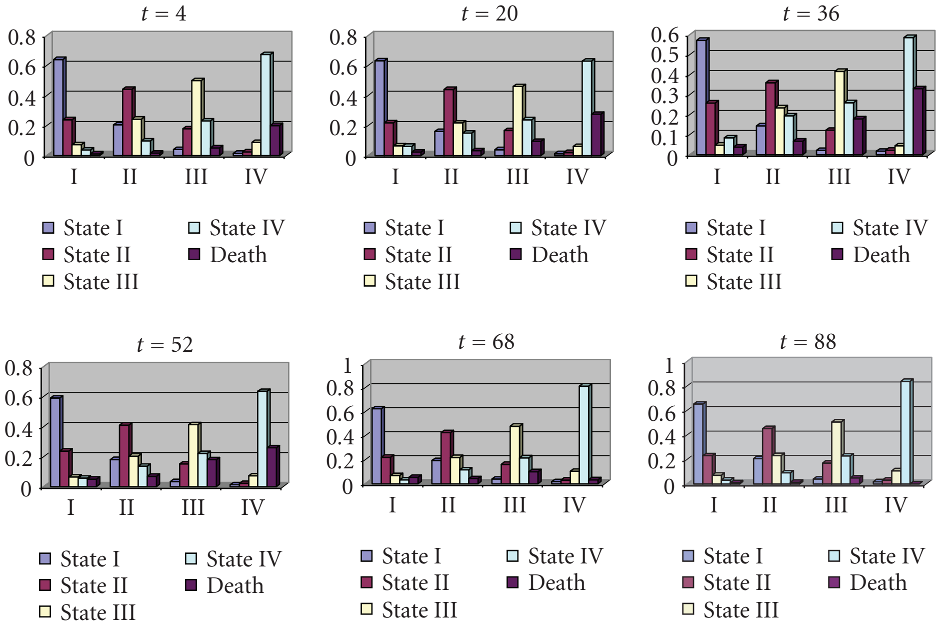


FIGURE 5.4. Conditional probabilities of developing state j of the disease at the next transition given that previous state occupied was i and no change occurred up to month t. The states occupied up to month t are in the axis categories.

This probability can be estimated by means of the following formula:

$$\widehat{\varphi}_{ij}(t) = \frac{\widehat{p}_{ij} - \widehat{Q}_{ij}(t)}{1 - \widehat{H}_i(t)}. \tag{5.2}$$

After definition (5.1) by means of SMP, it is possible to obtain the following result.

(4) $\widehat{\varphi}_{ij}(t)$ represents the estimated probability of developing stage j of the disease at the next transition if the previous state was i and no change of state occurred up to time t. In this way, in the case we studied, if the patient does not shift for a time t from state i, the probability of him being dead in the next transition can be computed easily. In Figure 5.4, a graphical representation of the probabilities of the first transition after a time t is shown. As for $\widehat{\phi}_{ij}(t)$, only the values corresponding to lapses of sixteen months are reported. They are all, however, available on request. A physician might read the probability of moving into state j of the disease (for each $j \in \{\text{I, II, III, IV, D}\}$) at the next transition if the previous state occupied was i (for each $i \in \{\text{I, II, III, IV}\}$) and no change occurred up to month t (for each t).

6. Concluding remarks

In this paper we have presented an HSMP approach to the dynamic evolution of the Human Immunodeficiency Virus Infection, as defined by CD4+ levels, and the probabilities of an infected patient's survival. By means of this approach, we cannot only consider randomness in the possible stages of seriousness which the disease may show but also the randomness of the duration of the waiting time in each state. The model starts from the idea that the disease evolution problem can be considered a special type of reliability problem and this idea allows the application of some semi-Markov reliability results to a healthcare environment.

We would like to point out that this paper does not show all the potential of the semi-Markov environment. Indeed, by means of the backward recurrence time process it is possible to assess different transition probabilities as a function of the duration inside the states. Moreover, it is possible to attach a reward structure to the process that allows the possibility of doing a cost analysis that considers, for example, the cost of antiretroviral treatment and/or other social costs related to the dynamic evolution of the HIV infection. These features will be the object of future research.

References

[1] P. Levy, "Processus semi-markoviens," in *Proceedings of the International Congress of Mathematicians, 1954, Amsterdam*, vol. 3, pp. 416–426, Erven P. Noordhoff N.V., Groningen, The Netherlands, 1956.

[2] W. L. Smith, "Regenerative stochastic processes," *Proceedings of the Royal Society of London. Series A.*, vol. 232, pp. 6–31, 1955.

[3] R. A. Howard, *Dynamic Probabilistic Systems. Vol. I: Markov Models*, John Wiley & Sons, New York, NY, USA, 1971.

[4] R. A. Howard, *Dynamic Probabilistic Systems. Vol. II: Semi-Markov and Decision Processes*, John Wiley & Sons, New York, NY, USA, 1971.

[5] N. Limnios and G. Oprişan, *Semi-Markov Processes and Reliability*, Statistics for Industry and Technology, Birkhäuser, Boston, Mass, USA, 2001.

[6] J. Janssen and R. Manca, *Applied Semi-Markov Processes*, Springer, New York, NY, USA, 2006.

[7] G. Di Biase, J. Janssen, and R. Manca, "Future pricing through homogeneous semi-Markov processes," *Applied Stochastic Models in Business and Industry*, vol. 21, no. 3, pp. 241–249, 2005.

[8] J. Janssen and R. Manca, "A realistic non-homogeneous stochastic pension fund model on scenario basis," *Scandinavian Actuarial Journal*, no. 2, pp. 113–137, 1997.

[9] G. D'Amico, J. Janssen, and R. Manca, "Homogeneous semi-Markov reliability models for credit risk management," *Decisions in Economics and Finance*, vol. 28, no. 2, pp. 79–93, 2006.

[10] O. Davidov, "The steady state probabilities for a regenerative semi-Markov processes with application to prevention and screening," *Applied Stochastic Models and Data Analysis*, vol. 15, no. 1, pp. 55–63, 1999.

[11] O. Davidov and M. Zelen, "Designing cancer prevention trials: a stochastic approach," *Statistics in Medicine*, vol. 19, no. 15, pp. 1983–1995, 2000.

[12] S. W. Lagakos, C. J. Sommer, and M. Zelen, "Semi-Markov models for partially censored data," *Biometrika*, vol. 65, no. 2, pp. 311–317, 1978.

[13] G. A. Satten and M. R. Sternberg, "Fitting semi-Markov models to interval-censored data with unknown initiation times," *Biometrics*, vol. 55, no. 2, pp. 507–513, 1999.

[14] M. R. Sternberg and G. A. Satten, "Discrete-time nonparametric estimation for semi-Markov models of chain-of-events data subject to interval-censoring and truncation," *Biometrics*, vol. 55, no. 2, pp. 514–522, 1999.

[15] M J. Sweeting, D. De Angelis, and O. O. Aalen, "Bayesian back-calculation using a multi-state model with application to HIV," *Statistics in Medicine*, vol. 24, no. 24, pp. 3991–4007, 2005.

[16] Y. Foucher, E. Mathieu, P. Saint-Pierre, J.-F. Durand, and J.-P. Daurès, "A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease," *Biometrical Journal*, vol. 47, no. 6, pp. 825–833, 2005.

[17] P. Joly and D. Commenges, "A penalized likelihood approach for a progressive three-state model with censored and truncated data: application to AIDS," *Biometrics*, vol. 55, no. 3, pp. 887–890, 1999.

[18] R. D. Gill, "Nonparametric estimation based on censored observations of a Markov renewal process," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 53, no. 1, pp. 97–116, 1980.

[19] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer Series in Statistics, Springer, New York, NY, USA, 1993.

[20] B. Ouhbi and N. Limnios, "Nonparametric estimation for semi-Markov processes based on its hazard rate functions," *Statistical Inference for Stochastic Processes*, vol. 2, no. 2, pp. 151–173, 1999.

[21] N. Limnios and B. Oubhi, "Nonparametric estimation for semi-Markov processes based on k-sample paths with application to reliability," in *Proceedings of the 11th Symposium on Applied Stochastic Models and Data Analysis (ASMDA '05)*, J. Janssen and P. Lenca, Eds., pp. 1061–1068, Brest, France, May 2005.

[22] D. M. Dabrowska and W. T. Ho, "Estimation in a semiparametric modulated renewal process," *Statistica Sinica*, vol. 16, no. 1, pp. 93–119, 2006.

[23] S. Iacobelli, A. Ullrich, N. Tinari, et al., "The 90K tumor-associated antigen and clinical progression in human immunodeficiency virus infection," *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, vol. 10, no. 4, pp. 450–456, 1995.

[24] E. Çinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1975.

[25] E. Çinlar, "Markov renewal theory: a survey," *Management Science*, vol. 21, no. 7, pp. 727–752, 1975.

[26] R. Pyke, "Markov renewal processes: definitions and preliminary properties," *Annals of Mathematical Statistics*, vol. 32, pp. 1231–1242, 1961.

[27] G. Corradi, J. Janssen, and R. Manca, "Numerical treatment of homogeneous semi-Markov processes in transient case—a straightforward approach," *Methodology and Computing in Applied Probability*, vol. 6, no. 2, pp. 233–246, 2004.

[28] A. Blasi, J. Janssen, and R. Manca, "Numerical treatment of homogeneous and non-homogeneous semi-Markov reliability models," *Communications in Statistics. Theory and Methods*, vol. 33, no. 3, pp. 697–714, 2004.

[29] S. Osaki, *Stochastic System Reliability Modeling*, vol. 5 of *Series in Modern Applied Mathematics*, World Scientific, Singapore, 1985.

[30] E. P. Gelmann, M. Popovic, D. Blayney, et al., "Proviral DNA of a retrovirus, human T-cell leukemia virus, in two patients with AIDS," *Science*, vol. 220, no. 4599, pp. 862–865, 1983.

[31] R. M. Anderson, G. F. Medley, R. M. May, and A. M. Johnson, "A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS," *IMA Journal of Mathematics Applied in Medicine and Biology*, vol. 3, no. 4, pp. 229–263, 1986.

[32] J. A. Levy, "Pathogenesis of human immunodeficiency virus infection," *Microbiological Reviews*, vol. 57, no. 1, pp. 183–289, 1993.

[33] Centres for Disease Control & Prevention, "1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults," *MMWR Recommendations and Reports*, vol. 41, no. RR-17, pp. 1–19, 1992.

[34] H. W. Jaffe and A. R. Lifson, "Acquisition and transmission of HIV," *Infectious Diseases Clinic of North America*, vol. 2, no. 2, pp. 299–306, 1988.

[35] UNAIDS/WHO, “AIDS Epidemic Update,” December 2006, http://www.unaids.org/en/HIV_data/epi2006/default.asp.

[36] S. Casari, A. Donasi, G. Paraninfo, et al., “Prognostic factors correlated with survival in AIDS patients,” *European Journal of Epidemiology*, vol. 15, no. 8, pp. 691–698, 1999.

[37] P. E. Sax, “Updated DHHS treatment guidelines,” *AIDS Clinical Care*, vol. 18, no. 12, p. 105, 2006.

Giuseppe Di Biase: Department of Science, University G. D’Annunzio of Chieti-Pescara,
viale Pindaro 42, 65127 Pescara, Italy
Email address: dibiase@sci.unich.it

Guglielmo D’Amico: Department of Drug Sciences, University G. D’Annunzio of Chieti-Pescara,
via dei Vestini, 66100 Chieti, Italy
Email address: g.damico@unich.it

Arturo Di Girolamo: Division of Infectious Diseases, Chieti Hospital,
via dei Vestini, 66100 Chieti, Italy
Email address: arturodigirolamo@aliceposta.it

Jacques Janssen: CESIAF, EURIA, Universite de Bretagne Occidentale, 6 avenue le Gorgeu,
CS 93837, 29238 Brest, Cedex 3, France
Email address: cesiaf@belgacom.net

Stefano Iacobelli: Department of Medical Oncology, University G. D’Annunzio of Chieti-Pescara,
via dei Vestini 66, 66100 Chieti, Italy
Email address: iacobelli@unich.it

Nicola Tinari: Department of Medical Oncology, University G. D’Annunzio of Chieti-Pescara,
via dei Vestini 66, 66100 Chieti, Italy
Email address: ntinari@unich.it

Raimondo Manca: Department of Mathematics for the Economics, Financial and Insurance
Decisions, University La Sapienza of Rome,
via del Castro Laurenziano 9, 00161 Rome, Italy
Email address: raimondo.manca@uniroma1.it

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 18609, 17 pages
doi:10.1155/2007/18609

Research Article

Models for Master-Slave Clock Distribution Networks with Third-Order Phase-Locked Loops

José Roberto Castilho Piqueira and Marcela de Carvalho Freschi

Received 9 September 2006; Accepted 16 February 2007

Recommended by José Manoel Balthazar

The purpose of this work is to study the processing and transmission of clock signals in networks of geographically distributed nodes, in order to derive conditions for frequency and phase synchronization between the nodes. The focus is on the master-slave architecture, which presents a priority scheme of clock distribution. One-way master-slave (OWMS) and two-way master-slave (TWMS) chains are studied, considering that the slave nodes are third-order phase-locked loops (PLLs). Third-order PLLs are chosen to improve the transient response but, if their parameters are not well adjusted, stability problems and chaotic behaviors appear, restricting the lock-in range of the network. Lock-in range for third-order PLLs with Sallen-Key filter is determined and it is verified whether this range is reduced when the PLLs are connected to a network. Numerical experiments show how chain size changes the lock-in ranges and the acquisition times.

1. Introduction

The distribution of clock signals is essential for several applications in control and communication [1] establishment of a worldwide time schedule system, synchronization of oscillators at different multiplexing points in digital telecommunication networks, control and monitoring of performance at specific instants in industrial processes and establishment of a synchronous state in a supercomputer composed of several processors.

In this work, the distribution of clock signals in telecommunications networks is considered. The problem concerns mainly to the distribution of phase and frequency signals through nodes distributed in a certain geographic area. There are three different implementation strategies: plesiochronous, master-slave (MS), and synchronous full-connected [2].

Plesiochronous networks are used when the synchronism is not critical. Oscillators with small frequency deviation are used in each node. They are manually adjusted and control signals are not needed. These networks are easy to implement, robust, though costly.

In synchronous full-connected networks, all the nodes have their own oscillator interchanging reference signals. They are more complex to be implemented, being used only in special applications, as in military communication networks.

Master-slave networks present a priority scheme of clock distribution, establishing a hierarchy between the nodes. There is a node with an extremely precise atomic oscillator, called master. The other nodes are controlled by the master's reference signal and are called slaves. As the control is centralized, if the master fails, the performance of the whole network is spoiled. However, due to its simple implementation and low cost, the master-slave networks are widely used in robotics and public telecommunications networks.

There are two types of MS networks, OWMS and TWMS. In the OWMS architecture, mainly used in telecommunication networks, the clock signal generated by the master is transmitted to the nodes sequentially, not considering the state of the slaves. In the TWMS, mainly used in process-control networks, the reference signal considers the master and the state of the slaves.

Here, the two distribution schemes are considered and compared from the lock-in point of view, with the slaves being third-order PLLs [3]. First, the isolated third-order PLL is considered and modeled, with an analytical determination of its lock-in range. Then, chains with third-order PLLs as slaves are explored by using numerical experiments.

2. Third-order PLL

The phase-locked loop (PLL) is an electronic device that has been used since 1932 in applications that demand automatic control of frequency. It is composed of a phase-detector (PD), a lowpass filter (F) and a voltage controlled oscillator (VCO) [4], as shown in Figure 2.1, and is used to extract the time basis in a reliable way, synchronizing the input signal $v_i(t)$ with the one of its internal oscillator (VCO) $v_o(t)$.

The nonlinear behavior of the PLL is due the phase detector (PD), which is represented by a signal multiplier that compares the phases of the input signal, $v_i(t)$, and the VCO output, $v_o(t)$. This operation is described by

$$v_d(t) = k_d v_i(t) v_o(t), \tag{2.1}$$

where k_d is the PD gain, and $v_i(t)$ and $v_o(t)$ have periodic expressions with central angular frequency $\omega_M(t)$ and instantaneous phases $\theta_i(t)$ and $\theta_o(t)$, respectively, as described

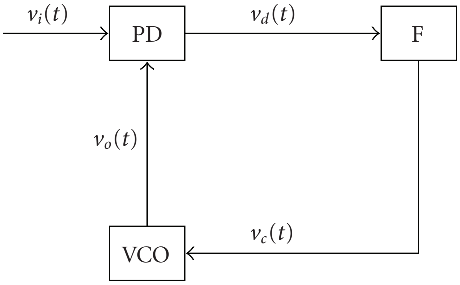


Figure 2.1. PLL block diagram.

below:

$$\begin{aligned} v_i(t) &= V_i \sin\left[\omega_M t + \theta_i(t)\right], \\ v_o(t) &= V_o \sin\left[\omega_M t + \theta_o(t)\right]. \end{aligned} \tag{2.2}$$

Combining (2.1), (2.2), $v_d(t)$ can be expressed as

$$v_d(t) = \frac{k_m V_i V_o}{2} \sin\left[2\omega_M t + \theta_i(t) + \theta_o(t)\right] + \frac{k_m V_i V_o}{2} \sin\left[\theta_i(t) - \theta_o(t)\right], \tag{2.3}$$

where k_m is the multiplier gain, being $k_d = k_m V_o/2$.

In expression (2.3), the presence of a second-harmonic term can be noticed. The signal $v_d(t)$ passes through the lowpass filter (F) to eliminate this high-frequency term called double-frequency jitter [5]; however, a small amplitude double-frequency term remains. This jitter is responsible for oscillations around the synchronous state, causing disturbances in the network performance [6].

The VCO signal is controlled by the filter output and its frequency is given by

$$\dot{\theta}_o(t) = k_o v_c(t), \tag{2.4}$$

where k_o is the VCO gain.

The VCO output has a phase $\theta_o(t)$. When the phase error $\varphi(t) = \theta_i(t) - \theta_o(t)$ has a constant value or, equivalently, the frequency error $\dot{\varphi}(t) = \dot{\theta}_i(t) - \dot{\theta}_o(t)$ is zero, the system is in the synchronous state.

First-order filters, implying second-order PLLs, are normally chosen due to their inherent stability and good lock-in range [7]. However, these PLLs frequently present high-level double-frequency terms in the PD output, as a phase-jitter, as it is difficult to adapt a first-order filter that eliminates these components in a satisfactory way. In order to eliminate this double-frequency jitter and to improve the transitory responses, higher-order filters are chosen, implying PLLs with order greater than 2.

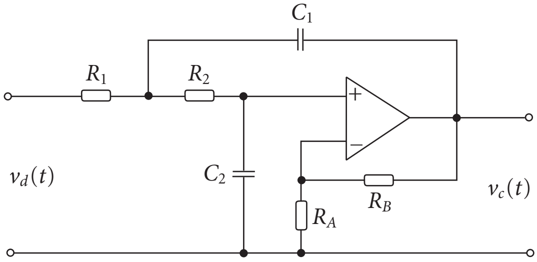


Figure 2.2. Sallen-Key second-order filter.

Here, a second-order Sallen-Key filter as shown in Figure 2.2, is chosen [7], resulting in a third-order PLL. The filter transfer function is given by

$$H(s) = \frac{V_c(s)}{V_d(s)} = \frac{K\omega_0^2}{s^2 + (\omega_0/Q)s + \omega_0^2}, \tag{2.5}$$

where $\omega_0^2 = 1/R_1R_2C_1C_2$, $K = 1 + R_B/R_A$ and $Q = 1/\omega_0[C_2(R_1 + R_2) + R_1C_1(1 - K)]$.

Considering the normalization of the cut-off frequency, that is, $R_1 = R_2 = C_1 = C_2 = 1$, and the PLL gain as $G = k_dk_0V_iV_o/2$, combining (2.3), (2.4) and (2.5), and neglecting the high-frequency terms of (2.4), the dynamic equation of the third-order PLL becomes

$$\dddot{\varphi} + (3 - K)\ddot{\varphi} + \dot{\varphi} + KG\sin\varphi = \dddot{\theta}_i + (3 - K)\ddot{\theta}_i + \dot{\theta}_i. \tag{2.6}$$

Considering phase-ramp inputs, $\theta_i = \Omega t + \psi$, the dynamic equation becomes

$$\dddot{\varphi} + (3 - K)\ddot{\varphi} + \dot{\varphi} + KG\sin\varphi = \Omega. \tag{2.7}$$

Equation (2.7) describes the third-order PLL, considering $\varphi \in (-\pi, \pi]$ and $\Omega > 0$. The synchronous state corresponds to a constant phase error φ, and to frequency and acceleration errors, $\dot{\varphi}$ and $\ddot{\varphi}$, equal to zero.

3. Lock-in range

The set of parameters and inputs corresponding to a reachable asymptotically stable synchronous state for (2.7) is called lock-in range. Consequently, the lock-in range is the set of filter gains K, input frequencies Ω, and PLL gains G corresponding to the existence of an asymptotically stable synchronous state $(\varphi, \dot{\varphi}, \ddot{\varphi}) = (\varphi^*, 0, 0)$ for (2.7).

Analyzing (2.7), the synchronous state is $(\varphi, \dot{\varphi}, \ddot{\varphi}) = (\arcsin(\Omega/KG), 0, 0)$, implying a first existence condition $\Omega \le KG$.

For $\Omega = KG$, there is the synchronous state $(\pi/2, 0, 0)$ that is nonhyperbolic [8]. For $\Omega < KG$, there are two synchronous states: $(\varphi_1, 0, 0)$ and $(\varphi_2, 0, 0)$, so that $\sin\varphi_1 = \sin\varphi_2 = \Omega/KG$, and $\cos\varphi_1 = -\cos\varphi_2 = \sqrt{1 - (\Omega/KG)^2}$. The first state can be stable depending on the parameters combination. The second one is unstable.

The stability of the synchronous states, $(\varphi_1, 0, 0)$ and $(\varphi_2, 0, 0)$ can be analyzed by using the characteristic polynomial related to the linear approximation of (2.7), around the equilibrium states

$$P(\lambda) = \lambda^3 + (3-K)\lambda^2 + \lambda + KG\cos\varphi_{\text{SS}}. \tag{3.1}$$

The stability of the synchronous state (SS) is given by the real-part of the roots of $P(\lambda)$. If they are all negative, the corresponding synchronous state is asymptotically stable. If there is a root with positive real part, the corresponding synchronous state is unstable.

According to the Routh-Hurwitz stability criterion [9], the number of positive real-part roots of the polynomial is equal to the number of signal changes in coefficients of the first column of Routh-Hurwitz matrix R_0. For $P(\lambda)$ given by (3.1), R_0 is as follows:

$$R_0 = \begin{bmatrix} 1 & 1 \\ 3-K & KG\cos\varphi \\ \dfrac{3-K-KG\cos\varphi}{3-K} & 0 \\ KG\cos\varphi & 0 \end{bmatrix}. \tag{3.2}$$

As the synchronous state $(\varphi_2, 0, 0)$ has a negative cosine, observing R_0 and according to the Routh-Hurwitz criterion, it can be seen that the first term of the first column is positive and the fourth term is negative. Consequently, there is at least a signal change, and, therefore, $(\varphi_2, 0, 0)$ is unstable.

The synchronous state $(\varphi_1, 0, 0)$ has a positive cosine, consequently, the conditions for its asymptotical stability are $3-K>0$ and $3-K-KG\sqrt{1-(\Omega/KG)^2}>0$. With these conditions, the lock-in range for a third-order PLL is

(i) $1 \leq K < 3$;
(ii) $G > \Omega/K$;
(iii) $G < \sqrt{9/K^2 - 6/K + 1 + \Omega^2/K^2}$.

Then there is only one synchronous state, $(\varphi_1, 0, 0)$ and the lock-in range results from two bifurcations: a saddle-node, related to the existence of the synchronous state, and Hopf, related to the stability of the synchronous state [8].

The condition $G = \Omega/K$ represents a saddle-node bifurcation, whose diagram is shown in Figure 3.1. Below the surface, there is no synchronous state, and on it the state is nonhyperbolic. Above the surface, there are two equilibrium points: $(\varphi_1, 0, 0)$ and $(\varphi_2, 0, 0)$, so that $\sin\varphi_1 = \sin\varphi_2 = \Omega/KG$ and $\cos\varphi_1 = -\cos\varphi_2 = \sqrt{1-(\Omega/KG)^2}$.

Condition $G = \sqrt{9/K^2 - 6/K + 1 + \Omega^2/K^2}$ represents a Hopf bifurcation, as shown in Figure 3.2. Below the surface, the state $(\varphi_1, 0, 0)$ is asymptotically stable, and above it, unstable.

4. OWMS and TWMS networks

In OWMS topology, the transmission of time signals follows only one direction. The master node has its own time basis that is independent of the other nodes. The time basis of all slave nodes depends on only one node, which can be the master or another slave. OWMS networks can be implemented in two topologies [10], single chain and single star.

TWMS networks have reference signals sent in the two ways of the network. The master has its own time basis, but the time basis of each slave depends on more than one node.

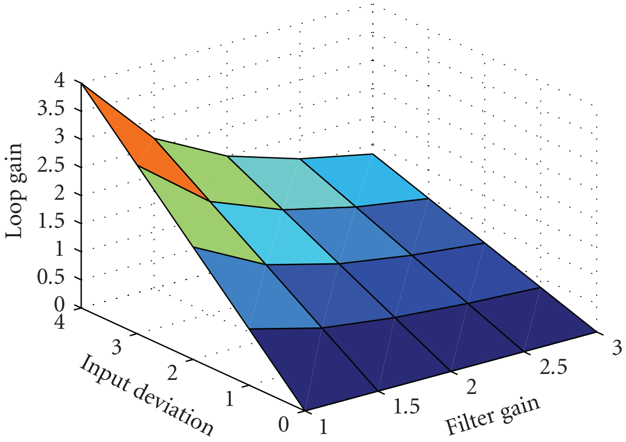


Figure 3.1. Saddle-node bifurcation.

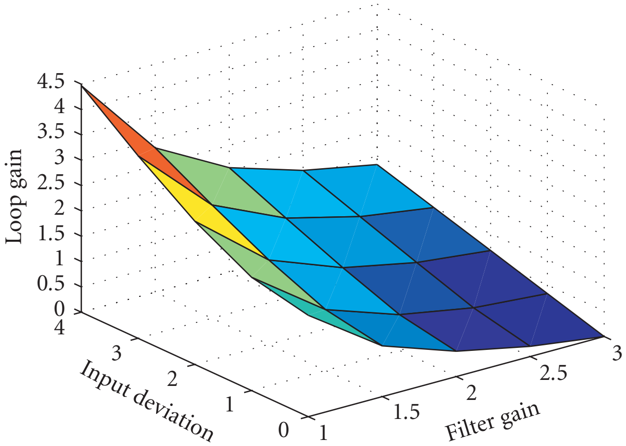


Figure 3.2. Hopf bifurcation.

They can be implemented in four topologies: double chain, double star, single loop, and double loop [10].

Here, only the single and double chains are studied because they are the most common architectures in commercial networks. The single chain topology is composed of a master node, which has an independent time basis, connected to slave nodes in a sequential way, as shown in Figure 4.1. Each slave is a third-order PLL with phase controlled by the node that precedes it in the chain.

The double chain topology is similar to the single chain, but the reference signal, which will be the input of a slave node n, considers the signals from nodes $n-1$ and $n+1$, as shown in Figure 4.2.

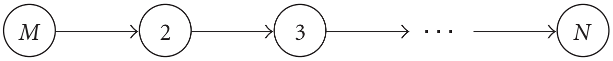


Figure 4.1. Single-chain topology.

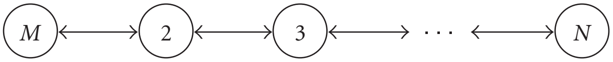


Figure 4.2. Double-chain topology.

In this topology, the time basis of the master node, called node 1, does not depend on the other nodes and is given by a precise and reliable oscillator whose phase is:

$$\phi_M = \omega t + \psi(t), \tag{4.1}$$

where ω represents the frequency of normal operation of master clock and $\psi(t)$ is a perturbation term.

As delays in the main commercial networks are small related to the time constants of the node filters, they are not considered. Consequently, the signal sent by the master to the first slave considers its own phase and the phase of the first slave. Then, the phase of the control signal sent for the first slave is given by the following equation:

$$\Phi_1(t) = 2\Phi_M(t) - \Phi_2(t), \tag{4.2}$$

where $\Phi_2(t)$ represents the phase of node 2, the first slave node.

For this node, the input phase is

$$\Phi_i^{(2)}(t) = \Phi_M(t) - 0.5\Phi_2(t) + 0.5\Phi_3(t). \tag{4.3}$$

From (4.2) and (4.3),

$$\Phi_i^{(2)}(t) = 0.5\Phi_1(t) + 0.5\Phi_3(t). \tag{4.4}$$

As can be seen in (4.4), the input phase of the n-slave node depends on the phase of the nodes $n-1$ and $n+1$. So, for each slave n of the chain, $n = 2,3,\ldots,N-1$,

$$\Phi_i^{(n)}(t) = 0.5\Phi_{n-1}(t) + 0.5\Phi_{n+1}(t). \tag{4.5}$$

For node N, the last of the chain, the input signal will be the output signal of node $N-1$.

5. Numerical experiments

In order to explore the several aspects of the lock-in range and performance parameters, by using MATLAB-Simulink, single and double chain architectures, as described above,

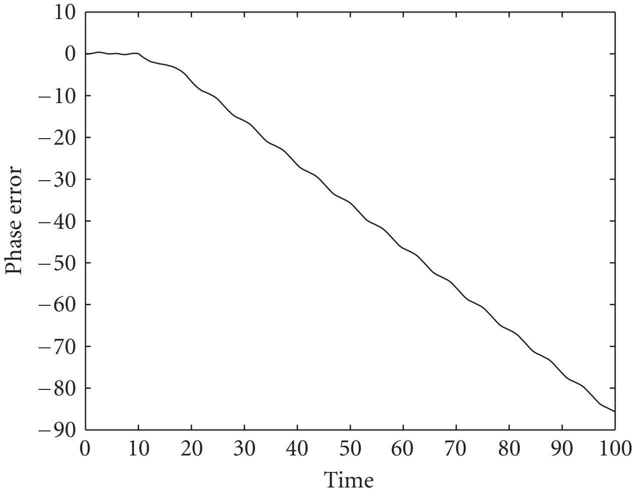


Figure 5.1. Phase error without synchronous state.

were mounted with periodic oscillators as master and built-in PLLs as slaves. Simulations with these architectures were conducted with "Ode-45-Dormand-Prince" [11] with variable step integration method and 10^{-6} of relative tolerance.

The simulations aimed to study the reachability of the synchronous state and to obtain acquisition parameters for the whole network. Input parameters and filter transfer functions were varied trying to confirm the analytical results obtained in Section 3.

The master node was simulated by a periodic signal generator, and a phase ramp starting at tenth second of simulation was added to the master phase in order to analyze the networks capacity in accommodating this perturbation. The input deviation Ω and the free-running angular frequency ω_M were set in 1 rad/s, and the parameters K and G were varied.

The synchronism is analyzed observing the phase error that must be constant in the synchronous state. Network performance includes the double-frequency jitter as it is not completely eliminated being responsible for the oscillations around the synchronous state.

5.1. OWMS.

The chains were mounted according to the single chain topology.

One-slave node chain. For $K = 1$, the lock-in range is analytically given by $G \in (1, \sqrt{5})$.

In the simulations, the synchronous state is reachable for $G \in (1.18, 2.2)$. For $G = 1$, the phase error goes to infinite, as illustrated in Figure 5.1, and for G lower than this value, the behavior is the same.

For values of G in the lock-in range, the phase error presents an equilibrium state with a small oscillation, the double-frequency jitter, around it, as can be seen in Figure 5.2 for $G = 1.2$.

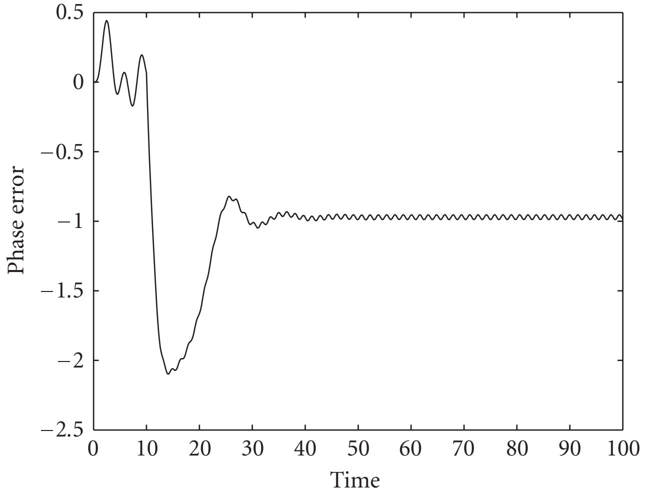


Figure 5.2. Phase error with PLL gain in the lock-in range.

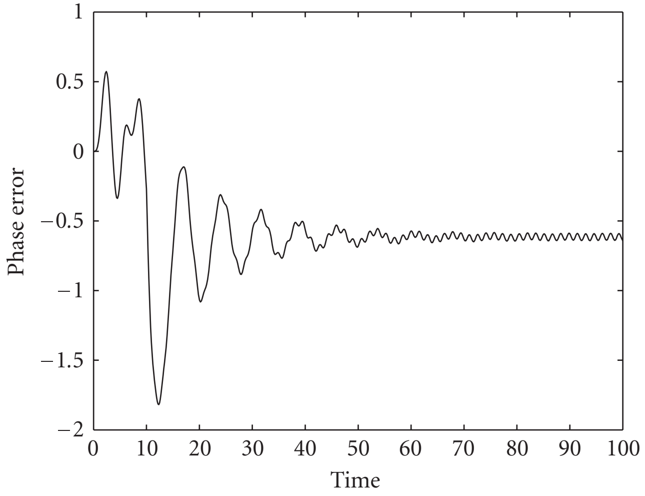


Figure 5.3. Phase error increasing PLL gain in the lock-in range.

When the PLL gain G is increased, the acquisition time and double-frequency jitter increase, as shown in Figure 5.3 for $G = 1.7$.

When the gain value approaches the lock-in range limit ($G = 2.2$), a large amplitude oscillation appears around the synchronous state, as shown in Figure 5.4. With the PLL gain out of this range, there is no synchronous state, as shown in Figure 5.5 for $G = 3.5$.

For $K = 2$, the lock-in range analytically determined is given by $G \in (0.5, \sqrt{2}/2)$.

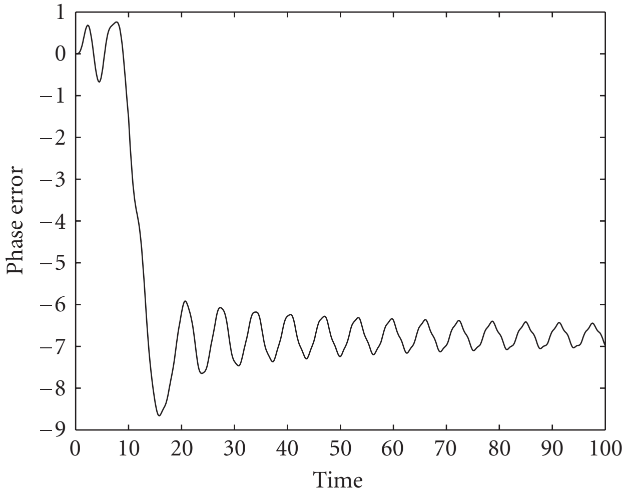


Figure 5.4. Phase error with PLL gain in the lock-in range limit.

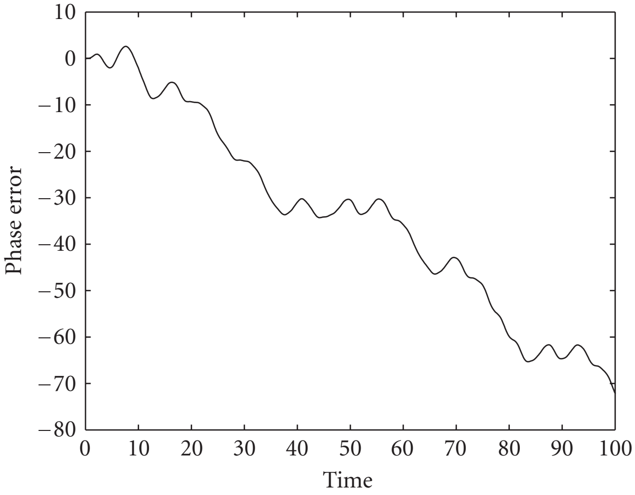


Figure 5.5. Phase error with PLL gain out of lock-in range.

When the simulations are conducted, varying G for the whole theoretical lock-in range, it is experimentally noticed that the synchronous state is not reachable for certain values of G, implying that the practical lock-in range is smaller than the theoretical one. This fact is probably due to nonrobustness of the model in this band of gains.

According to the numerical simulations, the real lock-in range seems to be $[0.5, 0.59) \cup (0.66, \sqrt{2}/2)$. In order to observe this fact, Figure 5.6 shows the result for $G = 0.6$, with the synchronous state not reachable. For $G = 0.67$, reachability is recovered as shown in Figure 5.7.

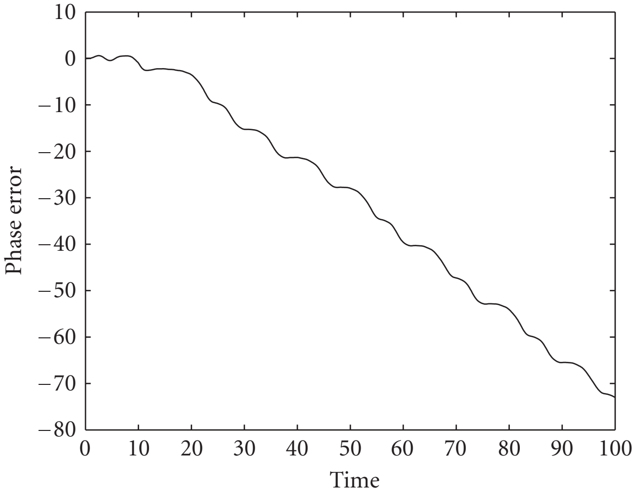


Figure 5.6. Phase error with synchronous state not reachable, but with PLL gain in the theoretical lock-in range.

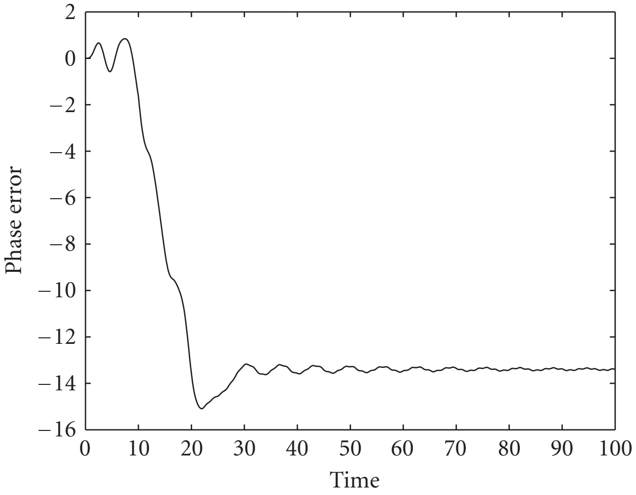


Figure 5.7. Phase error with PLL gain in the lock-in range.

Further experiences show that the higher the parameter K, the smaller the lock-in range. With K next to the limit of the lock-in range $(1 \leq K < 3)$, $K = 2.9$, for instance, corresponding to a theoretical range $(0.344, 0.346)$, the synchronous state is not reachable for any G.

The number of slave nodes was increased up to ten gradually and each chain was simulated in accordance with the methodology used in the previous item. It was observed that to each slave node added to the chain, the synchronizing ranges become smaller, and the settling time increases.

Therefore, the reference signal of the network loses quality while it is transmitted along the chain. Due to this fact, according to G.812 recommendation of ITU-T [12], the higher number of sequential nodes recommended for a single chain is ten, and in the following, results for this type of chain are presented.

Ten-slave node chain. With ten slave nodes, the synchronous state is still reachable, but for a smaller lock-in range, as listed below:

(i) for $K = 1$, the theoretical lock-in range is $G \in (1, \sqrt{5})$. In the simulations, the synchronous state is reachable for $G \in (1.63, 1.64)$;

(ii) for $K = 1.5$, the theoretical lock-in range is $G \in (0.667, 1.2)$, but the synchronous state is reachable for $G \in (0.97, 0.98)$;

(iii) for $K = 2$, the theoretical lock-in range is $G \in (0.5, \sqrt{2}/2)$, and in the simulations, the synchronous state is reachable only for $G = 0.63$ after 400 seconds of simulation;

(iv) for $K = 2.5$, the theoretical lock-in range is $G \in (0.4, 0.447)$, and the synchronous state is reachable for the following gain values: 0.416, 0.436, 0.438, 0.44, 0.443, and 0.444.

Figure 5.8 shows the phase error for $K = 2.5$ and $G = 0.416$. For the tenth node, the settling-time is high but the double-frequency jitter disappeared, due to the fact that the signal passed through the ten lowpass filters of the chain.

Then, considering the lock-in ranges, the behavior of the whole network with ten slave nodes is satisfactory, showing that the third-order PLL with a Sallen-Key filter for extraction of reference signal in OWMS provides good performance figures.

5.2. TWMS.

The analysis of TWMS networks [1] strongly depends on the number of nodes [13], consequently, the simulations follow an increasing sequence of the number of slave nodes.

Two-slave node chain. Running the simulations, the lock-in ranges are the following:

(i) for $K = 1$, the synchronous state is reachable only for $G = 1.3$;

(ii) for $K = 1.5$, the synchronous state is reachable for $G \in (0.8, 0.9)$;

(iii) for $K = 2$, in the simulations, the synchronous state is reachable only for $G = 0.6$;

(iv) for $K = 2.5$, the synchronous state is reachable for $G \in (0.4, 0.5)$, that is, for a lock-in range greater than the theoretical one.

Figure 5.9 shows the phase error in the two slave nodes for $K = 2.5$ and $G = 0.42$. The second node presents a higher settling-time but better double-frequency jitter performance, as expected.

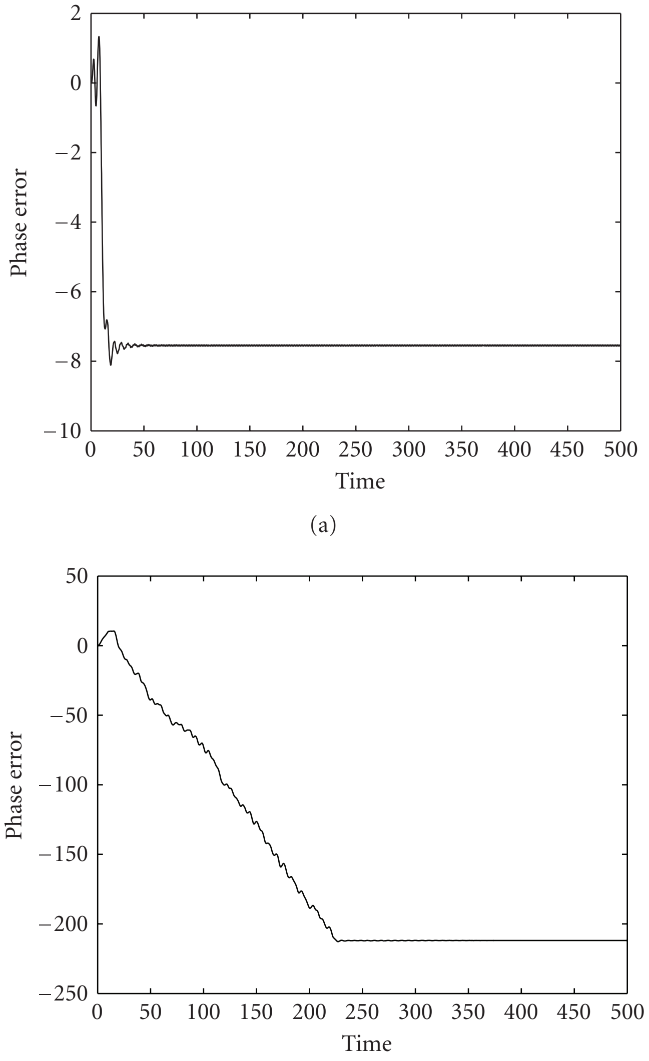
Three-slave node chain. Lock-in ranges are the following:

(i) for $K = 1$ there was no synchronization;

(ii) for $K = 1.5$ the synchronous state is reachable only for $G = 0.7$;

(iii) for $K = 2$, the synchronous state is reachable for $G \in (0.5, 0.6)$;

(iv) for $K = 2.5$, the synchronous state is reachable for $G \in (0.43, 0.48)$, surpassing the theoretical lock-in range again.



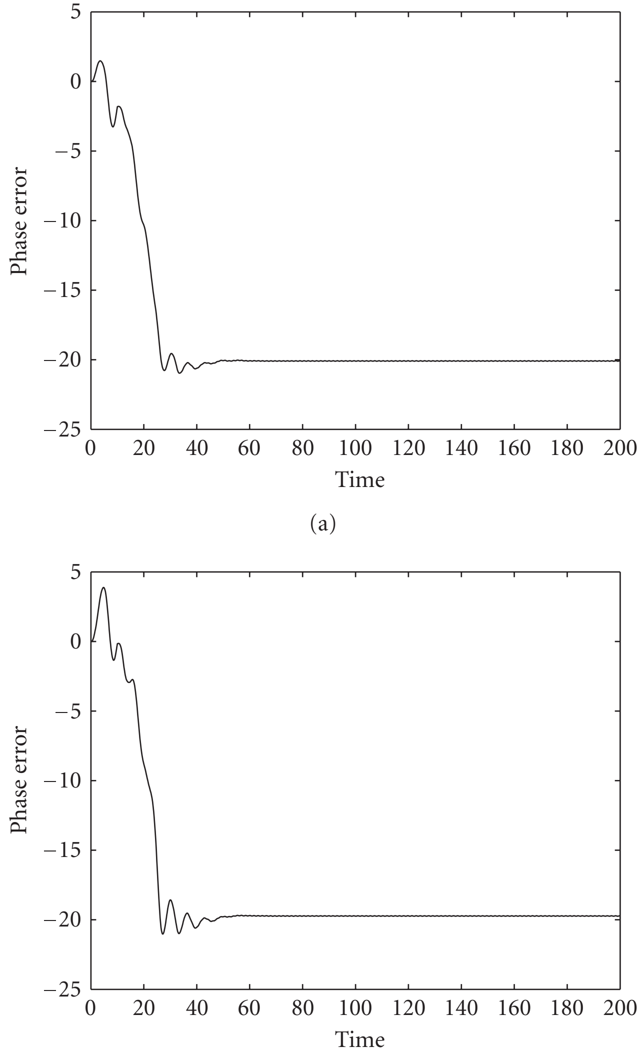
(a)

(b)

Figure 5.8. Phase error (a) first slave node (b) tenth slave node.

In this case, it is relevant to notice that the settling-time increased and varies from 100 to 250 seconds. Besides, the lock-in range was reduced as listed above.

Four-slave node chain. This chain approached the synchronous state only for $K = 2.5$ and $G = 0.43$, having a considerable oscillation around the synchronous state, as shown in Figure 5.10 for the first and fourth slave nodes. Also it is noticed that a small increase of this oscillation with the increase in the number of slaves. Jitter did not present any significant alteration.



(a)

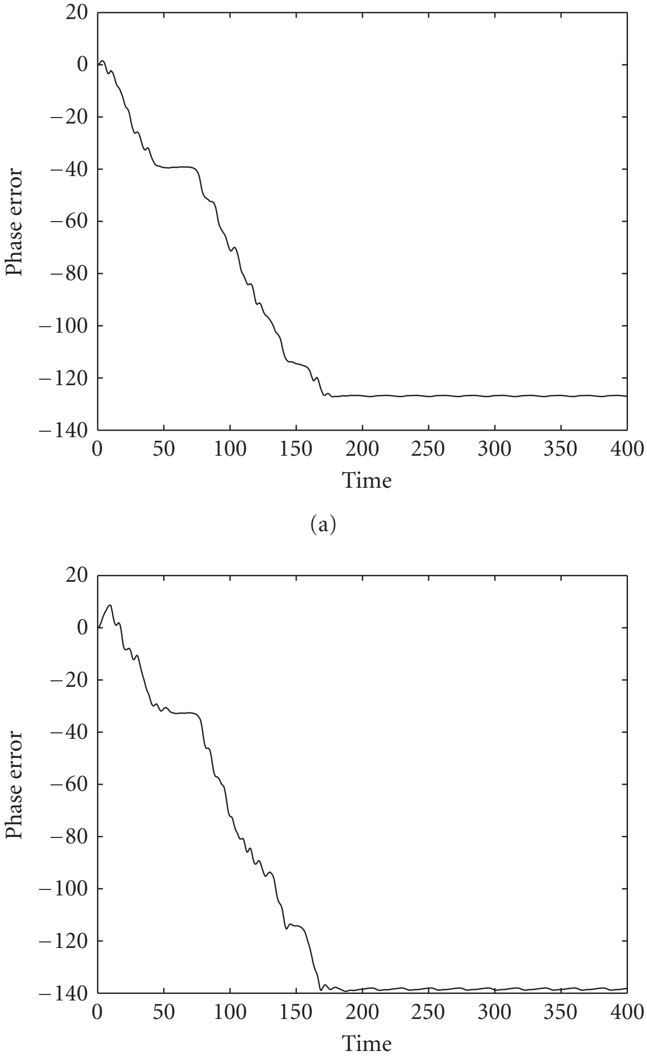
(b)

Figure 5.9. Phase error (a) first slave node (b) second slave node.

Five-slave node chain. In these conditions, the synchronous state is not reachable. The simulation time was increased to 500 seconds, but for all the values of parameters tested, the phase error went to infinite in all the slaves.

Then it is noticed that the increase in the chain makes the synchronization of TWMS networks difficult considerably, having a limited number of slave nodes, above which the behavior of the network becomes totally unstable.

This result is in accordance to [13] that claims that TWMS networks present limitations in the number of slave nodes that should not be higher than three when a first-order



(a)

(b)

Figure 5.10. Phase error (a) first slave node (b) fourth slave node.

lag filter is used. Simulations have shown that the critical number of slave nodes for third-order PLLs with Sallen-Key filter is four.

6. Conclusions

Third-order PLLs provide satisfactory conditions of synchronism, transient response and double-frequency jitter attenuation when used as slave nodes in MS chain networks. Simulations confirmed analytical results, also showing the reliability of the usual PLL models.

Increasing the number of nodes in a chain makes synchronization more difficult reducing the lock-in range of the OWMS and the TWMS networks. As some theoretical studies show [13], for TWMS chains there is a maximum number of nodes above which the synchronous state is not reachable. This number, when third-order PLLs with Sallen-Key filter are used, is four.

Comparing single-chain (OWMS) with double-chain (TWMS), the OWMS architecture is more indicated for precise clock distribution, since it supports more slave nodes and presents a larger lock-in range.

The limitation in the number of nodes and in the lock-in range for TWMS is the main reason to use this architecture only for process-control in local area networks. In this case, in spite of these problems, transient responses and jitter performance are considerably improved.

Acknowledgment

The authors are supported by CNPq.

References

[1] W. C. Lindsey, F. Ghazvinian, W. C. Hagmann, and K. Dessouky, "Network synchronization," *Proceedings of the IEEE*, vol. 73, no. 10, pp. 1445–1467, 1985.

[2] S. Bregni, *Synchronization of Digital Networks*, Jonh Wiley & Sons, England, UK, 1st edition, 2002.

[3] W. C. Lindsey, *Syncronization Systems in Communication and Control*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1972.

[4] R. E. Best, *Phase-Locked Loops*, McGraw-Hill, New York, NY, USA, 4th edition, 1999.

[5] J. R. C. Piqueira and L. H. A. Monteiro, "Considering second-harmonic terms in the operation of the phase detector for second-order phase-locked loop," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 6, pp. 805–809, 2003.

[6] J. R. C. Piqueira, E. Y. Takada, and L. H. A. Monteiro, "Analyzing the effect of the phase-jitter in the operation of second order phase-locked loops," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 52, no. 6, pp. 331–335, 2005.

[7] J. R. C. Piqueira, "Aplicação da Teoria Qualitativa de Equações Diferenciais a Problemas de Sincronismo de Fase," Tese de doutorado, Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil, 1987.

[8] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, vol. 42 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1983.

[9] K. Ogata, *Modern Control Engineering*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1997.

[10] J. R. C. Piqueira, "Uma contribuição ao estudo das redes com malhas de sincronismo de fase," Tese de Livre-Docência, Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil, 1997.

[11] S. Lynch, *Dynamical Systems with Applications Using MATLAB*, Birkhäuser Boston, Boston, Mass, USA, 2004.

[12] *G.812 Timing Requirements of Slave Clocks Suitable for Use as Node Clocks in Synchronization Networks*, ITU-T: 1997.

[13] J. R. C. Piqueira, S. A. Castillo-Vargas, and L. H. A. Monteiro, “Two-way master-slave double-chain networks: limitations imposed by linear master drift for second order PLLs as slave nodes,” *IEEE Communications Letters*, vol. 9, no. 9, pp. 829–831, 2005.

José Roberto Castilho Piqueira: Escola Politécnica, Universidade de São Paulo, Avenida Prof. Luciano Gualberto, travessa 3, no. 158, 05508-900 São Paulo, SP, Brazil
Email address: piqueira@lac.usp.br

Marcela de Carvalho Freschi: Escola Politécnica, Universidade de São Paulo, Avenida Prof. Luciano Gualberto, travessa 3, no. 158, 05508-900 São Paulo, SP, Brazil
Email address: marcela.freschi@poli.usp.br