Complexity

Parallel Analysis, Control, and Intelligence of Cyber-Physical-Social Systems

Lead Guest Editor: Ning Cai
Guest Editors: Jian-Xiang Xi, Roberto Sabatini, Yi-Sheng Lv, and Muhammad Junaid Khan

Parallel Analysis, Control, and Intelligence of Cyber-Physical-Social Systems

Complexity

Parallel Analysis, Control, and Intelligence of Cyber-Physical-Social Systems

Lead Guest Editor: Ning Cai
Guest Editors: Jian-Xiang Xi, Roberto Sabatini, Yi-Sheng Lv, and Muhammad Junaid Khan

Contents

Hindawi
Complexity
Volume 2021, Article ID 6684849, 13 pages
https://doi.org/10.1155/2021/6684849

Research Article

Distributed Event-Triggered Circle Formation Control for Multiagent Systems with Nonuniform Quantization

Jiayan Wen,[1,2] **Haijiang Zhang**,[1,2] **Guangxing Tan**,[1] **Ning Cai**,[3] **and Guangming Xie**[1,4]

[1]*School of Electrical and Information Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China*
[2]*Guangxi Key Laboratory of Automobile Components and Vehicle Technology in Guangxi University of Science and Technology, Liuzhou, China*
[3]*College of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[4]*College of Engineering, Peking University, Beijing 100871, China*

Correspondence should be addressed to Jiayan Wen; wenjiayan2012@126.com

Received 10 December 2020; Accepted 28 July 2021; Published 9 August 2021

Academic Editor: Abdellatif Ben Makhlouf

This article focuses on circle formation control problem of multiagent systems based on event-triggered strategy under limited communication bandwidth. In such system, each agent can only perceive the angular distance of its nearest neighbor in the counterclockwise direction, and the angular distance of the nearest neighbor in the clockwise direction needs to be obtained by communicating with each other. In order to address the aforementioned problem, a novel distributed algorithm based on the combination of nonuniform quantitative communication technology and event-triggered control is proposed. Sufficient conditions on circle formation control are derived under which the states of all agents can be confirmed to converge to some desired equilibrium point. Different from the traditional uniform quantization communication framework, nonuniform quantization can be beneficial for handling small signals and improving the performance of multiagent systems concerned. Furthermore, under the proposed policy, all the designed quantizers do not emerge saturated. Numerical simulation results are provided to verify the effectiveness of the proposed algorithm.

1. Introduction

According to the new research literature, formation control of multiagent systems (MASs), which is oriented to design an appropriate control protocol so that all agents can maintain a prescribed geometric shape, has received significant attention owing to its broad applications [1–3]. Circle formation control as a datum problem in formation control of MASs is widely investigated in sundry areas [4–6]. In many actual application scenarios, the components of the system will be scattered in a wide area, and the exchange of information between components (including controllers, actuators, and wireless sensors) is mainly implemented via digital communication networks [7]. In theoretical analysis, the system is usually regarded as an ideal state. More specifically, for this ideal situation, it is assumed that each agent does not experience packet loss and distortion when performing information interaction with its neighbors [8]. The requirement of the basic assumption is that all communication channels between agents possess a sufficiently large bandwidth or unlimited capacity. However, it is contradictory to actual system applications, that is to say, for the practical application case, digital network resources often may be limited by different degrees of energy and communication bandwidth due to some reasons [9]. In particular, when it comes up to the large-scale MASs, the limitation of network communication bandwidth will not only affect the quality of data transmission but also cause the overall performance of the system to decrease. As a result, considering the limited communication capacity of the whole digital network, it is necessary to pay much attention when designing the control protocols with close to practical applications.

In view of the limitation of network resources of MASs, a variety of control algorithms involving information quantification have gradually emerged. In the pioneering works

[10, 11], the quantization communication based on integer-valued and real-valued has been analyzed, respectively. Aiming at the shortcomings of the static quantizer designed in [10, 11], the authors of [12] have further proposed a dynamic encoder-decoder quantization algorithm. Then, Li et al. [13] characterized a novel control method that can be symmetrically compensated by adjusting the corresponding parameters of the controller to reduce the number of bits transmitted by each digital channel to only 1 bit. Thereafter, the theory that is closely related to the quantization algorithm of networked systems has been systematically studied in [14], where the authors provided a comprehensive analysis of the average consensus of MASs with a finite number of quantization levels by introducing a scaling function into the encoding-decoding quantizer. The extensions of the authors' work in [13] are further investigated for cases: the agents being governed by the second-order linear dynamics [15, 16], discrete linear systems [17, 18], and nonlinear systems [19, 20], respectively, the presence of communication channel time-delays, and so on. Xu et al. [21] have considered the leader-following fixed-time quantitative consensus problem of the nonlinear multiagent systems with a novel way of impulse control.

Owing to the limited bandwidth and energy resource, circle formation control of multiagent systems over directed graph with quantized communications is significant from both theoretical and engineering points of view. Furthermore, as the questions arising pointed out in [14], the designed protocols in the existing literature aforementioned are executed synchronously, and they need to be updated at each time step. This may cause much unnecessary energy consumption, especially in an environment with limited resources. With the growing demand in industry on systematic methods to model, analyze, and design systems, event-triggered control (ETC) has been proposed as a promising control mode to solve the above problems [22, 23]. Fortunately, this ETC framework has been applied to deal with the quantized consensus problems of MASs. In [24, 25], the authors introduced ETC into the second-order system, which proved that it can effectively reduce the calculation amount of the system and reduce the update frequency of the controller. In [26], a nonlinear decomposition method of asymmetric hysteresis quantizer is proposed by using the fuzzy logic system to estimate random perturbation term and unknown nonlinear function, and the sector constraint property is used. In [27–30], the authors have investigated the quantized consensus problem of the general linear systems and nonlinear systems, respectively, with the couple of ETC fashion. To the best knowledge of the authors, most studies have focused on the analysis and design of uniform quantification combined with ETC strategy to solve the formation control problem of MASs, and few results are devoted to concentrate on the co-design structure between nonuniform quantification and ETC scheme compared with the former counterpart, especially few for the datum problem of circle formation. Thus, it is natural to motivate us to consider the novel combination involving nonuniform quantizer together with the ETC mechanism, in which nonuniform quantization can enhance the quantization signal-to-noise ratio of the small signal and the key information, which is extracted in the small signal, can be ignored easily [31–33].

The core of this paper is dedicated to consider the circle formation problem of MASs under the constraints of communication bandwidth and limited energy as close to reality as possible. Here, we focus on the special case that each agent can only perceive the angular distance from itself to the nearest neighbor in the counterclockwise direction, while the counterpart in the clockwise direction be acquired via the digital communication network. The main contributions of this paper mainly are threefold. First, a novel algorithm is proposed to tackle the circle formation problem of MASs, in which the nonuniform dynamic quantizer plays a crucial role in data compression and transmission. Second, a distributed event-triggered condition that only relies on the local information of neighboring agents is constructed. Finally, the proof that the designed nonuniform dynamic quantizers will never be saturated is strictly provided, and the comprehensive comparison of advantages and disadvantages of uniform and nonuniform quantizer is shown in detail. In summary, the algorithm proposed in this paper may be favorable to reveal the practical constraints that originated from physical control systems.

An outline of this paper is organised as follows. In Section 2, some basic background and problem statements are given. In Section 3, a new encoder-decoder nonuniform quantizer is designed and the ETC conditions are proposed. Section 4 analyzes the convergence of the proposed control law and proves that the nonuniform quantizer is always in an unsaturated state. Section 5 verifies the feasibility and superiority of the new algorithm through comparative analysis. The conclusion of this article and the prospect of subsequent research directions are referred in Section 6.

2. Preliminaries and Problem Statement

In this section, some symbols and basic concepts of algebraic graph theory are collected together. Then, the definition of circle formation and its related properties is mentioned.

2.1. Preliminaries. $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{N}$ denote the sets of real, positive real, nonnegative real, and positive integers, respectively. $\mathscr{A}^T$, $\|\mathscr{A}\|$, and $\|\mathscr{A}\|_\infty$ denote the transpose, Euclidean norm, and infinite norm for a vector or matrix $\mathscr{A}$, respectively. The set of all real matrices with m rows and n columns is denoted as $\mathbb{R}^{m\times n}$. For an arbitrary vector x, $x_{\geq 0}$ means that each element in the vector x is nonnegative. N-dimensional column vector whose elements are 1 and 0 is denoted as 1_N and 0_N, respectively. I_N is defined as an N-dimensional identity matrix. Given a positive number z, $\lfloor z \rfloor$ represents the integer round-down of x. $\otimes$ represents the Kronecker product, which has the following properties:

(1) $(A\otimes B)(C\otimes D) = (AC)\otimes(B\ D)$

(2) $(A\otimes B)^T = A^T\otimes B^T$

For a weighted directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E}_{\mathscr{G}}, \mathscr{A}_{\mathscr{G}})$, it contains a set of vertices $\mathscr{V} = \{1, 2, \ldots, N\}$, the edge set

$\mathscr{E}_{\mathscr{G}} \subseteq \mathscr{V} \times \mathscr{V}$, and weighted adjacency matrix $\mathscr{A}_{\mathscr{G}} \in \mathbb{R}_{\geq 0}^{N\times N}$. Self-edges (i,i) are not allowed, i.e., $(i,i) \notin \mathscr{E}_{\mathscr{G}}$ for all $i \in \mathscr{V}$. For $\mathscr{G}$, the corresponding weighted adjacency matrix $\mathscr{A}_{\mathscr{G}} = [a_{ij}] \in \mathbb{R}_{\geq 0}^{N\times N}$ is defined element-wise with $a_{ij} > 0$ if $(i,j) \in \mathscr{E}_{\mathscr{G}}$ and $a_{ij} = 0$ otherwise. If there exists an edge $(i,j) \in \mathscr{E}_{\mathscr{G}}$, then we refer to j as an out-neighbor of i and i as an in-neighbor of j. For a given node i, the sets of out- and in-neighbors are denoted by N_i^+ and N_i^-, respectively. A directed path from vertex i to j in a digraph is an ordered sequence of edges starting with i and ending with j. If there exists a directed path for any two distinct vertices i and j in $\mathscr{G}$, the digraph $\mathscr{G}$ is called strongly connected. Accordingly, both the out- and in-degree matrices $\mathscr{D}^+ = [d_1^+, d_2^+, \ldots, d_N^+]$ and $\mathscr{D}^- = [d_1^-, d_2^-, \ldots, d_N^-]$ are diagonal matrices, where $d_i^+(t) = \sum_{j\in N_i^+}^{N} a_{ij}$ and $d_i^-(t) = \sum_{j\in N_i^-}^{N} a_{ij}$. A digraph is called weight-unbalanced if $\mathscr{D}^+ \neq \mathscr{D}^-$, and the degree of $\mathscr{G}$ is defined as $d^* = \max_{i\in\mathscr{V}} d_i^+$. The Laplacian matrix is defined as $\mathscr{L} = \mathscr{D}^+ - \mathscr{A}$, which is given in this paper.

Lemma 1 (see [4]). *For a weighted directed graph $\mathscr{G}$, there exists the following properties:*

(1) The Laplace matrix $\mathscr{L}$ has a zero eigenvalue with associated eigenvector 1_N.

(2) If the graph $\mathscr{G}$ contains a spanning tree, the algebraic multiplicity of eigenvalue zero is simple, and the rest of eigenvalues have positive real parts.

(3) If graph $\mathscr{G}$ is strongly connected, then its Laplacian matrix $\mathscr{L}$ is irreducible and satisfies $\mathscr{L}1_N = 0_N$. And, there exists a vector $\xi = [\xi_1, \xi_2, \ldots, \xi_N]^T > 0$ satisfying $\xi^T\mathscr{L} = 0_N^T$ and $\xi^T 1_N = 1$.

Lemma 2 (see [13]). *If the unbalanced directed graph $\mathscr{G}$ is strongly connected and defines $\rho_\lambda = \max_{2\leq i\leq N|1-h\lambda_i|}$, $h \in (0, 1/d^*)$, then the Laplacian matrix $\mathscr{L}$ can be decomposed into $\mathscr{L} = \mathscr{T}^*$, where $\mathscr{T}^* = [\xi, \phi_2, \ldots, \phi_N] \in \mathbb{R}^{N\times N}$ and $\mathscr{T} = [1_N, \upsilon_2, \ldots, \upsilon_N] \in \mathbb{R}^{N\times N}$ are both nonsingular matrices. $\mathscr{M} \in \mathbb{R}^{N\times N}$ is the matrix $\mathscr{L}$ corresponding to the Jordan standard shape, and its first diagonal element is zero. A variable $\rho(\zeta, h) = \rho_1(h)/\zeta$, $\zeta \in (\rho_\lambda, 1)$, is defined, where $\rho_1(h)$ represents the matrix $I_N - h\mathscr{M}$. A submatrix is removed from the first row and the first column, and existing real numbers are as follows:*

(1) $\mathscr{M}_{\lambda'}(\xi, h) = \max_{l\in N}\|\rho^l(\xi, h)\|$

(2) $\mathscr{M}_{\lambda}(\xi, h) = \lim_{l\longrightarrow\infty}\sum_m^l \|\rho^l(\xi, h)\|$

2.2. Problem Statement. Consider a system composed of N agents, $N \geq 2$, in which each agent is an independent individual with autonomous ability and moves along a preset circle. Initially, all agents are randomly located in different positions on the circle. Here, for the ease of analysis, we marked the agents as 1 to N in the counterclockwise direction, denoted as $x = [x_1, x_2, \ldots, x_N]^T \in \mathbb{R}^{\mathbb{T}}$, as shown in Figure 1. In a fixed coordinate system, the state of the angular position of the agent i at time t is denoted by $x_i(t)$. Then, the initial positions of all agents are set to satisfy

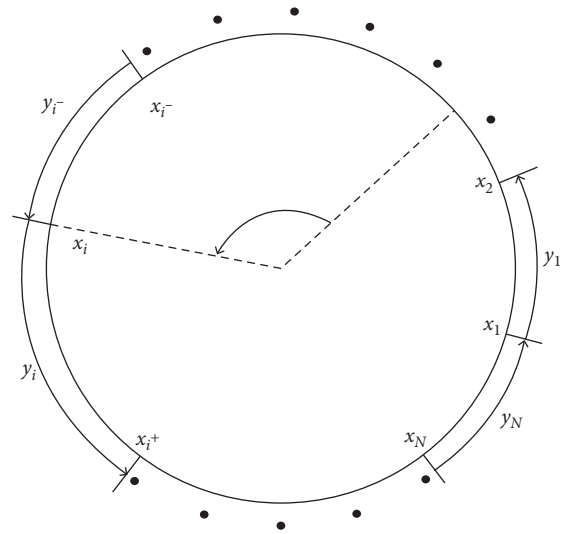


Figure 1: Mobile agents located on a circle.

$$0 \leq x_1(0) \leq \cdots \leq x_i(0) \leq x_{i+1}(0) \leq \cdots \leq x_N(0) \leq 2\pi. \quad (1)$$

In this paper, we mainly focus on the case that each agent only has two neighbors that are immediately in front of or behind itself. We denote the set of agent i's two neighbors by $N_i = [i^+, i^-]$, where

$$i^+ = \begin{cases} i+1, & i = 1, 2, \ldots, N-1, \\ 1, & i = N, \end{cases}$$
$$i^- = \begin{cases} N, & i = 1, \\ i-1, & i = 2, 3, \ldots, N. \end{cases} \quad (2)$$

Denote $d_i \in \mathbb{R}$ as the desired angular distance between individual i and its neighbor i^+. The information exchange relationship between agents is further established in such a network, where each agent i (equipped with a unidirectional sensor) can only perceive an angular distance from i to i^+ and from i to i^-. The counterpart is obtained through the shared communication network. In this setting, the communication network between agents can be described by a weight-unbalanced directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E}_{\mathscr{G}}, \mathscr{A}_{\mathscr{G}})$, where $\mathscr{V} = \{1, 2, \ldots, N\}$, $\mathscr{E}_{\mathscr{G}} = \{(1,2), (2,3), \ldots, (N-1, N)\}$, and $\mathscr{A}_{\mathscr{G}} = [a_{ij}] \in \mathbb{R}_{\geq 0}^{N\times N}$, as shown in Figure 2.

We consider the network of N agents with the dynamics

$$x_i(t+1) = x_i(t) + hu_i(t), \quad t = 0, 1, 2, \ldots; i = 1, 2, \ldots, N, \quad (3)$$

where $x_i(t) \in \mathbb{R}$ is the scalar state and $u_i(t) \in \mathbb{R}$ is the control input of the agent i. Let $y_i(t) \in (0, 2\pi), i \in \{1, 2, \ldots, N\}$, denote the actual angular distance between the agent i and its counterclockwise neighboring agent i^+. Thus,

$$y_i(t) = \begin{cases} x_{i^+}(t) - x_i(t) & i = 1, 2, \ldots, N-1, \\ x_{i^+}(t) - x_i(t) + 2\pi & i = N. \end{cases} \quad (4)$$

Stack vector $y(t) = [y_1(t), y_2(t), \ldots, y_N(t)]^T \in \mathbb{R}^{\mathbb{T}}$. Note that $y_i(t)$ is local information that can be acquired by a

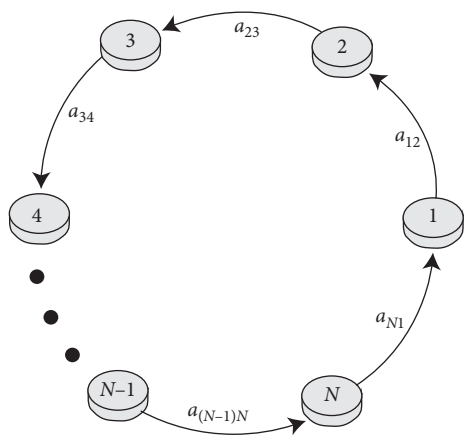


Figure 2: A weight-unbalanced digraph $\mathscr{G}$ with N agents.

unidirectional sensor equipped with agent i. In addition, $\sum_{i=1}^{N} y_i(t) \equiv 2\pi$ is always satisfied. The desired circle formation of the MAS is determined by the vector $d = [d_1, d_2, \ldots, d_N]^T \in \mathbb{R}^{\mathbb{T}}$, where d_i represents the required angular distance between agents i and i^+. If d satisfies $d_i > 0, \forall i \in \mathscr{V}$, and $\sum_{i=1}^{N} d_i = 2\pi$, it means that the formation of this desired circle is acceptable.

Now, we are ready to draft the definition of the circle formation problem.

Definition 1 (circle formation problem, see [9]). Given an admissible circle formation characterized by $d \in \mathbb{R}_{\geq 0}^{\mathbb{N}}$, design distributed control laws $u_i(t) \in \mathbb{R}, i = 1, 2, \ldots, N$, such that, under any initial condition (1), the solution to system (2) converges to some equilibrium point x^*, which satisfies $y^* = d$.

Based on the interaction between agents, our main goal in this paper is to explore a meaningful hybrid design nonuniform between quantization techniques and event-triggered mechanisms to solve the circle formation problems with limited communication bandwidth and limited energy:

$$L = \begin{bmatrix} \frac{d_2}{d_2+d_1} + \frac{d_1}{d_1+d_{\mathcal{N}}} & -\frac{d_1}{d_2+d_1} & 0 & \cdots & 0 & -\frac{d_1}{d_1+d_N} \\ -\frac{d_2}{d_2+d_1} & \frac{d_3}{d_3+d_2} + \frac{d_2}{d_3+d_1} & -\frac{d_2}{d_3+d_2} & \cdots & 0 & 0 \\ 0 & -\frac{d_3}{d_3+d_2} & \frac{d_4}{d_4+d_3} + \frac{d_3}{d_4+d_3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{d_{\mathcal{N}}}{d_{\mathcal{N}}+d_{\mathcal{N}-1}} + \frac{d_{\mathcal{N}-2}}{d_{\mathcal{N}-1}+d_{\mathcal{N}-2}} & -\frac{d_{\mathcal{N}-1}}{d_1+d_{\mathcal{N}-1}} \\ -\frac{d_{\mathcal{N}}}{d_1+d_{\mathcal{N}}} & 0 & 0 & \cdots & -\frac{d_{\mathcal{N}}}{d_{\mathcal{N}}+d_{\mathcal{N}-1}} & \frac{d_1}{d_1+d_{\mathcal{N}}} + \frac{d_{\mathcal{N}-1}}{d_{\mathcal{N}}+d_{\mathcal{N}-1}} \end{bmatrix}. \quad (5)$$

3. Distributed Control Law Design

With reference to the communication protocol based on sampling data proposed in [4–6, 9], the following definition is obtained:

$$u_i(t) = \frac{d_{i^-}}{d_i + d_{i^-}} y_i(t) - \frac{d_i}{d_i + d_{i^-}} y_{i^-}(t), \quad t = 0, 1, 2, \ldots; i = 1, 2, \ldots, N, \quad (6)$$

where i^- represents the neighbor of agent i. In (6), for the agent i, the fulfillment of the control input requires the exact state information of the agent i^-. However, the case considered in this paper is that the sensor is unidirectional, and we assume that exact information on the angular distance from the agent i to i^- is unavailable. To this end, the agents in each pair (i, i^-) need to use digital communication channels to exchange information. As described in [4], compared with

analog signals, digital signals show many advantages, such as signal security, robustness, strength, and less noise interference. Therefore, before transmission, the actual value of the corresponding number of agents i^- should be quantified. The quantitative communication scheme between agents i and i^- consists of a dynamic encoder-decoder pair. Then, the corresponding information exchange can be regarded as such a process, at each time step, the relevant state value of the sender is first encoded into symbol data and then transmitted. After receiving the data, the receiver will activate the decoder to obtain an estimate of the relevant state value of the sender.

Following the design of distributed control law (6), some intermediate variables need to be introduced. Define $\delta_j(t) = y_j(t)/d_j$ and $\hat{\delta}_j(t) = \hat{y}_j(t)/d_j = \xi_j(t_k^j)$, where $\xi_j(t_k^j)$ represents the sample value of agent j at t_k^jth event instant, $j \in \mathscr{V}$. It is worth pointing out that variable conversion will not affect the design of the encoder to the decoder nor will it affect the system stability analysis of the entire encoder.

Next, in a sense, a collaborative design idea is adopted, which is to propose an event-triggered law of nonuniform quantization algorithms combined to solve the problem of circle formation. For each agent $j = \{1, 2, \ldots, N\}$, the corresponding event-triggered encoder Φ_j is designed as follows:

$$\begin{cases} \xi_j(0) = 0, \\ \xi_j(t) = g(t)s_j(t) + \xi_j\left(t_k^j\right), \\ s_j(t) = q_t\left(\dfrac{1}{g(t)}\left(x_j(t) - \xi_j\left(t_k^j\right)\right)\right). \end{cases} \tag{7}$$

We use $q_t(z)$ to represent a nonuniform quantizer with a finite number of levels. Its role is to map the state deviation value of the agent to the discrete level value of the quantizer. Let $z = [z_1, z_2, \ldots, z_N]^T \in \mathbb{R}^T$ and $Q(z) = [q_t(1), q_t(2), \ldots, q_N(z)]^T \in \mathbb{R}^T$. $e_{qi} = (s_j(t) - (1/g(t))(x_j(t) - t\xi_j n(t_k^j)))$ denotes the quantization error. $q_t(z)$ can be described as

$$q_t(z) = \begin{cases} 0, & e^{-(1/2)\beta} - 1 < z < e^{(1/2)\beta} - 1, e^{((2k-1)/2)\beta} - 1 \le z < e^{((2k+1)/2)\beta} - 1, \\ k\beta, & k \in \{1, 2, \ldots, \Omega\}, \\ \Omega\beta, & z \ge e^{(\Omega+(1/2)\beta)} - 1, \\ -q_t(-z) & z \le e^{-(1/2)\beta} - 1, \end{cases} \tag{8}$$

where $\beta > 0$ is the quantization interval and Ω is the number of quantization level. In each iteration, the communication channel $(i, j) \in \mathscr{E}_{\mathscr{G}}, i \ne j$, is required to have ability of transmitting $\log_2(2\Omega + 1)$ bits. If the conditions $z \le e^{(\Omega+(1/2)\beta)} - 1$ and $|q_t(z) - z| \le e^{(\Omega+(1/2)\beta)} - 1$ are satisfied in the process of information interaction between agents, then the quantizer is not saturated.

Further, the neighboring agent i receives $s_j(t)$, which will be estimated by the decoder Ψ_{ji} to obtain the state value of agent j. There is a fact that the communication channel is considered to have no noise in the encoding-decoding process. Ψ_{ji} is described as

$$\begin{cases} \delta_{ji}(0) = 0, \\ \delta_{ji}(t) = g(t-1)\left(e^{s_j(t)} - 1\right) + \delta_{ji}\left(t_k^j\right), \end{cases} \tag{9}$$

where $t = 1, 2, \ldots, N$ and $\delta_{ji}(t) \in \mathbb{R}$ denotes the output of Ψ_{ji}. Relying on the information transmission scheme of the dynamic encoder-decoder described above, the distributed event-triggered circle formation control law of agent i is further designed as

$$u_i(t) = \frac{d_i d_{i^-}}{d_i + d_{i^-}}\left(\xi_i\left(t_k^i\right) - \delta_{i^-}\left(t_k^{i^-}\right)\right). \tag{10}$$

In above formula (10), i^- represents the counterclockwise neighbor of the agent i in the clockwise direction. The latest update time and the next update time of the agent i are, respectively, represented as t_k^i and t_{k+1}^i, $t \in [t_k^i, t_{k+1}^i)$.

Remark 1. It can be observed from encoder-decoder pair (7) and (9) that the advantages are as follows:

(1) At zero initial conditions, the encoder Φ_j and decoder Ψ_{ji} meet the condition $\Phi_j = \Psi_{ji}$, which can ensure that both sender and receiver sides have the same estimate of each sender's state.

(2) The term $\delta_j(t) - \xi_j(t_k^j)$ is quantified rather than the state $\delta_j(t)$, which can save communication bits and enhance the communication robustness.

(3) The aperiodic event-triggered state $\delta_j(t_k^j)$ is used to construct the encoder-decoder pair. Compared with the periodic event-triggered, the aperiodic event-triggered only needs to store the key information when triggering the event conditions rather than each period, and it can save memory for the device equipped with each agent.

Remark 2. In the process of designing the dynamic nonuniform encoder-decoder pair, a scaling function $g(t)$ that has a monotonously decreasing characteristic and attenuates to 0 as t approaches ∞ is introduced. It enables the quantizer to be continuously stimulated to strengthen the information

interaction between agents. At the same time, $g(t)$ should be large enough to ensure that the quantizer is never saturated.

Remark 3. Uniform quantization is also realized by encoder-decoder pairs and an integral way to quantize information in a uniform and fixed interval, which allows both large and small signals to have the same signal-to-noise ratio [34]. However, nonuniform quantization is to quantify the signal in the nonuniform and unfixed interval through its unique compression and spread characteristics, which can improve the signal-to-noise ratio of small signals and thus interpret the information contained in small signals, especially when the event-triggered control strategy is adopted, the event will not be triggered when the signal is small, so it will be ignored, which leads to a large final error of the system. Nonuniform quantization can effectively avoid this problem and improve the consistency of the system.

Then, substituting the protocol (6)–(9) into system (3) and noting (10), we obtain the following closed-loop system:

$$\begin{aligned} x_i(t+1) &= x_i(t) + hu_i(t) \\ &= x_i(t) + h\left[\sum_{j\in N_i^+}^{N} a_{ij}\hat{x}_j\left(t_{kj}^j\right) - \sum_{j\in N_i^-}^{N} a_{ij}\hat{x}_i\left(t_{ki}^i\right)\right] \\ &= x_i(t) + h\left[\sum_{j\in N_i^+}^{N} a_{ij}\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_i\left(t_{ki}^i\right)\right)\right]. \end{aligned} \tag{11}$$

Assume that the following conditions exist:

$$\begin{aligned} ha_{ii} &= 1 - \sum_{j=1, j\neq i}^{N} ha_{ij}, \\ \omega_{ij} &= ha_{ij}, \\ \hat{e}_i(t) &= \hat{x}_i\left(t_{ki}^i\right) - x_i(t). \end{aligned} \tag{12}$$

System (11) can be simplified as

$$\begin{aligned} x_i(t+1) &= x_i(t) + h\sum_{j=1}^{N} a_{ij}\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_i\left(t_{ki}^i\right) \\ &= -\hat{e}_i(t) + \sum_{j=1}^{N} w_{ij}\hat{x}_j\left(t_{kj}^j\right). \end{aligned} \tag{13}$$

Denote

$$\begin{aligned} X(t) &= \left[x_1(t), x_2(t), \ldots, x_n(t)\right]^T \in \mathbb{R}^{Nn\times 1}, \\ \hat{X}(t) &= \left[\hat{x}_1(t), \hat{x}_2(t), \ldots, \hat{x}_n(t)\right]^T \in \mathbb{R}^{Nn\times 1}, \\ \hat{e}(t) &= \left[\hat{e}_1(t), \hat{e}_2(t), \ldots, \hat{e}_n(t)\right]^T \in \mathbb{R}^{Nn\times 1}, \\ \theta(t) &= X(t) - \vartheta_N X(t) \in \mathbb{R}^{Nn\times 1}. \end{aligned} \tag{14}$$

4. Convergence Analysis

In this section, we need to prove the issues raised in this paper. Towards this end, some assumptions need to be clearly made at the start.

Assumption 1. $M_x \leq \max\|x_i(0)\|$ and $M_\theta \leq \max\|\theta_i(0)\|$, where M_x and M_θ are known nonnegative constants. For the reason that each agent is defined to accept symbol data only from its neighbors, event-triggered calculations depend only on the local information available to each agent. We propose that the following distributed sampling event-triggered condition for the ith agent satisfies

$$\hat{e}_i^2(t) \geq \sum_{j=1, j\neq i}^{N} w_{ij}\alpha_i(w_{ii} - \alpha_i)\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_i\left(t_{ki}^i\right)\right)^2. \tag{15}$$

Remark 4. Zeno behavior generally refers to event-triggered control, where the control is triggered infinitely within a limited time. If Zeno behavior occurs in the system, it will cause a certain degree of delay to the system, which will result in poor system stability and even system hardware failure. It is necessary to strictly consider avoiding Zeno behavior in a continuous-time ETC system, but not in a discrete counterpart. Because there is at least one time step between two consecutive samples in a discrete-time system, Zeno behavior does not occur.

Theorem 1. *Based on the condition of strongly connected weight-unbalanced digraph $\mathcal{G}$, combining of the system (2), encoder-decoder pair (4) and (6), and the designed control law (7) as well as the event-triggered condition (11) is achieved. The circle formation problem can be realized when the assumption is valid and the numbers of quantization levels are satisfied.*

Proof. A Lyapunov function candidate can be taken as

$$V = \sum_{i=1}^{N} x_i^2(t). \tag{16}$$

Then,

$$\begin{aligned} \Delta V &= V(t+1) - V(t) \\ &= \sum_{i=1}^{N} x_i^2(t+1) - \sum_{i=1}^{N} x_i^2(t) \\ &= \sum_{i=1}^{N}\left[\hat{e}_i^2(t) - 2\sum_{j=1}^{N}\hat{e}_i(t)w_{ij}\hat{x}_j\left(t_{kj}^j\right) + \left(\sum_{j=1}^{N} w_{ij}\hat{x}_j\left(t_{kj}^j\right)\right)^2\right] \\ &\quad - \sum_{i=1}^{N}\hat{x}_i^2(t). \end{aligned} \tag{17}$$

Since

$$\begin{aligned}
&\sum_{j=1}^{N} w_{ij}\hat{e}_i(t)\hat{x}_j\left(t_{kj}^{j}\right)\\
&= \sum_{j=1,j\neq i}^{N} w_{ij}\hat{e}_i(t)\hat{x}_j\left(t_{kj}^{j}\right)\\
&\quad + \left(1-\sum_{j=1,j\neq i}^{N} w_{ij}\right)\hat{e}_i(t)\hat{x}_i\left(t_{ki}^{i}\right)\\
&= \sum_{j=1,j\neq i}^{N} w_{ij}\hat{e}_i(t)\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_i\left(t_{ki}^{i}\right)\right)\\
&\quad - \hat{e}_i(t)\hat{x}_i\left(t_{ki}^{i}\right)\left(\sum_{j=1}^{N} w_{ij}\hat{x}_j\left(t_{kj}^{j}\right)\right)^2\\
&= \sum_{j=1}^{N} w_{ij}^2\hat{x}_j^2\left(t_{kj}^{j}\right)+2\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\hat{x}_j\left(t_{kj}^{j}\right)\hat{x}_r\left(t_{kr}^{r}\right),
\end{aligned} \tag{18}$$

we have

$$\begin{aligned}
\Delta V &= \sum_{i=1}^{N}\left[\sum_{j=1}^{N} w_{ij}^2\hat{x}_j^2\left(t_{kj}^{j}\right)+2\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\hat{x}_j\left(t_{kj}^{j}\right)\hat{x}_r\left(t_{kr}^{r}\right)\right]\\
&\quad -2\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} w_{ij}\hat{e}_i(t)\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_i\left(t_{ki}^{i}\right)\right)\\
&\quad +\sum_{i=1}^{N}\hat{e}_i^2(t)-\hat{x}_i^2(t)+2\hat{e}_i(t)\hat{x}_i\left(t_{kj}^{j}\right)\\
&= \sum_{i=1}^{N}\left[\sum_{j=1}^{N} w_{ij}^2\hat{x}_j^2\left(t_{ij}^{j}\right)+\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\left(\hat{x}_j^2\left(t_{kj}^{j}\right)+\hat{x}_r^2\left(t_{kr}^{r}\right)\right)\right]\\
&\quad +\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\left(-\hat{x}_j^2\left(t_{kj}^{j}\right)\right)\\
&\quad -\hat{x}_r^2\left(t_{kr}^{r}\right)+2\hat{x}_j\left(t_{kj}^{j}\right)\hat{x}_r\left(t_{kr}^{r}\right)\Big)\\
&\quad -2\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} w_{ij}\hat{e}_i(t)\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_i\left(t_{ki}^{i}\right)\right)\\
&\quad +\sum_{i=1}^{N}\hat{e}_i^2(t)-\hat{x}_i^2(t)+2\hat{e}_i(t)\hat{x}_i\left(t_{ki}^{i}\right).
\end{aligned} \tag{19}$$

Note that

$$\begin{aligned}
&\sum_{i=1}^{N}\left[\sum_{j=1}^{N} w_{ij}^2\hat{x}_j^2\left(t_{kj}^{j}\right)+\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\left(\hat{x}_j^2\left(t_{kj}^{j}\right)+\hat{x}_r^2\left(t_{kr}^{r}\right)\right)\right]\\
&= \sum_{i=1}^{N}\left[\sum_{j=1}^{N} w_{ij}^2\hat{x}_j^2\left(t_{kj}^{j}\right)+\sum_{j=1}^{N}\sum_{r=1,r\neq j}^{N} w_{ij}w_{ir}\hat{x}_j\left(t_{kj}^{j}\right)\right.\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{r=1}^{N} w_{ij}w_{ir}x_j^2\left(t_{kj}^{j}\right)\\
&= \sum_{i=1}^{N}\hat{x}_j^2\left(t_{kj}^{j}\right),
\end{aligned} \tag{20}$$

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\left(-\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_r\left(t_{kr}^{r}\right)+2\hat{x}_j\left(t_{kj}^{j}\right)\hat{x}_r\left(t_{kr}^{r}\right)\right)\\
&= -\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{r<j}^{N} w_{ij}w_{ir}\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_r\left(t_{kr}^{r}\right)\right)^2\\
&= -\sum_{i=1}^{N}\left[\sum_{j=1,j\neq i}^{N}\sum_{r=j,r\neq i}^{N} w_{ij}w_{ir}\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_r\left(t_{kr}^{r}\right)\right)^2\right.\\
&\quad \left.+\sum_{j=1,j\neq i}^{N} w_{ij}w_{ii}\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_r\left(t_{kr}^{r}\right)\right)\right].
\end{aligned} \tag{21}$$

Under the Young inequality [35], for any x and $y\in\mathbb{R}$ and $\tau\in\mathbb{R}$, $\tau>0$, it has the following properties $xy\leq(\tau/2)x^2+(1/2\tau)y^2$, and we have

$$\begin{aligned}
&-2\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} w_{ij}\hat{e}_i(t)\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_i\left(t_{ki}^{i}\right)\right)\\
&\leq \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} 2w_{ij}\left[\frac{\hat{e}_i^2(t)}{2\alpha_i}+\frac{\alpha_i}{2}\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_i\left(t_{ki}^{i}\right)\right)^2\right]\\
&\leq \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} w_{ij}\left[\frac{\hat{e}_i^2(t)}{\alpha_i}+\alpha_i\left(\hat{x}_j\left(t_{kj}^{j}\right)-\hat{x}_i\left(t_{ki}^{i}\right)\right)^2\right],
\end{aligned} \tag{22}$$

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\hat{e}_i^2(t)-\hat{x}_i^2(t)-2\hat{e}_i(t)\hat{x}_i\left(t_{ki}^{i}\right)\right)\\
&= \sum_{i=1}^{N}\left[\hat{e}_i^2(t)-\left(\hat{x}_i\left(t_{ki}^{i}\right)-\hat{e}_i(t)\right)^2-2\hat{e}_i(t)\hat{x}_i\left(t_{ki}^{i}\right)\right]\\
&= -\sum_{i=1}^{N}\hat{x}_i^2\left(t_{ki}^{i}\right).
\end{aligned} \tag{23}$$

Combining (20)–(23), we can obtain

$$\begin{aligned}\Delta V \le & \sum_{j=1}^{N} \hat{x}_j^2\left(t_{kj}^j\right) - \sum_{i=1}^{N}\left[\sum_{j=1,j\neq i}^{N}\sum_{k=j,r\neq i}^{N} w_{ij}w_{ir}\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_r\left(t_{kr}^r\right)\right)^2\right. \\ & \left. + \sum_{j=1,j\neq i}^{N} w_{ij}w_{ii}\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_r\left(t_{kr}^r\right)\right)\right] \\ & - \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} w_{ij}\left(\frac{\hat{e}_i^2(t)}{\alpha_i} + \alpha_i\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_i\left(t_{ki}^i\right)\right)^2\right) - \sum_{i=1}^{N}\hat{x}_i^2\left(t_{ki}^i\right) \\ \le & \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} w_{ij}\left[\frac{\hat{e}_i^2(t)}{\alpha_i} - (w_{ii}-\alpha_i)\left(\hat{x}_j\left(t_{kj}^i\right) - \hat{x}_i\left(t_{ki}^i\right)\right)^2\right] \\ \le & \sum_{i=1}^{N}\left[\frac{\hat{e}_i^2(t)}{\alpha_i} - \sum_{j=1,j\neq i}^{N} w_{ij}(w_{ii}-\alpha_i)\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_i\left(t_{kt}^i\right)\right)^2\right].\end{aligned} \tag{24}$$

Based on the event-triggered conditions (15) given above, we can obtain that the system is stable when $\hat{e}_i(t)$ meets the following conditions:

$$\hat{e}_i^2(t) < \sum_{j=1,j\neq i}^{N} w_{ij}\alpha_i(w_{ii}-\alpha_i)\left(\hat{x}_j\left(t_{kj}^j\right) - \hat{x}_i\left(t_{ki}^i\right)\right)^2. \tag{25}$$

Theorem 2. *If Assumption 1 is satisfied, given any admissible circle formation characterized by d, consider the system in (3) with the combination between dynamic encoder-decoder pair (7) and (9), designed control law (10), and scaling function $s_j(t) - s_0\xi$ over a strongly connected weight-unbalanced digraph $\mathcal{G}$. For given $h \in (0, 1/d^*)$ and $\xi \in (\rho_\lambda, 1)$, the solvability of the circle formation problem is achieved when the upper bound of quantization $\mathcal{H}_1(\xi, h)$ satisfies*

$$\mathcal{H}_1(\xi,h) = \lfloor E(\xi,h) - \frac{1}{2}\beta\rfloor + \beta, \tag{26}$$

where

$$E(\xi,h) = \frac{(1+2hd^*)}{\xi}\left(e^{((2\Omega+1)/2)\beta} - 1\right) + \frac{2h^2d^{*2}\|\nu\|_\infty\|\phi\|_\infty}{\xi(\xi-\rho_\lambda)}, \tag{27}$$

and the condition

$$s_0 > \max\left[\frac{2M_x}{(2\mathcal{H}-1)\varepsilon}, \frac{(\xi-\rho_\lambda)(\xi M_\theta + 2hd^*M_x)}{hd^*}\right], \tag{28}$$

holds simultaneously, where β represents the quantization interval and $\mathcal{H} \le (\mathcal{H}_1(\xi,h)/\beta)$. That is to say, system (3) satisfies $\lim_{t\longrightarrow\infty} y(t) = d$. Furthermore, all the quantizers will never be saturated.

Proof. Combining equations (7)–(9), we have

$$\begin{cases} X(t+1) = ((I_N - hL)\otimes I_n)X(t) - (hL\otimes I_n)\hat{e}(t), \\ \hat{X}(t+1) = g(t)\left(e^{q_t((1/g(t))(x(t+1)-\hat{x}(t)))} + 1\right) + \hat{X}(t). \end{cases} \tag{29}$$

According to $\mathcal{J}_N L = L\mathcal{J}_N = 0$, we can get

$$\begin{aligned} & X(t+1) - \hat{X}(t) \\ & = ((I_N - hL)\otimes I_n)X(t) + (hL\otimes I_n)\hat{e}(t) + g(t)\left(e^{q_t((1/g(t))(X(t+1)-\hat{X}(t)))} + 1\right) + \hat{X}(t) \\ & = ((I_N - hL)\otimes I_n)\hat{e}(t) - (hL\otimes I_n)(X(t) - J_N X(t)) \\ & = ((I_N - hL)\otimes I_n)\hat{e}(t) - (hL\otimes I_n)\theta(t). \end{aligned} \tag{30}$$

Then,

$$\begin{cases} \theta(t+1) = ((I_N - hL)\otimes I_n)\theta(t) + (hL\otimes I_n)\hat{e}(t), \\ \hat{e}(t+1) = ((I_N - hL)\otimes I_n)\hat{e}(t) + (hL\otimes I_n)\theta(t) - g(t)\left(e^{q_t((1/g(t))(X(t+1)-\hat{X}(t)))} + 1\right). \end{cases} \tag{31}$$

To make it easier to calculate, some new variables are introduced as

$$\begin{aligned} \varphi(t) &= \frac{\hat{e}(t)}{g(t)}, \\ \mu(t) &= \frac{\theta(t)}{g(t)}, \\ g(t) &= s_0 \xi^t. \end{aligned} \tag{32}$$

From $s_j(t) - s_0\xi$ and (31), we can get

$$\begin{cases} \mu(t+1) = \xi^{-1}((I_N - hL)\oplus I_n)\theta(t) + \xi^{-1}(hL \otimes I_n)\varphi(t), \\ \varphi(t+1) = \xi^{-1}((I_N - hL)\oplus I_n)\varphi(t) - (hL \otimes I_n)\mu(t) - \left(e^{q_t((1/g(t))(X(t+1) - \widetilde{X}(t)))} + 1\right). \end{cases} \tag{33}$$

For a clear expression, we summarize the procedure of proof by dividing into two cases:

Case 1: when $t = 0$ and $X(0) = 0_N$, we get

$$\begin{aligned} \|\varphi(0)\|_\infty &= \left\|\frac{\hat{e}(0)}{s_0}\right\|_\infty \le \frac{M_x}{s_0}, \\ \|((I_N - hL) \otimes I_n)\varphi(0) - (hL \otimes I_n)\mu(0)\|_\infty &= \left\|\frac{X_0}{s_0}\right\|_\infty \le \frac{M_x}{s_0} \le e^{((2\Omega+1)/2)\beta} - 1. \end{aligned} \tag{34}$$

Obviously, when $t = 0$, the quantizer is not saturated.

Case 2: when $t \ge 0$, define a nonnegative integer $r = 1, 2, \ldots, t$; then, we assume that

$$\sup_{0<t<r} \left\| \left[((I_N - hL) \otimes I_n)\varphi(t) - (hL \otimes I_n)\mu(t) - \left(\exp\left(q_t \left(\frac{1}{g(t)} ((I_N - hL) \otimes I_n)\varphi(t) - (hL \otimes I_n)u(t) \right) \right) - 1 \right) \right] \right\|_\infty \le e^{(1/2)\beta} - 1, \tag{35}$$

then

$$\sup_{0\le r} \|\varphi(t)\|_\infty \le \frac{1}{\xi}\left(e^{(1/2)\beta} - 1\right). \tag{36}$$

Let $\overline{\mu}(t) = T^*\mu(t)$, relying on Lemma 2, $\overline{\mu}(t) = [\overline{\mu}_1(t), \overline{\mu}_2(t)]^T$, $\overline{\mu}_1(t) \in \mathbb{R}^{1\times Nn}$, and $\overline{\mu}_2(t) \in \mathbb{R}^{(N-1)n\times Nn}$. Here, $\overline{\mu}_1(t) = 0$, which is available:

$$\begin{aligned} \overline{\mu}(t+1) &= T^*\mu(t+1) \\ &= \xi^{-1}((T^* - hMT^*) \otimes I_n)\mu(t) + \xi^{-1}(hMT^* \otimes I_n)\varphi(t). \end{aligned} \tag{37}$$

Let $\phi = [\phi_2, \ldots, \phi_N] \in \mathbb{R}^{(N-1)n\times Nn}$ and $v = [v_2, \ldots, v_N] \in \mathbb{R}^{Nn\times(N-1)n}$. Define $t = r + 1$, $\overline{\mu}_2(t) = \phi\mu(t)$, and $\mu(t) = v\overline{\mu}_2(t)$, and formula (40) can be organized as

$$\mu(r+1) = v\rho^{r+1}(\xi, h)\phi\overline{\mu}_2(0) - v\rho^r(\xi, h)\xi^{-1}(h\phi L \otimes I_n)\varphi(0) - v\sum_{m=0}^{r-1} \rho^m(\xi, h)\xi^{-1}(h\phi L \otimes I_n)\varphi(r-m), \tag{38}$$

where

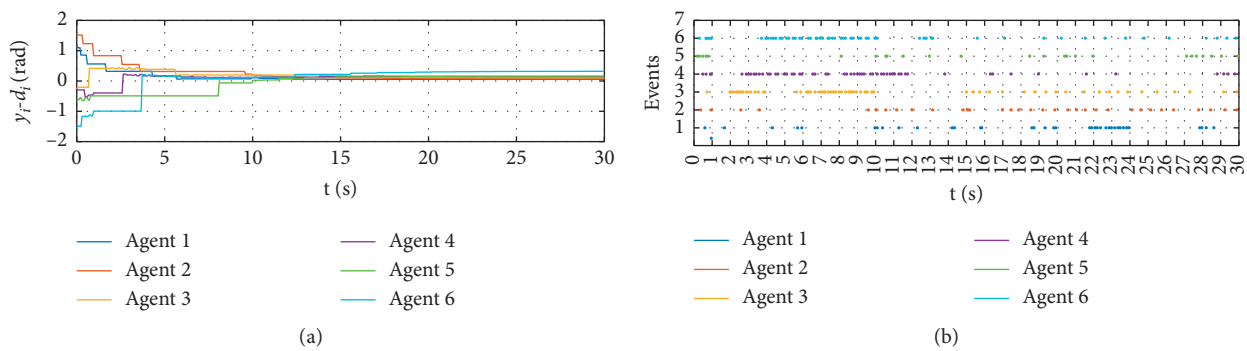


Figure 3: Simulation results of the uniform quantification algorithm in [9]. (a) Represents the evolution of the difference between current angular distance and the desired one between each pair of neighboring agents. (b) Represents the event sequences of $N = 6$ agents.

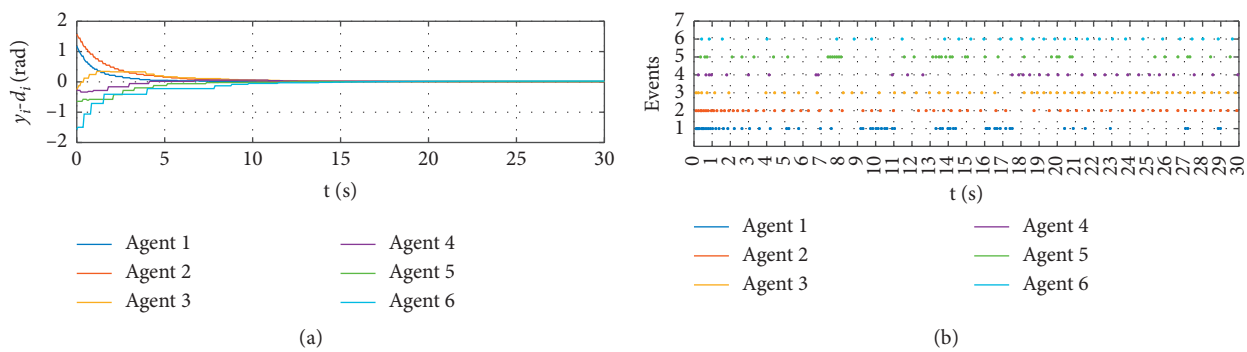


Figure 4: Simulation results of the nonuniform quantification algorithm. (a) Represents the evolution of the difference between current angular distance and the desired one between each pair of neighboring agents. (b) Represents the event sequences of $N = 6$ agents.

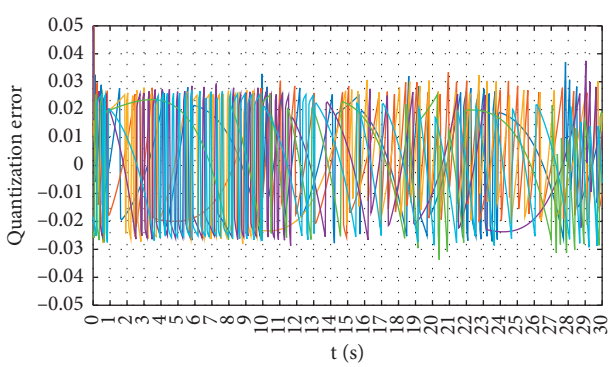


Figure 5: Uniform quantization error $e_{qi}(t)\,(i = 1, 2, \ldots, 6)$.

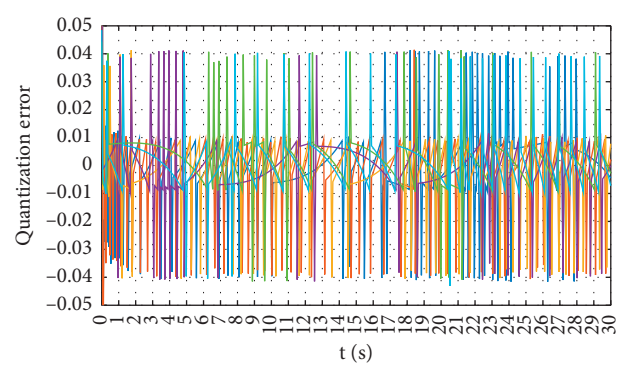


Figure 6: Nonuniform quantization error $e_{qi}(t)\,(i = 1, 2, \ldots, 6)$.

$$\left\|\nu\rho^{r+1}(\xi, h)\phi\overline{\mu}_2(0)\right\|_\infty \le \|\nu\|_\infty \left\|\rho^{r+1}(\xi, h)\right\|_\infty \|\phi\|_\infty \left\|\overline{\mu}_2(0)\right\|_\infty \le \frac{\|\nu\|_\infty \|\phi\|_\infty M_\theta}{s_0}\left(\frac{\rho_\lambda}{\xi}\right)^r. \tag{39}$$

Based on $\|L\|_\infty = d^*$, $\xi \in (\rho_\lambda, 1)$, and we have

$$\begin{aligned}&\left\|v\rho^{r}(\xi,h)\xi^{-1}(h\phi L\otimes I_n)\varphi(0)\right\|_{\infty}\\&\le\|v\|_{\infty}\left\|\rho^{r}(\xi,h)\right\|_{\infty}\xi^{-1}h\|\phi\|_{\infty}\left\|L\otimes I_n\right\|_{\infty}\|\varphi(0)\|_{\infty}\\&\le\frac{2hd^{*}\|v\|_{\infty}\|\phi\|_{\infty}M_{\infty}}{s_0\xi}\left(\frac{\rho_\lambda}{\xi}\right)^{r}.\end{aligned}\tag{40}$$

Next,

$$\left\|v\sum_{m=0}^{r-1}\rho^{m}(\xi,h)\xi^{-1}(h\phi L\otimes I_n)\varphi(r-m)\right\|_{\infty}\le\frac{2hd^{*}\|u\|_{\infty}\|\phi\|_{\infty}}{\xi(\xi-\rho_\lambda)}\left(1-\left(\frac{\rho_\lambda}{\xi}\right)^{r}\right).\tag{41}$$

Depending on Theorem 2 and formulas (39)–(41), we have

$$\begin{aligned}\|\mu(r+1)\|_{\infty}&\le\frac{\|v\|_{\infty}\|\phi\|_{\infty}M_{\theta}}{s_0}\left(\frac{\rho_\lambda}{\xi}\right)^{r}+\frac{2hd^{*}\|v\|_{\infty}\|\phi\|_{\infty}M_{x}}{s_0\xi}\left(\frac{\rho_\lambda}{\xi}\right)^{r}+\frac{2hd^{*}\|v\|_{\infty}\|\phi\|_{\infty}}{\xi(\xi-\rho_\lambda)}\left(1-\left(\frac{\rho_\lambda}{\xi}\right)^{r}\right)\\&\le\max\left(\frac{\xi\|\mathbf{\mu}\|_{\infty}\|\phi\|_{\infty}M_{\theta}+2hd^{*}\|v\|_{\infty}\|\phi\|_{\infty}M_{x}}{s_0\xi}\xi^{*}+\frac{hd^{*}\|v\|_{\infty}\|\phi\|_{\infty}}{\xi(\xi-\rho_\lambda)}(1-\xi^{*})\right).\end{aligned}\tag{42}$$

Then,

$$\begin{aligned}&\left\|((I_N-hL)\otimes I_n)\varphi(r-1)-(hL\otimes I_n)\mu(r-1)\right\|_{\infty}\\&\le\left\|(I_N-hL)\otimes I_n)\varphi(r-1)\right\|_{\infty}+\left\|(hL\otimes I_n)\mu(r-1)\right\|_{\infty}\\&\le\frac{(1+2hd^{*})}{\xi}\left(e^{((2\Omega+1)/2)\beta}-1\right)+2hd^{*}\max\left[\frac{\xi\|v\|_{\infty}\left\|\right\|_{\infty}\left\|\right\|_{\infty}M_{\theta}+2hd^{*}\|v\|_{\infty}\left\|\right\|_{\infty}\left\|M\right\|_{\infty}M_{\infty}}{s_0\xi}\xi^{*}+\frac{hd^{*}\|v\|_{\infty}\|\phi\|_{\infty}}{\xi(\xi-\rho_\lambda)}(1-\xi^{*})\right]\\&=\frac{(1+2hd^{*})}{\xi}\left(e^{((2\Omega+1)/2)\beta}-1\right)+\frac{2h^{2}d^{*2}\|v\|_{\infty}\|\phi\|_{\infty}}{\xi(\xi-\rho_\lambda)}\\&=E(\xi,h)\le e^{E(\xi,h)}-1\le e^{\lfloor E(\xi,h)-(1/2)\beta\rfloor+(3/2)\beta}-1\\&=e^{\mathscr{H}_1(\xi,h)+(1/2)\beta}-1=e^{((2\Omega+1)/2)\beta}-1.\end{aligned}\tag{43}$$

That is to say, all the quantizers are not saturated when the above condition is satisfied by quantizer level number $(2\Omega-1)$.

5. Numerical Examples

In this section, a multiagent system composed of $N=6$ agents is used to simulate and verify the superiority of the nonuniform quantization algorithm. Under the condition that the initial conditions of each agent in the system satisfy (1), $h=0.06\,\mathrm{s}$ is selected as the sampling period of the system. Under the condition that the proposed algorithm can be realized, the expected angular distance between agents is set to $d=[(\pi/8),(\pi/6),(\pi/4),(\pi/3),(3\pi/8),(3\pi/4)]$. We set the quantization interval correlation amount $\beta=0.2$, $s_0=10$, and $\xi=0.98288$. The simulation results are shown in the figure.

Comparing the simulation results of uniform quantization algorithm (Figure 3) and nonuniform quantization algorithm (Figure 4) can get some conclusions. Due to the characteristics of nonuniform quantization, a better quantized signal-to-noise ratio can be obtained when processing small signals. The error of nonuniform quantization will be more smaller. In other words, nonuniform quantization can effectively improve the system accuracy. It is obvious that the nonuniform quantization tracking effect is better and the

final error is significantly reduced when the event-triggered mechanism is considered. The reason why nonuniform quantization triggers less than uniform quantization is that nonuniform quantization can improve the SNR of small signals and trigger events to update the control protocol even when the signal is small. However, uniform quantization tends to ignore this signal, which leads to the accumulation of errors and ultimately increases the frequency of events. In summary, the nonuniform quantization algorithm proposed in this paper can better improve the system accuracy and can improve the system performance under the same conditions.

As shown in Figures 5 and 6, the quantization error of each agent is $e_{qi}(t) = s_j(t) - ((1/g(t))(x_j(t) - t\xi_j n(t_k^j)))$. In the nonuniform quantization algorithm, $s_j(t)$ is log-scaled, so the quantization error exhibits the phenomenon in Figure 6. It is worth noting that the quantization error of each agent's quantizer is less than its upper limit and saturation will never occur.

6. Conclusions

This paper mainly explores the existing uniform quantization method further and designs a new nonuniform quantization algorithm combined with the event-triggered mechanism. First of all, in order to improve the information exchange between agents, we propose a scheme based on nonuniform dynamic codec, which effectively solves the problem of small signal loss during information exchange between agents. Then, a distributed control algorithm based on the coordinated control of the event-triggered mechanism and nonuniform encoder is given to reduce system energy consumption. In addition, all designed quantizers do not appear to be saturated. Finally, numerical simulation results verify the effectiveness of the algorithm. Future work will focus on solving practical problems, such as the rapid convergence of multiagents under limited input and the consistency of general nonlinear systems.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by grants from the National Natural Science Foundation of China (nos. 61963006, 61563005, 61563006, 61973007, and 61633002), Natural Science Foundation of Guangxi Province of China (Grant nos. 2018GXNSFAA050029 and 2018GXNSFAA294085), Innovation Project of Guangxi University of Science and Technology Graduate Education (no. GKYC202006), and Project of Guangxi Key Laboratory of Automobile Components and Vehicle Technology (no. 2017GKLACVTZZ02).

References

[1] B. D. O. Anderson, Z. Sun, T. Sugie, S.-I. Azuma, and K. Sakurama, "Formation shape control with distance and area constraints," *IFAC Journal of Systems and Control*, vol. 1, pp. 2–12, 2017.

[2] K.-K. Oh, M.-C. Park, and H.-S. Ahn, "A survey of multi-agent formation control," *Automatica*, vol. 53, pp. 424–440, 2015.

[3] X. Yang, L. Liao, Q. Yang, B. Sun, and J. Xi, "Limited-energy output formation for multiagent systems with intermittent interactions," *Journal of the Franklin Institute*, 2021.

[4] J. Wen, C. Wang, P. Xu, and G. Xie, "Decentralized event-triggered circle formation control for multiagent systems via synchronous periodic event detection," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 3, pp. 910–925, 2020.

[5] C. Wang and G. Xie, "Limit-cycle-based decoupled design of circle formation control with collision avoidance for anonymous agents in a plane," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6560–6567, 2017.

[6] P. Xu, J. Wen, C. Wang, and G. Xie, "Distributed circle formation control over directed networks with communication constraints," *IFAC-PapersOnLine*, vol. 52, no. 3, pp. 108–113, 2019.

[7] L. Zhang and L. Sun, "Multi-objective service restoration for blackout of distribution system with distributed generators based on multi-agent ga," *Energy Procedia*, vol. 12, pp. 253–262, 2011.

[8] W. Liu, S. Zhou, S. Yan, and Q. Wu, "LQR_based consensus algorithms of muti-agent systems with a prescribed convergence speed," in *Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference*, pp. 868–873, Yantai, China, August 2014.

[9] J. Wen, P. Xu, C. Wang, G. Xie, and Y. Gao, "Distributed event-triggered circle formation control for multi-agent systems with limited communication bandwidth," *Neurocomputing*, vol. 358, pp. 211–221, 2019.

[10] A. Kashyap, T. Basar, and R. Srikant, "Consensus with quantized information updates," in *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 2728–2733, San Diego, CA, USA, December 2006.

[11] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus on networks with quantized communication," *International Journal of Robust and Nonlinear Control*, vol. 19, no. 16, pp. 1787–1816, 2009.

[12] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Efficient quantized techniques for consensus algorithms," in *Proceedings of the NeCST workshop*, pp. 1–8, Nancy, France, June 2007.

[13] H. Li, G. Chen, T. Huang, and Z. Dong, "High-performance consensus control in networked systems with limited bandwidth communication and time-varying directed topologies," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1043–1054, 2016.

[14] T. Li, M. Fu, L. Xie, and J.-F. Zhang, "Distributed consensus with limited communication data rate," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 279–292, 2010.

[15] T.-F. Ding, M.-F. Ge, C.-H. Xiong, and J. H. Park, "Bipartite consensus for networked robotic systems with quantized-data interactions," *Information Sciences*, vol. 511, pp. 229–242, 2020.

[16] Z. Zeng, X. Wang, Z. Zheng, and L. Zhao, "Edge agreement of second-order multi-agent system with dynamic quantization

via the directed edge laplacian," *Nonlinear Analysis: Hybrid Systems*, vol. 23, pp. 1–10, 2017.

[17] J. Li, D. W. C. Ho, and J. Li, "Adaptive consensus of multi-agent systems under quantized measurements via the edge laplacian," *Automatica*, vol. 92, pp. 217–224, 2018.

[18] Z.-M. Li, X.-H. Chang, K. Mathiyalagan, and J. Xiong, "Robust energy-to-peak filtering for discrete-time nonlinear systems with measurement quantization," *Signal Processing*, vol. 139, pp. 102–109, 2017.

[19] C.-E. Ren, L. Chen, C. L. P. Chen, and T. Du, "Quantized consensus control for second-order multi-agent systems with nonlinear dynamics," *Neurocomputing*, vol. 175, pp. 529–537, 2016.

[20] Q. Zheng, S. Xu, and Z. Zhang, "Asynchronous nonfragile H_∞ filtering for discrete-time nonlinear switched systems with quantization," *Nonlinear Analysis: Hybrid Systems*, vol. 37, Article ID 100911, 2020.

[21] Z. Xu, C. Li, and Y. Han, "Leader-following fixed-time quantized consensus of multi-agent systems via impulsive control," *Journal of the Franklin Institute*, vol. 356, no. 1, pp. 441–456, 2019.

[22] G. Guo, L. Ding, and Q.-L. Han, "A distributed event-triggered transmission strategy for sampled-data consensus of multi-agent systems," *Automatica*, vol. 50, no. 5, pp. 1489–1496, 2014.

[23] L. Hetel, C. Fiter, H. Omran et al., "Recent developments on the stability of systems with aperiodic sampling: an overview," *Automatica*, vol. 76, pp. 309–335, 2017.

[24] T. Li, H. Zhao, and Y. Chang, "A novel event-triggered communication strategy for second-order multiagent systems," *ISA Transactions*, vol. 97, pp. 93–101, 2020.

[25] T. Li, Z. Li, S. Fei, and Z. Ding, "Second-order event-triggered adaptive containment control for a class of multi-agent systems," *ISA Transactions*, vol. 96, pp. 132–142, 2020.

[26] G. Liu, Y. Pan, H.-K. Lam, and H. Liang, "Event-triggered fuzzy adaptive quantized control for nonlinear multi-agent systems in nonaffine pure-feedback form," *Fuzzy Sets and Systems*, vol. 416, pp. 27–46, 2021.

[27] Y. Xie and Z. Lin, "Event-triggered global stabilization of general linear systems with bounded controls," *Automatica*, vol. 107, pp. 241–254, 2019.

[28] X. Dong and G. Hu, "Time-varying formation control for general linear multi-agent systems with switching directed topologies," *Automatica*, vol. 73, pp. 47–55, 2016.

[29] X. Yao, Y. Lian, and J. H. Park, "Disturbance-observer-based event-triggered control for semi-markovian jump nonlinear systems," *Applied Mathematics and Computation*, vol. 363, Article ID 124597, 2019.

[30] P. Wang, G.-H. Yang, and Y. Pan, "Event-triggered reliable dissipative filtering for nonlinear networked control systems," *Neurocomputing*, vol. 360, pp. 120–130, 2019.

[31] J. Lin, L. Li, M. Bi, J. Li, M. Hu, and W. Hu, "A study on performance improvement of IMDD-UFMC with modified k-means non-uniform quantization," *Optics Communications*, vol. 476, Article ID 126324, 2020.

[32] A. G. Dimitrov, J. P. Miller, Z. Aldworth, and T. Gedeon, "Non-uniform quantization of neural spike sequences through an information distortion measure," *Neurocomputing*, vol. 38–40, pp. 175–181, 2001.

[33] T. Kitayabu, H. Ishikawa, M. Hagiwara, and H. Shirai, "Effect of input-signal statistical property in delta-sigma modulator with non-uniform quantization," in *Proceedings of the 2012 IEEE Radio and Wireless Symposium*, pp. 183–186, Santa Clara, CA, USA, January 2012.

[34] S. Liu, L. Xie, and D. E. Quevedo, "Event-triggered quantized communication-based distributed convex optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 167–178, 2016.

[35] X.-L. Hu, "An extension of young's inequality and its application," *Applied Mathematics and Computation*, vol. 219, no. 12, pp. 6393–6399, 2013.

Hindawi
Complexity
Volume 2021, Article ID 6635242, 11 pages
https://doi.org/10.1155/2021/6635242

Research Article

I-GANs for Infrared Image Generation

Bing Li,[1] **Yong Xian**,[1] **Juan Su**,[1] **Da Q. Zhang**,[1] **and Wei L. Guo**[2]

[1]*Xi'an High-Tech Research Institute, Xi'an 710025, China*
[2]*Xi'an Satellite Control Center, Xi'an 710043, China*

Correspondence should be addressed to Bing Li; libingbenyi@163.com

Received 11 December 2020; Revised 24 February 2021; Accepted 2 March 2021; Published 23 March 2021

Academic Editor: Ning Cai

The making of infrared templates is of great significance for improving the accuracy and precision of infrared imaging guidance. However, collecting infrared images from fields is difficult, of high cost, and time-consuming. In order to address this problem, an infrared image generation method, infrared generative adversarial networks (I-GANs), based on conditional generative adversarial networks (CGAN) architecture is proposed. In I-GANs, visible images instead of random noise are used as the inputs, and the D-LinkNet network is also utilized to build the generative model, enabling improved learning of rich image textures and identification of dependencies between images. Moreover, the PatchGAN architecture is employed to build a discriminant model to process the high-frequency components of the images effectively and reduce the amount of calculation required. In addition, batch normalization is used to optimize the training process, and thereby, the instability and mode collapse of the generated adversarial network training can be alleviated. Finally, experimental verification is conducted on the produced infrared/visible light dataset (IVFG). The experimental results reveal that high-quality and reliable infrared data are generated by the proposed I-GANs.

1. Introduction

Due to the limitations of the application background and support capabilities, the template used in infrared imaging guidance is usually a visible image, while the real-time image itself is infrared. The imaging principles of infrared and visible are different, which results in a large feature disparity between the infrared image and the visible image. As a result, the difficulty of scene matching in infrared imaging guidance increases. If the infrared image is used as the reference image for matching, the matching accuracy and precision can be improved. Moreover, the matching difficulty can be reduced. However, relying solely on an off-site field to obtain infrared reference maps is time-consuming, and it is also arduous to obtain infrared images of targets in complex environments and harsh climates. Compared with testing in the field, the use of infrared image simulation technology to generate the infrared characteristics of the scene in the environment of interest can not only effectively reduce the cost of acquiring infrared data but also generate a large amount of infrared data that is difficult to obtain in the field under a variety of natural environments and scene conditions. In this way, the generated infrared data can be used in the fields of aviation, aerospace, navigation, meteorology, geology, and agriculture by providing basic and reliable data for detection [1], classification [2], positioning, identification, tracking purposes, etc. Therefore, generating infrared reference maps through infrared image simulation technology is highly significant for military and civilian applications.

In recent years, with the continuous improvement of computer performance [3, 4] and the rapid development of deep learning theory, many new neural network-based generation models have been proposed. Among these, generative adversarial networks (GANs) [5] have demonstrated a unique capacity to meet research and application needs in many fields and have accordingly become one of the most critical research hotspots in the field of artificial intelligence [6, 7]. Antipov et al. used conditional generative adversarial networks (CGAN) to generate face images [8]. Through applying GANs to the field of face turning (which refers to a technique for synthetizing high definition (HD) frontal face images from a single-sided face image), Huang

and Tran proposed two-pathway generative adversarial networks (TP-GANs) [9] and disentangled representation learning-generative adversarial networks (DR-GANs) [10], respectively. The Markov-based Markovian generative adversarial networks (MGANs) [11] have the same synthesis speed as texture network [12] in generating image textures. Isola et al. demonstrated that pix2pix approach could realize the conversion of black and white to colour, satellite to map, semantic to street view, and edge to photo [13]. Moreover, the image textures and backgrounds generated by BigGAN [14] are more realistic, although the computation complexity of this approach is high. Subsequently, in order to improve the learning performance by taking advantage of the improvement in image generation quality, Donahue and Simonyan proposed BigBiGAN based on the BigGAN model, extending this approach to the image learning context by adding encoders and modifying the identifier [15]. Image super resolution generative adversarial networks (SRGAN) used residual networks (ResNets) and VGG networks [16] as generators and discriminators, respectively, to attain a better texture detail learning effect [17]. In order to solve the lifelong learning problem of the generative model, Zhai et al. presented the Lifelong GAN [18]. He et al. proposed a dual learning mechanism in which the neural machine translation system can automatically learn from unlabeled data through a dual learning game [19]. Following the idea of dual learning, Yi et al. used the DualGAN model of dual learning to achieve cross-domain image generation [20], and Zhu et al. introduced cycle consistency into GANs to extend the image-to-image conversion work [21]. Choi et al. first proposed a novel and scalable method, StarGAN, which is capable of converting images to images translation for multiple domains from using only one model [22]. Beginning with RGB images from Kinect and curve normal maps, Karras et al. proposed a generative adversarial model called Style-GAN, which takes normal surface as the basis for the generative adversarial networks used to generate images [23]. Based on Style-GAN model, Yang and Lim proposed a framework capable of generating face images that fall into the same distribution as that of a given one-shot example [24]. Besides, Richardson et al. presented a generic image-to-image translation framework Pixel2Style2Pixel (pSp). The pSp framework is based on a new encoder network that directly generates a series of style vectors which are fed into a pretrained Style-GAN generator, forming the extended $W+$ latent space [25]. Chen et al. presented a domain adaptive image-to-image conversion (DAI2I) framework, which is suitable for the I2I model of samples outside the domain [26].

At present, the majority of GANs-based image generation researches have applied GANs to face synthesis, texture generation, sketch-to-photo applications, transforming visible images to night vision images, etc. However, few studies have been published on the use of GANs models in the field of infrared image simulation. In view of the high cost, comparatively small quantities, and the relative difficulty of obtaining infrared data in the off-site field, this paper proposes an infrared image generation method based on generative adversarial networks (infrared generative adversarial networks, or I-GANs), which is capable of simulating and generating infrared images on the basis of visible images. Besides, the generated infrared images can be used to create infrared reference maps, which provide reliable infrared data and expand infrared databases. Based on CGAN architecture, the I-GANs algorithm employs the D-LinkNet network to build the generation network, using visible images and infrared simulation samples as the inputs and outputs, respectively. Then, the real target sample and the generated simulation sample are utilized to train the PatchGAN-based discrimination network, which outputs the probability of a generated sample belonging to the corresponding category. Through alternating iterative training of the generation network and the discriminant network, the final generated infrared simulation samples have essentially the same data distribution as the real samples.

The novelty of the work in this paper can be summarized as follows: (1) innovation of research background. We present a novel generation adversarial network algorithm (i.e., I-GANs) with infrared image simulation as the research background, which has a reliable reference value for the subsequent infrared image generation researches; (2) we introduce a D-LinkNet module into conditional GANs. Armed with D-LinkNet, the generator can better preserve the spatial details of the images and achieve multiscale feature fusion.

2. Related Work

Generative adversarial networks (GANs) were first proposed by Goodfellow et al. at the 28th International Conference on Neural Information Processing Systems in 2014 [5]. The generative adversarial networks are a new generative model developed on the basis of a deep generative model. The significant difference between this model and other generative models lies in its use of an adversarial approach. It first learns the difference between the generated sample and the training sample through the discriminator and then guides the generator to reduce this difference rather than to directly target the differences between the data distribution and the model distribution. At present, GANs are one of the most significant research hotspots in the field of artificial intelligence.

2.1. Generative Adversarial Networks. The key concept behind GANs involves setting up a zero-sum game to achieve learning through the confrontation between two players. In the zero-sum game, one player acts as the generator while the other acts as the discriminator. The generator's main task is to generate samples that appear as identical as possible to the training samples, thereby deceiving the other player. For the discriminator, the goal is to accurately determine whether the input samples belong to the set of real training samples. In GANs, the generation network and the adversarial network are often thought of as analogous to a counterfeiter of banknotes and a detector of forged currency. The GANs training process thus resembles the following

procedure: the counterfeiter continues to increase the sophistication of their forged banknotes in order to produce counterfeit banknotes that are as identical as possible to real currency, in the hope that the forgery detector will fail to spot the forgery; for their part, the money detector constantly improves their ability to identify counterfeit banknotes. As the GANs training process continues, both the counterfeiter's ability to manufacture convincing counterfeit notes and the money detector's ability to identify forgeries will continually increase [20].

The GANs consist of two networks, a generative network (generator G) and an adversarial network (discriminator D), which corresponds to the generative and the adversarial model, respectively. The basic framework of the original generative adversarial networks is illustrated in Figure 1.

In the original GANs, the value function $V(G, D)$ [5, 27] is defined as follows:

$$V(G, D) = \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z})))], \tag{1}$$

where $x \sim p_{\text{data}}$ represents the distribution of x taken from real data, $z \sim p_z$ indicates that the random noise z comes from simulated data (such as a Gaussian noise distribution), $\mathbb{E}(\cdot)$ is the expected value, and G tries to minimize this objective while an adversarial D tries to maximize it; i.e., $G* = \arg\min_G \max_D V(G, D)$.

2.2. Conditional Generative Adversarial Networks. With the goal of remedying the original GANs' inability to generate pictures with specific attributes, Mirza and Osindero proposed the conditional generative adversarial networks (CGAN) [28]. The core concept of the CGAN involves integrating condition information y into the generator and discriminator. Condition y can be any label information, such as the facial expressions of face images and image categories. The CGAN network structure is presented in Figure 2.

The objective of a CGAN can be expressed as follows:

$$\ell_{\text{CGAN}}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]. \tag{2}$$

3. Methods

3.1. Objective. In this section, based on the CGAN framework, we proposed the I-GANs algorithm which uses images as input rather than random noise. In order to make better use of the structural information contained in the input image, the L1 objective function is introduced into the loss function as follows:

$$\ell_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]. \tag{3}$$

The loss function of I-GANs is then finally defined as follows:

$$G^* = \arg\min_G \max_D \ell_{\text{CGAN}}(G, D) + \lambda\mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]. \tag{4}$$

3.2. Generative Networks. The network of the common encoder-decoder structure operates by first downsampling to a low dimension and then upsampling to the original resolution. By contrast, D-LinkNet [29], which uses LinkNet as the basic framework and then introduces a residual network [30], has the advantages of employing skip connection (used to retain pixel-level detailed information at different resolutions), residual blocks, and encoder-decoder systems, thus increasing the receptive fields of the network, retaining the spatial detail information of the image, and realizing multiscale feature fusion.

In the proposed I-GANs algorithm, D-LinkNet is used to construct a generative network. More specifically, in this article, D-LinkNet is designed to receive images of size 256×256 as input. As shown in Figure 3, D-LinkNet is composed of three parts, A, B, and C, which are the encoder part, the central part, and the decoder part, respectively. In the encoder part, ResNet34 [30], which is trained on the ImageNet dataset, is used as the encoder. In the central part, dilated convolution with shortcut is added to enhance the network's recognition ability, expand the receptive field, and fuse multiscale information. Finally, the decoder part uses transposed convolution [31] layers to conduct upsampling, restoring the resolution of the feature map from 8×8 to 256×256.

The center dilation part of this D-LinkNet can be unrolled into the structure illustrated in Figure 4. From top to bottom in the figure, if the dilation rates of the stacked dilated convolution layers are 2, 1, and 0, respectively, then the corresponding numbers of receptive fields are 7, 3, and 1; finally, the results of each branch are added together, and the characteristics of the fusion are obtained. Since the encoder part of the D-LinkNet contains five downsampling layers, while the size of the input data is 256×256, the encoder output feature map will be of the size 8×8. In this case, D-LinkNet uses dilated convolution layers with a dilation rate of 1 and 2 in the center part. Thus, the feature points on the last center layer will yield 7×7 points on the first center feature map, covering the main part of the first center feature map.

3.3. Adversarial Networks. In the I-GANs, the adversarial network is constructed using the convectional PatchGAN classifier. The main idea behind PatchGAN is as follows: since GANs are used to build high-frequency information, there is no need to input the entire image into the discriminator; instead, the discriminator can make true or false judgments about each block of the image, which penalises the structure only on the scale of the image block. Therefore, the I-GANs' discriminator only needs to pay attention to the local structure of the image (which can effectively reduce the number of parameters in training), model the high-frequency components of the image, and rely on the L1 items to ensure the accuracy at low frequencies.

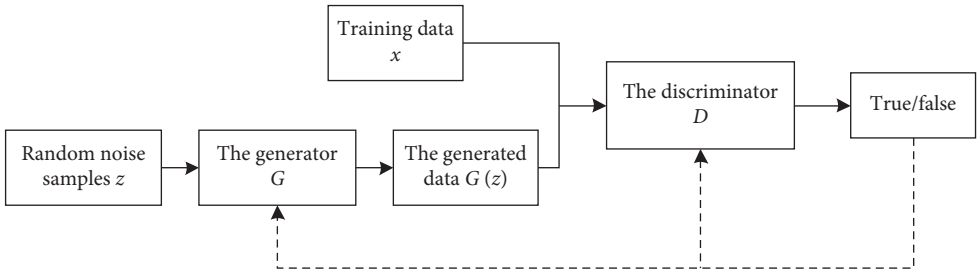


Figure 1: The basic framework of the original GANs. The GANs consist of two networks: a generative network (generator G) and an adversarial network (discriminator D).

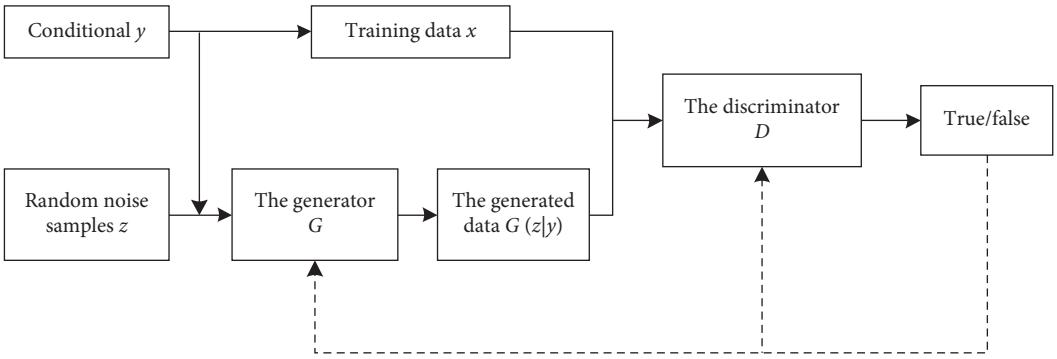


Figure 2: The basic framework of CGAN. CGAN has an additional condition y.

4. Results and Discussion

4.1. Datasets. UAV is equipped with a thermal infrared camera and a visible camera (both of which are coaxially installed) to capture the desired target and scene in the designated area. In brief, the designated area is photographed using a coaxial infrared camera and a visible-light camera simultaneously. Targets in the data include buildings (with materials including steel, concrete, cement, and various types of bricks), vehicles (including trucks and buses), radar covers, power stations (e.g., thermal and hydroelectric), oil depots, highways (with materials including cement and asphalt), runways, grasslands (both real and artificial), trees, and rivers (or ponds). Scenes in the data include cities, campuses, streets, factories, residential areas, transportation hubs, and rivers. Meteorological conditions identified in the data collection include sunny, cloudy, hazy, and rainy. We name this dataset "IVFG."

4.2. Subjective Evaluation. In order to evaluate the proposed I-GANs methods, we conducted a large number of experiments on the IVFG dataset. The generation effect of infrared-generated images is evaluated by means of subjective observation and objective index verification.

Next, infrared-generated images of buildings, chimneys, and cooling towers, generated by the I-GANs algorithm, are presented in Figures 5–7. The building materials in Figure 5 include steel, concrete, cement, and various types of bricks. Through visual interpretation and subjective evaluation, it can be determined that the grey information and contour information of the infrared-generated images are closer to those of the real infrared images. In addition, the similarity between the two is higher, and the infrared generation effect is superior.

4.3. Objective Evaluation. Generally speaking, the greater the similarity of the grey characteristics between generated infrared images and those obtained in real time, the better the infrared image generation results. In order to objectively evaluate the I-GANs algorithm's effectiveness at generating infrared images, we calculate the Root Mean Square Error (RMSE) and feature similarity (Feature SIMilarity, FSIM) [32] between infrared generation-based templates (which are split off from infrared-generated results via human-computer interaction) and infrared real-time maps, respectively.

The RMSE is a measure of the degree of information change between the two images, which reflects the difference in grey values. In general, the smaller the RMSE value, the smaller the greyscale difference between the two, that is, the better the generation effect of the infrared-generated images. On the contrary, the larger the RMSE value, the worse the generation effect of the infrared-generated image. Moreover, FSIM represents an improvement of structural similarity, which not only uses phase consistency to extract rich texture, edge, and structure information, but also introduces the contrast information of the gradient amplitude to extract images, enabling the structural differences between images to be evaluated. Generally speaking, the greater the FSIM value, the higher the similarity between images (i.e., the better

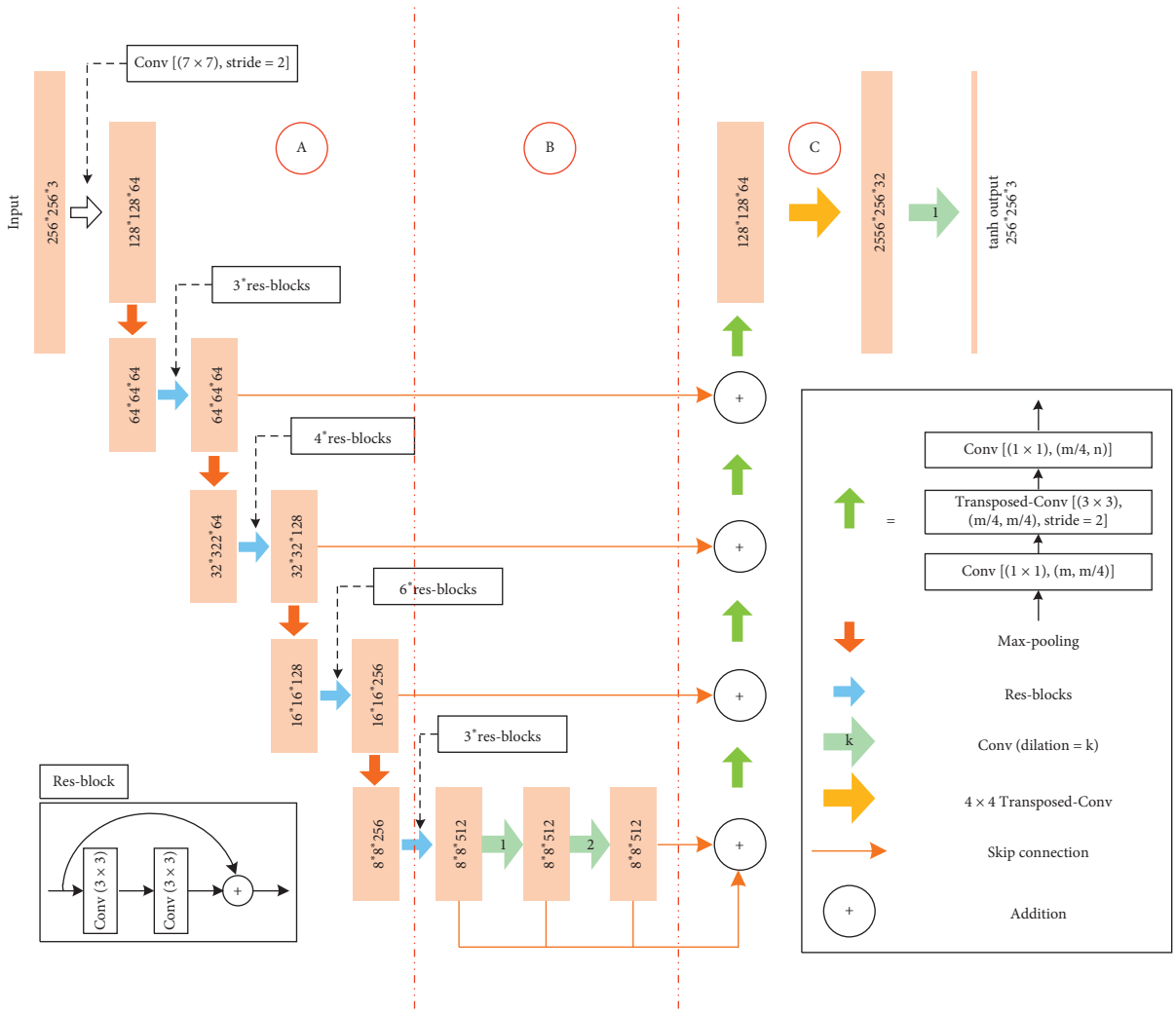


Figure 3: A diagram of D-LinkNet. Each orange rectangular block represents a multichannel features map. Part A and Part C are the encoder and the decoder of D-LinkNet, respectively. Here, D-LinkNet uses ResNet34 as encoder.

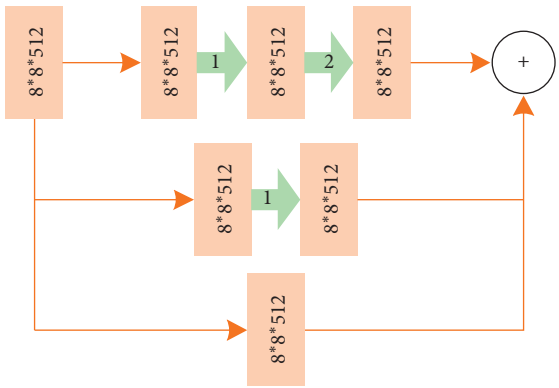


Figure 4: The unrolled structure of the center dilation part of D-LinkNet. From top to bottom, the receptive fields are 7, 3, and 1, respectively.

the infrared generation). Because the user tends to pay more attention to the infrared generation effect of the target, this paper only calculates the RMSE and FSIM between the target's infrared real-time map and the infrared generation map. The RMSE and FSIM are calculated according to the following equations:

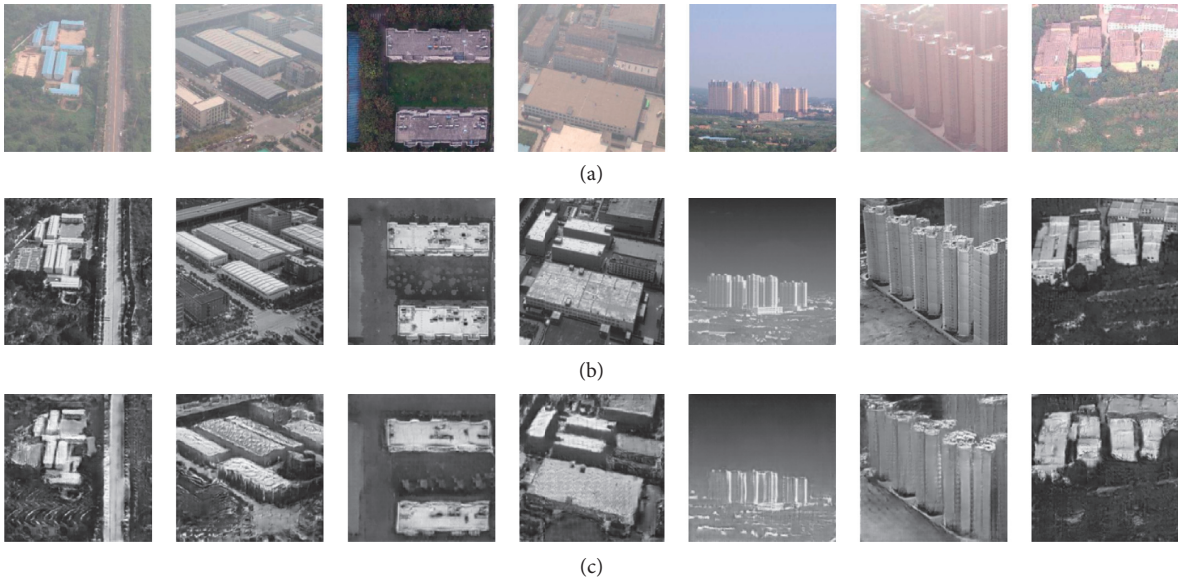


Figure 5: Examples of infrared-generated images of buildings produced on the basis of the I-GANs algorithm. (a) Visible images. (b) Real infrared images. (c) Infrared-generated images.

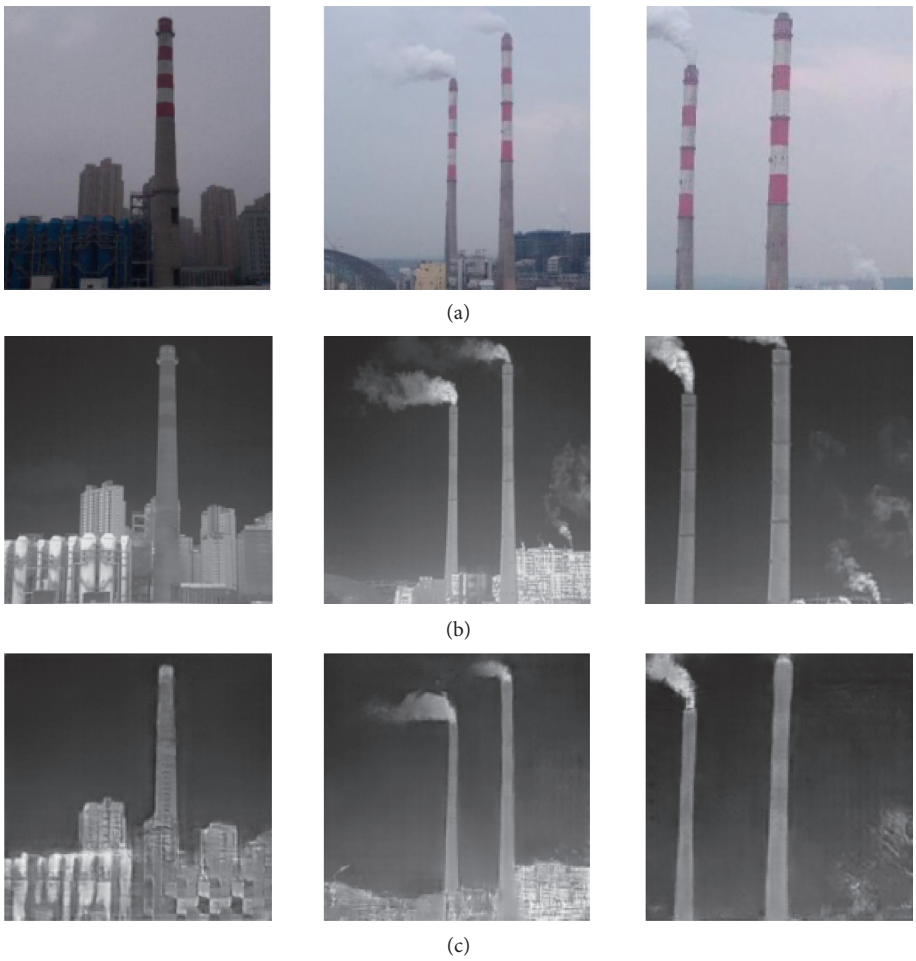


Figure 6: Examples of infrared-generated images of chimneys produced on the basis of the I-GANs algorithm. (a) Visible images. (b) Real infrared images. (c) Infrared-generated images.

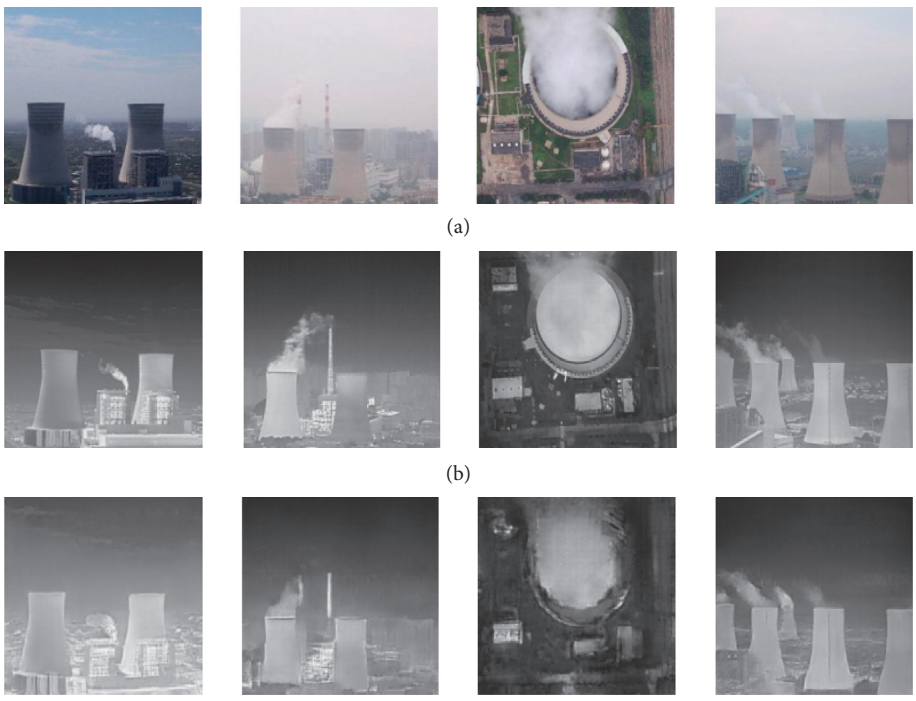


Figure 7: Examples of infrared-generated images of cooling towers produced on the basis of the I-GANs algorithm. (a) Visible images. (b) Real infrared images. (c) Infrared-generated images.

$$\mathrm{RMSE} = \left(\frac{1}{\mathrm{MN}} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(I(x,y) - S(x,y) \right)^2 \right)^{1/2},$$

$$\begin{cases} S_{PC} = \dfrac{2\mathrm{PC}_1(I) \cdot \mathrm{PC}_2(S) + T_1}{\mathrm{PC}_1^2(I) + \mathrm{PC}_2^2(S) + T_1}, \\ S_G = \dfrac{2G_1(I) \cdot G_2(S) + T_2}{G_1^2(I) + G_2^2(S) + T_2}, \\ S_L = [S_{\mathrm{PC}}]^{\alpha} \cdot [S_G]^{\beta}, \\ \mathrm{FSIM} = \dfrac{\sum S_L \cdot \max(\mathrm{PC}_1(I), \mathrm{PC}_2(S))}{\sum \max(\mathrm{PC}_1(I), \mathrm{PC}_2(S))}, \end{cases} \tag{5}$$

where I and S represent the infrared measure of the target and the infrared simulation chart, respectively. Moreover, $\mathrm{PC}_1(I)$ and $\mathrm{PC}_2(S)$ represent the phase consistency of I and S, respectively, while $G_1(I)$ and $G_2(S)$ represent the gradient amplitude of I and S, respectively.

In this paper, in order to verify the generation results, the proposed I-GANs algorithms are compared with three GANs-based algorithms, the generators of which are U-Net256, ResNet9, and ResNet34, respectively. Among them, the algorithm with U-Net256 as generator is the classic Pix2pix algorithm [13], and the following are all described with "Pix2pix". Besides, in the following, the GANs-based algorithms construct generators with ResNet9 and ResNet34, respectively, are called "Resnet9" and "Resnet34," respectively. The network structure of the four algorithms participating in the experimental comparison is shown in Table 1.

There are 1374 sets of infrared/visible light images (1374 infrared images and 1374 visible images) in the dataset involved in the experiment in this paper. The training samples and test samples are constructed according to the ratio of 1070 : 304. For the RMSE index, smaller value is superior; among the FSIM index, larger value is superior. We make statistics on the number of superior and inferior values of the actual values of the image quality evaluation indexes and define the statistical result as the ratio of superiority and inferiority (RSI).

We count the RMSE and FSIM values between all infrared images generated by these four algorithms and the corresponding real infrared images. We also calculate the average value of each index value (represented by mRMSE and mFSIM) and the RSI of the index values between the four algorithms. The statistical results are shown in Table 2. RMSE needs to consider the grey value of the corresponding points of the two images. However, there are differences (such as scale transformation, rotation, and angle) between the visible image and the real infrared image—it is not possible to fully pair the corresponding points of the target's infrared generation reference map and the same coordinates in the real infrared image. This affects the calculation of the square root error, which may lead to a larger RMSE value.

According to the experimental data given in Table 2, it can be concluded that

Table 1: The network structure of the four GANs algorithms.

Method	Networks				
	Generator				Discriminator
	U-net256	ResNet9	ResNet34	D-LinkNet34	PatchGAN
Pix2pix	√				√
Resnet9		√			√
Resnet34			√		√
Our method				√	√

Table 2: The average and the superior/inferior sample numbers of the evaluation indexes.

Method	The average of the evaluation indexes		The superior/inferior sample numbers of the evaluation indexes			
			RMSE samples		FSIM samples	
	mRMSE	mFSIM	Our'<Other'	Our'>Other'	Our'>Other'	Our'<Other'
Pix2pix	35.01	0.721	207	97	220	84
Resnet9	34.42	0.722	180	124	220	84
Resnet34	37.04	0.700	228	76	243	61
Our method	33.82	0.737	—	—	—	—

(a). Among the four algorithms, our method has the smallest mRMSE value of 33.82 and the largest mFSIM value of 0.737, which means that the quality of the infrared images generated by our method is the best;

(b) In the 304 groups of comparative data, the numbers of samples where our method's RMSE index values are better than Pix2pix, Resnet9, and Resnet34 are 207, 180, and 228, respectively;

(c) In the 304 groups of comparative data, the numbers of samples where our method's FSIM index values are better than Pix2pix, Resnet9, and Resnet34 are 220, 220, and 243, respectively.

According to the above analysis, the quality of the infrared image generated by our method is better than the other three GANs-based algorithms.

4.3.1. Statistical Results of RMSE. In order to express the experimental results more intuitively, based on the ascending order of the 304 RMSE values obtained by our algorithm, a comparison chart of the experimental results of our method and Pix2pix is drawn. As shown in Figure 8, the experimental results of our method are represented by the curve "", and the experimental results of Pix2pix are represented by the scattered points "".

It can be seen from Figure 8 that the number of "" above the curve "" is obviously more than those below the curve. Among the RMSE index results of our method, 207 index values are superior to the Pix2pix, and 97 index values are inferior to the Pix2pix. That is, the RMSE index RSI of the two algorithms is 207 : 97, indicating that, among the infrared images generated by our method, 207 images are with better quality than the Pix2pix algorithm.

According to the drawing standard in Figure 8, the RMSE index results obtained by our method, Resnet9, and Resnet34 algorithms are drawn, as shown in Figure 9. In Figure 9, the RMSE values of our method, Resnet9, and Resnet34 are represented by the curve "", the scattered point "", and the scattered point "", respectively.

As demonstrated in Figure 9, the number of "" and "" distributed above the curve "" is obviously more than those below the curve. The RMSE index RSI of our method and Resnet9 algorithm is 180 : 124, and the RSI of our method and Resnet34 algorithm is 228 : 76. These illustrate that the quality of infrared images generated by our method is significantly better than Resnet9 and Resnet34 algorithms.

4.3.2. Statistical Results of FSIM. According to the drawing standard in Figure 8, the FSIM index results obtained by our method and Pix2pix are drawn, as shown in Figure 10. In Figure 10, the FSIM values of our method and Pix2pix are represented by the curve "" and the scattered point "", respectively.

As shown in Figure 10, the number of "" below the curve "" is obviously more than those above the curve. Among the FSIM index results of our method, 220 index values are superior to the Pix2pix, and 84 index values are inferior to the Pix2pix. This indicates that the FSIM index RSI of the two algorithms is 220 : 84, which means that among the infrared images generated by our method, 220 images are with better quality than the Pix2pix algorithm.

Similarly, we draw the FSIM index results obtained by our method, Resnet9, and Resnet34 algorithms. As shown in Figure 11, the FSIM values of our method, Resnet9, and Resnet34 are represented by the curve "", the scattered point "", and the scattered point "", respectively.

As shown in Figure 11, the number of "" and "" distributed below the curve "" is obviously more than those above the curve. The FSIM index RSI of our method and Resnet9 algorithm is 220 : 84, and the RSI of our method and Resnet34 algorithm is 243 : 61. These also show that the

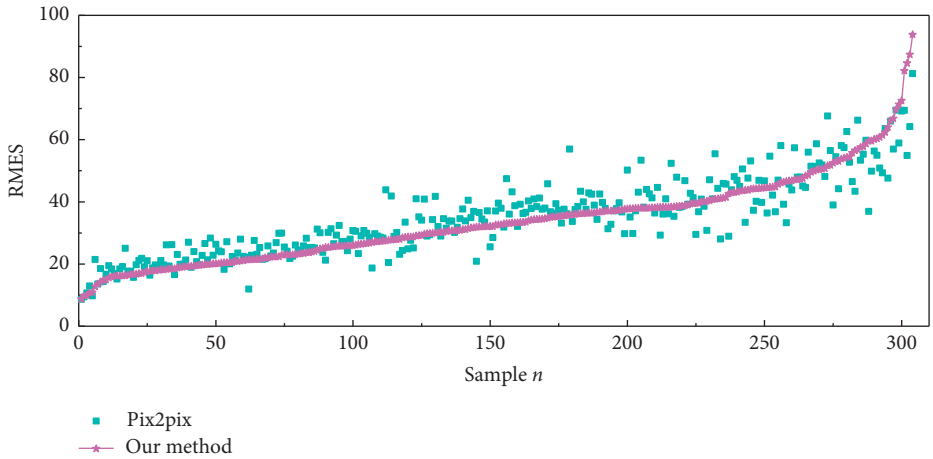


Figure 8: The RMSE results between the real infrared images and the infrared images generated by our method and the Pix2pix. The X-axis represents different test samples and the Y-axis represents the RMSE value corresponding to the sample.

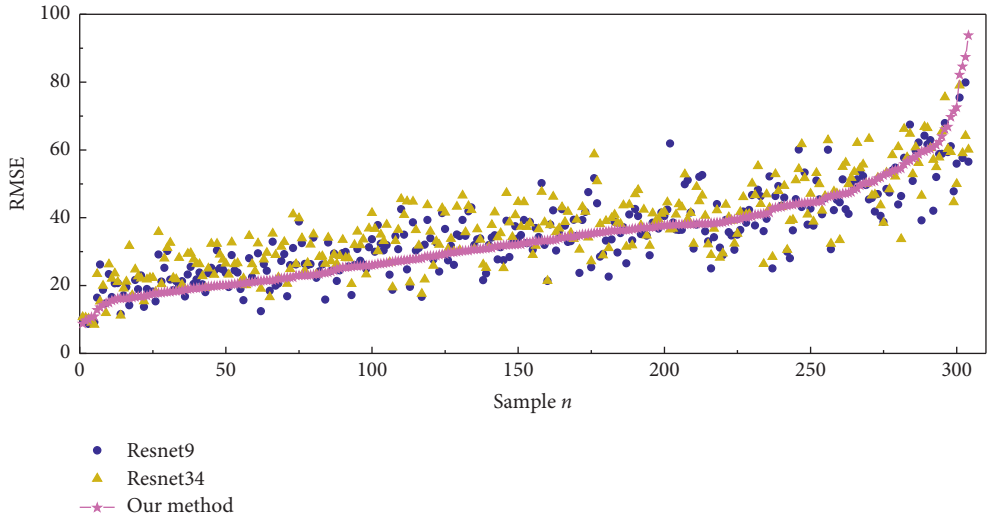


Figure 9: The RMSE results between the real infrared images and the infrared images generated by our method and the Resnet9 and Resnet34. The X-axis represents different test samples and the Y-axis represents the RMSE value corresponding to the sample.

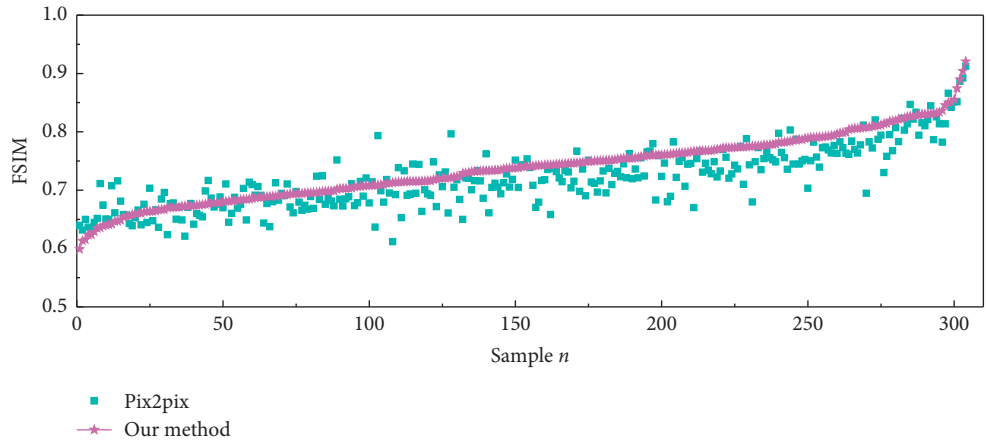


Figure 10: The FSIM results between the real infrared images and the infrared images generated by our method and the Pix2pix. The X-axis represents different test samples and the Y-axis represents the FSIM value corresponding to the sample.

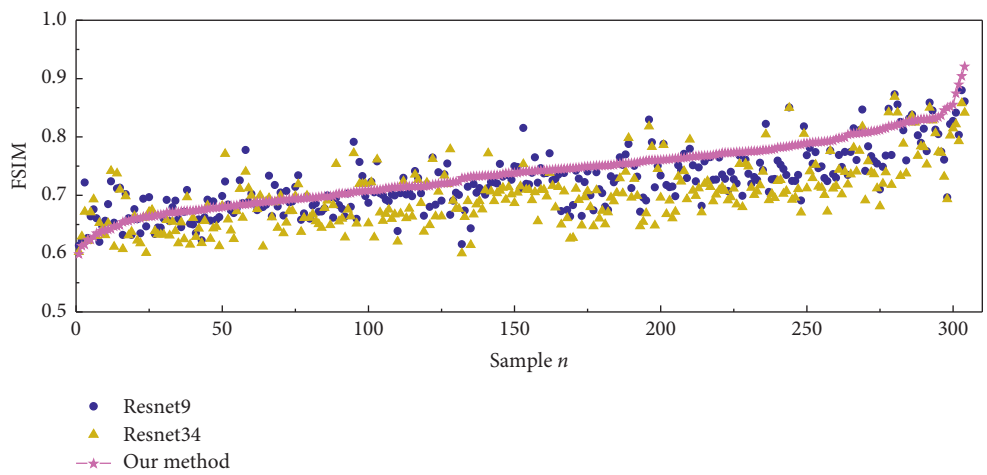


Figure 11: The FSIM results between the real infrared images and the infrared images generated by our method and the Resnet9 and Resnst34. The X-axis represents different test samples and the Y-axis represents the FSIM value corresponding to the sample.

quality of infrared images generated by our method is significantly better than Resnet9 and Resnet34 algorithms.

Based on subjective interpretation and objective analysis, it can be determined that the infrared images generated by our method (that is, I-GANs algorithm) are similar to the real infrared images; i.e., the infrared generation effect is well.

5. Conclusions

Infrared reference map preparation plays an important role in improving the accuracy and precision of infrared imaging guidance. This paper proposes an infrared image generation algorithm based on generative adversarial networks, which is named I-GANs. The algorithm introduces the D-LinkNet network to build a generation network for the purpose of learning image textures and discovering the dependencies between images. Furthermore, PatchGAN is adopted to construct a discriminant model, which can effectively process the high-frequency components of the image and reduce the amount of calculation required. In the training process, batch normalization and the Adam are utilized to optimize the training process in order to alleviate training instability and mode collapse. The simulation on the produced infrared/visible light image data (IVFG) reveals that the proposed I-GANs algorithm can generate high-quality infrared images, which are more realistic and similar to the real infrared images.

Data Availability

The data used to support this research was collected by the authors through UAV, which is equipped with a thermal infrared camera and a visible camera (both of which are coaxially installed) to capture the desired target and scene in the designated area; in brief, the designated area is photographed using a coaxial infrared camera and a visible-light camera simultaneously. Targets in the data include buildings (with materials including steel, concrete, cement, and various types of bricks), vehicles (including trucks and buses), radar covers, power stations (e.g., thermal and hydroelectric), oil depots, highways (with materials including cement and asphalt), runways, grasslands (both real and artificial), trees, and rivers (or ponds). Scenes in the data include cities, campuses, streets, factories, residential areas, transportation hubs, and rivers. Meteorological conditions identified in the data collection include sunny, cloudy, hazy, and rainy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 41574008, 61302195, and 41774156.

References

[1] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4585–4598, 2020.

[2] X. Yang, G. Lin, Y. Liu, F. Nie, and L. Lin, "Fast spectral embedded clustering based on structured graph learning for large-scale hyperspectral image," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.

[3] N. Cai, M. He, Q. Wu, and M. Khan, "On almost controllability of dynamical complex networks with noises," *Journal of Systems Science and Complexity*, vol. 32, pp. 1125–1139, 2017.

[4] Z.-Y. Tan, N. Cai, J. Zhou, and S.-G. Zhang, "On performance of peer review for academic journals: analysis based on distributed parallel system," *IEEE Access*, vol. 7, pp. 19024–19032, 2019.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets,," in *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS)*, pp. 2672–2680, Montreal, Canada, December 2014.

[6] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances:

extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[7] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 2442–2454, 2020.

[8] G. Antipov, M. Baccouche, and J. L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2089–2093, Beijing, China, September 2017.

[9] R. Huang, S. Zhang, T. Y. Li, and R. He, "Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis," in *Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV)*, pp. 2458–2467, Venice, Italy, October 2017.

[10] L. Tran, X. Yin, and X. M. Liu, "Disentangled representation learning GAN for pose-invariant face recognition," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1283–1292, Honolulu, HI, USA, July 2017.

[11] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Proceedings of the 14th European Conference on Computer Vision (ECCV)*, pp. 702–716, Amsterdam, The Netherlands, October 2016.

[12] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: feed-forward synthesis of textures and stylized images," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 1349–1357, New York, NY, USA, June 2016.

[13] P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, Honolulu, HI, USA, June 2017.

[14] A. Brock, J. Donahue, and K. Simonyan: Large Scale GAN Training for High Fidelity Natural Image Synthesis," 2018, https://arxiv.org/abs/1809.11096.

[15] J. Donahue and K. Simonyan, "Large scale adversarial representation learning," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS)*, Vancouver, Canada, February 2019.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 2015.

[17] C. Ledig, L. Theis, F. HuszSr et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, Honolulu, HI, USA, July 2017.

[18] M. Zhai, L. Chen, F. Tung et al., "Lifelong GAN: continual learning for conditional image generation," in *Proceedings of the 32th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2759–2768, Long Beach, CA, USA, September 2019.

[19] D. He, Y. Xia, T. Qin et al., "Dual learning for machine translation," in *Proceedings of the 29th International Conference on Neural Information Processing Systems (NIPS)*, pp. 820–828, San Juan, Puerto Rico, December 2016.

[20] Z. Yi, H. Zhang, P. Tan, and M. L. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proceedings of the 16th IEEE International Conference on Computer Vision (CVPR)*, pp. 2868–2876, Venice, Italy, July 2017.

[21] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the 16th IEEE International Conference on Computer Vision (CVPR)*, pp. 2223–2232, Venice, Italy, July 2017.

[22] Y. Choi, M. Choi, M. Kim et al., "StarGAN: unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the 31th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8789–8797, Salt Lake City, UT, USA, June 2018.

[23] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the 32th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, Long Beach, CA, USA, September 2019.

[24] C. Yang and S.-N. Lim, "One-Shot domain adaptation for face generation," in *Proceedings of the 33th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5920–5929, Seattle, WA, USA, August 2020.

[25] E. Richardson, Y. Alaluf, O. Patashnik et al., Encoding in Style: A StyleGAN Encoder for Image-To-Image Translation," 2020, https://arxiv.org/abs/2008.00951.

[26] Y. T. Chen, X. G. Xu, and J. Y. Jia, "Domain adaptive image-to-image translation," in *Proceedings of the 33th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5273–5282, Seattle, WA, USA, August 2020.

[27] I. J. Goodfellow, "NIPS 2016 tutorial: generative adversarial networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*, San Juan, Puerto Rico, December 2016.

[28] M. Mirza and S. Osindero: Conditional Generative Adversarial Nets," 2014, https://arxiv.org/abs/1411.1784.

[29] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, June 2018.

[30] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings f the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[31] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proceedings of the 13th IEEE International Conference on Computer Vision (ICCV)*, pp. 2018–2025, Barcelona, Spain, March 2011.

[32] L. Zhang, L. Zhang, X. Q. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2378–2386, 2011.

Hindawi
Complexity
Volume 2021, Article ID 6694587, 14 pages
https://doi.org/10.1155/2021/6694587

Research Article

Intermittent Time-Varying Formation Control for High-Order Networked Agents Subject to Discontinuous Communications

Lixin Wang, Zhe Luo, Xiaoqiang Li, Xinsan Li, and Xiaogang Yang

High-Tech Institute of Xi'an, Xi'an 710025, China

Correspondence should be addressed to Xiaogang Yang; doctoryxg@163.com

Received 10 December 2020; Revised 1 February 2021; Accepted 23 February 2021; Published 3 March 2021

Academic Editor: Ning Cai

This paper investigates the leaderless and leader-follower time-varying formation design and analysis problems for a group of networked agents subject to discontinuous communications. Firstly, a leaderless time-varying formation control protocol is proposed via the intermittent control strategy, where the control input of each agent is constructed by the distributed local state information and formation instructions in the communication time unit, but it is zero in the noncommunication time unit. Then, an explicit formulation of the formation center function is determined to describe the formation movement trajectory of the whole networked agents. Leaderless time-varying formation design and analysis with discontinuous communications are given in the form of linear matrix inequalities. Moreover, the main results of the leaderless cases are extended to the leader-follower cases. Finally, two numerical examples are provided to illustrate the theoretical results of leaderless and leader-follower cases, respectively.

1. Introduction

Distributed cooperative control has received more attention from scholars in the last two decades, which can be applied in many circumstances, including flocking [1], consensus [2–6], formation control [7–10], distributed computation [11–13], and multisource data analysis [14, 15]. Distributed formation control indicates that a group of networked intelligent agents form the desired geometrical shape via the distributed control protocol, which is constructed by the local information among neighboring agents. It was shown in [16] that the consensus-based formation control is distributed and can be achieved by utilizing the consensus algorithms and tools. Recently, the consensus theory developed fast, and many interesting research results emerged, as shown in [17–22]. As a result, the distributed formation control has aroused many researches, which can be divided into time-invariant formation and time-varying formation according to the time-dependent characteristics of the formation shape.

For the time-invariant formation, the relative position among agents remains unchanged after forming the formation structure, which means that the geometrical shape of the whole networked agents is time-invariant. From the graph theory perspective, the time-invariant formation control conditions were derived in the form of the Nyquist criterion in [23]. Jafarian et al. [24] investigated the time-invariant formation keeping problems for nonholonomic wheeled robots, where the disturbance rejection is achieved. Finite-time time-invariant formation control was achieved in [25], where a class of nonlinear control protocol was utilized. For the time-varying formation, the formation shape can be time-varying, as shown in [26, 27], which is more flexible than the time-invariant formation, and can be utilized in many practical applications. Dong et al. [28] provided a time-varying formation tracking scheme for second-order networked agents and applied it to the formation flying of a team of quadrotors. Wang et al. [29] proposed a robust time-varying formation control protocol with a distributed extend state observer, which can compensate the external disturbances actively. For a group of agents with multiple leaders, a leader-follower time-varying formation control method was shown in [30], where necessary and sufficient conditions with the formation tracking

feasibility conditions were given. Generally speaking, the time derivative of the time-varying formation is not zero and cannot be ignored in formation design and analysis, so the time-varying cases are more challenging than the time-invariant ones.

Due to the temporary interrupt of communication links, the sensing device failures, and the silent period of communications, the networked agents may suffer discontinuous communications. On the one hand, the communication topologies may be switched since the links of the network are changed. Some interesting works regarding switching topologies can be found in [31, 32], where it was shown that the communication topologies switch at some moments by the switching signal. On the other hand, the discontinuous communications are intermittent in the sequence of the time units; that is, the communication time units and the noncommunication time units appear alternately. Wang et al. [33] investigated the limited-budget consensus problems with intermittent interactions, where the consensus protocol was codesigned by the limited budget and the performance index and can guarantee the weighting optimization between the consensus performance and the energy consumption. Sun and Wang [34] proposed a new sampling-based time unit method to solve the consensus problems for nonlinear networked agents with intermittent interactions. The time-invariant formation control with intermittent interactions was studied in [35]. However, to the best of our knowledge, the time-varying formation control problems for networked agents with discontinuous communication in terms of both the switching topologies and intermittent interactions are still open.

In this paper, we investigate the leaderless and leader-follower time-varying formation design and analysis for high-order networked homogeneous agents with discontinuous communications caused by switching topologies and intermittent interactions. Firstly, a new time-varying formation control protocol is proposed via the intermittent control strategy, which only adopts the local intermittent information and formation instructions among neighboring agent. Secondly, by the nonsingular transformation and the orthonormal transformation, the closed-loop dynamics of the whole network for both leaderless cases and leader-follower cases are decomposed into two subdynamics, which, respectively, describe the formation movement trajectory of the networked agents as a whole and the relative movement among agents. Thirdly, leaderless and leader-follower time-varying formation design and analysis criteria are derived under the condition of discontinuous communications, where the convergency of the Lyapunov function is analyzed in the communication time units and noncommunication time units, and it can be guaranteed by satisfying the formation feasibility condition, the discontinuous communication condition, and the linear matrix inequality condition, simultaneously.

Compared with the related work regarding the time-varying formation control, the contribution of this paper is twofold. Firstly, different from the works in [26–30], this paper considers the discontinuous communications of both switching topologies and intermittent communications. In this case, the right-hand side of the closed-loop system is piecewise continuous. To solve this problem, a new intermittent time-varying formation control method is proposed. However, the analysis and design method in [26–30] cannot be adopted in this paper. Secondly, this paper determines the formation movement trajectory of the networked agents for both leaderless and leader-follower cases. For the leaderless case, the formation movement trajectory is determined by the formation center function, and it is determined by the zero-input response of the leader in the leader-follower case. Besides, it is shown that the intermittent communication and the switching topology do not affect the formation movement trajectory. In contrast, the authors in [26–30] did not determine the formation movement trajectory in the situation of the discontinuous communications.

The main body of this paper is arranged as follows. The model of discontinuous communications and the dynamics of the agents are established in Section 2. Leaderless time-varying formation design and analysis criteria with discontinuous communications are given in Section 3, where the explicit formulation of the formation center function is also determined. Section 4 extends the main results of the leaderless time-varying formation design and analysis to the leader-follower cases. Two numerical simulation examples are provided in Section 5, and Section 6 concludes the whole paper.

Throughout this paper, $\mathfrak{R}^{n\times d}$ stands for the $n\times d$-dimensional real matrix space. $\aleph$ is the set of natural numbers. N is utilized to denote the number of the agents, and 1_N represents the N-dimensional column vector with all components 1. The number, vector, and matrix of zero value are collectively called as 0. $Q^T = Q > 0$ means that matrix Q is symmetric and positive definite.

2. Problem Formulation and Preliminaries

2.1. Communication Constraint Modeling. In this paper, we consider the discontinuous communication among agents. On the one hand, the agent cannot communicate with each other in some noncommunication time units. On the other hand, the communication topologies of the networks are switched in some communication time units. To show the abovementioned discontinuous communication type from the time-domain perspective, it is supposed for $\forall s \in \aleph$ that there exists a nonoverlapping time unit sequence $[T_s, T_{s+1}) = [T_s, \widetilde{T}_s) \cup [\widetilde{T}_s, T_{s+1})$, where $T_s = T_s^1 < T_s^2 < \cdots < T_s^{r_s} = \widetilde{T}_s < \widetilde{T}_s + \varepsilon_s = T_{s+1}$ with r_s and ε_s being positive integers. Notice that $[T_s, \widetilde{T}_s)$ and $[\widetilde{T}_s, T_{s+1})$ represent the communication time unit and the noncommunication time unit, respectively. $T_s^1, T_s^2, \ldots, T_s^{r_s} \ldots$ denotes the switching time over which the communication topologies are switched. Without loss of generality, the initial time is assumed to be $T_0 = 0$. The length of time unit $[T_s, T_{s+1})$ satisfies that $0 < T_{\min}^* \le T_s^* = T_{s+1} - T_s \le T_{\max}^*$. It should be noted that $0 < T_{\text{dwell}} \le T_s^{r_s} - T_s^{r_{s-1}} \le \overline{T}_{\text{dwell}}$, where T_{dwell} is the minimum dwell time. The noncommunication rate is defined as $\sigma_s = (T_{s+1} - \widetilde{T}_s)/(T_{s+1} - T_s)$, where $0 < \sigma_s \le \sigma_{\max} < 1$ and $\sigma_{\max}$ is called the maximum noncommunication rate. Note that the discontinuous communication is aperiodic

since the length of each time unit $[T_s, \widetilde{T}_s)$ $(s \in \aleph)$ can be unequal.

The switching topologies are modeled as $\mathbb{G} = \{G^1, G^2, \ldots, G^k\}$ with the switching signal $\varpi(t)$: $[0, +\infty) \longrightarrow \{1, 2, \ldots, k\}$, where G is the digraph. For each digraph G, the vertex set is denoted by $\mathscr{V} = \{v_1, v_2, \ldots, v_H\}$, and the edge set is represented by $\mathscr{E} = \{(v_l, v_m)\colon\ v_l, v_m \in \mathscr{V}\}$ with the edge weight $b_{lm}^{\varpi(t)}$. Note that if there exists an edge (v_l, v_m), $v_l, v_m \in \mathscr{V}$, from the vertex v_m to v_l, then the edge weight $b_{lm}^{\varpi(t)} > 0$. Otherwise, $b_{lm}^{\varpi(t)} = 0$. The neighboring set of the vertex v_l is defined as $\mathcal{N}_l^{\varpi(t)} = \{v_m \in \mathscr{V}\colon\ (v_m, v_l) \in \mathscr{E}\}$, and $L^{\varpi(t)} = [l_{lm}^{\varpi(t)}]_{H\times H}$ stands for the Laplacian matrix of the topologies with $l_{ll}^{\varpi(t)} = \sum_{m\in\mathcal{N}_l^{\varpi(t)}} b_{lm}^{\varpi(t)}$ and $l_{lm}^{\varpi(t)} = -b_{lm}^{\varpi(t)}$ $(l \neq m)$. More details about the graph theory can be found in [36]. It should be pointed out that the switching topology does not affect the requirement of the noncommunication rate. In this paper, both leaderless and leader-follower communication topologies are considered, which satisfy the following assumption.

Assumption 1. It is assumed that the leaderless communication topology is represented by connected undirected graph, and the leader-follower communication topology is denoted by digraph containing a spanning tree with the leader locating at the root of the spanning tree.

Lemma 1. *For the connected undirected graph, the Laplacian matrix L is symmetric and positive semidefinite, and zero is the simple eigenvalue of L.*

2.2. Network Dynamics Modeling. The dynamics of the networked agents with leaderless structures are described as follows:

$$\dot{x}_l(t) = Ax_l(t) + Bu_l(t), \tag{1}$$

where $l = 1, 2, \ldots, N$, $A \in \mathfrak{R}^{n\times n}$, $B \in \mathfrak{R}^{n\times d}$, $x_l(t)$ are the states of agent l, and $u_l(t)$ is the control protocol.

Definition 1. The expected formation shape of the network is described by a vector-valued function $z(t) = [z_1^T(t), z_2^T(t), \ldots, z_N^T(t)]^T$, where $z_l(t)$ $(l = 1, 2, \ldots, N)$ is the piecewise continuous differentiable and is called the formation instruction.

According to the formation instruction, we propose a time-varying formation control protocol via the intermittent control strategy as follows:

$$u_l(t) = \begin{cases} K \displaystyle\sum_{m\in\mathcal{N}_l^{\varpi(t)}} b_{lm}^{\varpi(t)}\left(x_m(t) - z_m(t) - x_l(t) + z_l(t)\right), & t \in \left[T_s, \widetilde{T}_s\right), \\ 0, & t \in \left[\widetilde{T}_s, T_{s+1}\right), \end{cases} \tag{2}$$

where $l = 1, 2, \ldots, N$, $K \in \mathfrak{R}^{d\times n}$ is the gain matrix. Then, the definitions of the leaderless time-varying formation design and analysis are given as follows.

Definition 2. (*leaderless time-varying formation analysis*). For any given gain matrix $K \in \mathfrak{R}^{d\times n}$ and bounded initial states $x_l(0) - z_l(0)$, $l = 1, 2, \ldots, N$, if there exists a function $h(t) \in \mathfrak{R}^n$ such that $\lim_{t\longrightarrow+\infty}(x_l(t) - z_l(t) - h(t)) = 0$, $l = 1, 2, \ldots, N$, then it is said that network (1) with protocol (2) reaches leaderless time-varying formation, where $h(t)$ is said to be the formation center function.

Definition 3. (*leaderless time-varying formation design*). If there exists a gain matrix K such that network (1) with protocol (2) reaches leaderless time-varying formation, then it is said to be leaderless time-varying formation reachable.

In this paper, we mainly focus on designing the gain matrix K such that network (1) subject to discontinuous communications reaches leaderless time-varying formation with protocol (2). Then, the main results on the leaderless time-varying formation are extended to the leader-follower cases.

Remark 1. It should be noticed that protocol (2) is constructed via the intermittent control strategy; that is, the control input is intermittent over the nonoverlapping time unit sequence $[T_s, T_{s+1})$, $\forall s \in \aleph$. In the communication time unit, the control input is established according to the local state information among neighboring agents and the formation instructions. However, in the noncommunication time unit, the control input is set to be zero since it is missing. This kind of intermittent control strategy will lead to the piecewise continuous right-hand side of the closed-loop networks, which is challenging to be dealt with in the stability analysis of the networked agents.

Remark 2. Note that, for the consensus, it requires that all the agents reach an agreement of states, where the formation structure is not needed. For the time-invariant formation, the formation instruction is invariable, whose time derivate is zero; i.e., $\dot{z}_l \equiv 0$ $(l = 1, 2, \ldots, N)$. Compared with the consensus and the time-invariant formation, the main difficulty in designing the time-varying formation protocol is that the time derivate of the formation instruction $z_l(t)$ $(l = 1, 2, \ldots, N)$ affects the analysis of the convergency of the

formation. In this case, the formation feasibility condition is introduced to overcome the challenging problems in designing the gain matrix of the time-varying formation protocol.

3. Leaderless Time-Varying Formation Design and Analysis

In this section, we give leaderless time-varying formation design and analysis criteria under the condition of discontinuous communications, and then, we determine an explicit formulation of the formation center function.

For $l = 1, 2, \ldots, N$, let $\varphi_l(t) = x_l(t) - z_l(t)$, substituting (2) into (1) gives that

$$\dot{\varphi}_l(t) = \begin{cases} A(\varphi_l(t) + z_l(t)) + BK \sum_{m \in \mathcal{N}_l^{\varpi(t)}} b_{lm}^{\varpi(t)} (\varphi_m(t) - \varphi_l(t)) - \dot{z}_i(t), & t \in [T_s, \widetilde{T}_s), \\ A(\varphi_l(t) + z_l(t)) - \dot{z}_i(t), & t \in [\widetilde{T}_s, T_{s+1}). \end{cases} \tag{3}$$

Denote $\varphi(t) = [\varphi_1^T(t), \varphi_2^T(t), \ldots, \varphi_N^T(t)]^T$, then we can rewrite equation (3) as

$$\dot{\varphi}(t) = \begin{cases} (I_N \otimes A)(\varphi(t) + z(t)) - \left(L^{\varpi(t)} \otimes BK\right)\varphi(t) - (I_N \otimes I_n)\dot{z}(t), & t \in [T_s, \widetilde{T}_s), \\ (I_N \otimes A)(\varphi(t) + z(t)) - (I_N \otimes I_n)\dot{z}(t), & t \in [\widetilde{T}_s, T_{s+1}). \end{cases} \tag{4}$$

Since Assumption 1 holds, we can find from Lemma 1 that there exists an orthonormal matrix $W^{\varpi(t)} = [1_N/\sqrt{N}, \widetilde{W}^{\varpi(t)}]$ such that

$$\left(W^{\varpi(t)}\right)^T L^{\varpi(t)} W^{\varpi(t)} = \operatorname{diag}\left\{\lambda_1^{\varpi(t)}, \lambda_2^{\varpi(t)}, \ldots, \lambda_N^{\varpi(t)}\right\}, \tag{5}$$

where $0 = \lambda_1^{\varpi(t)} < \underline{\lambda} \le \lambda_2^{\varpi(t)} \le \lambda_3^{\varpi(t)} \le \cdots \le \lambda_N^{\varpi(t)} \le \overline{\lambda}$ are the eigenvalues of the Laplacian matrix. Denote $\Omega^{\varpi(t)} = \operatorname{diag}\left\{\lambda_2^{\varpi(t)}, \lambda_3^{\varpi(t)}, \ldots, \lambda_N^{\varpi(t)}\right\}$ and $\upsilon(t) = ((W^{\varpi(t)})^T \otimes I_n)x(t) = [\upsilon_1^T(t), \kappa^T(t)]^T$, where $\kappa^T(t) = [\upsilon_2^T(t), \upsilon_3^T(t), \ldots, \upsilon_N^T(t)]^T$, and then, we can transform network (4) into the following form:

$$\dot{\upsilon}_1(t) = A\upsilon_1(t) + \left(\frac{1_H^T}{\sqrt{H}} \otimes A\right) z(t) - \left(\frac{1_H^T}{\sqrt{H}} \otimes I_n\right)\dot{z}(t), \quad t \in [T_s, T_{s+1}), \tag{6}$$

$$\dot{\kappa}(t) = \begin{cases} \left(I_{H-1} \otimes A - \Omega^{\varpi(t)} \otimes BK\right)\kappa(t) + \left(\left(\widetilde{W}^{\varpi(t)}\right)^T \otimes A\right) z(t) - \left(\left(\widetilde{W}^{\varpi(t)}\right)^T \otimes I_n\right)\dot{z}(t), & t \in [T_s, \widetilde{T}_s), \\ (I_{H-1} \otimes A)\kappa(t) + \left(\left(\widetilde{W}^{\varpi(t)}\right)^T \otimes A\right) z(t) - \left(\left(\widetilde{W}^{\varpi(t)}\right)^T \otimes I_n\right)\dot{z}(t), & t \in [\widetilde{T}_s, T_{s+1}). \end{cases} \tag{7}$$

From the above orthonormal transformation, network (4) is decomposed into two subnetworks; i.e., subnetworks (6) and (7). In the following, we will provide a theorem to describe the explicit formulation of the formation center function according to the dynamics of subnetwork (6).

Theorem 1. *If network (1) with protocol (2) reaches the time-varying formation $z(t)$, then the formation center function satisfies that*

$$\lim_{t \longrightarrow +\infty} (h(t) - h_c(t) + h_z(t)) = 0, \tag{8}$$

where

$$h_c(t) = e^{At}\left(\frac{1_N^T}{N} \otimes I_n\right) x(0), \qquad h_z(t) = \left(\frac{1_N^T}{N} \otimes I_n\right) z(t). \tag{9}$$

Proof. Define the following functions:

$$\Gamma_c(t) \triangleq W^{\varpi(t)} e_1 \otimes v_1(t) = \frac{1}{\sqrt{H}} 1_H \otimes v_1(t), \tag{10}$$

$$\Gamma_z(t) \triangleq \sum_{l=2}^{H} W^{\varpi(t)} e_l \otimes v_l(t), \tag{11}$$

where e_l $(l \in \{1, 2, \ldots, N\})$ is a N-dimensional unit column vector, whose l th element is 1. According to equation (10), we have

$$\Gamma_c(t) = \left(W^{\varpi(t)} \otimes I_n\right)\left[v_1^T(t), 0\right]^T. \tag{12}$$

Since

$$\sum_{l=2}^{N} W^{\varpi(t)} e_l \otimes v_l(t) = \left[0, \kappa^T(t)\right]^T, \tag{13}$$

we can derive that

$$\Gamma_z(t) = \left(W^{\varpi(t)} \otimes I_n\right)\left[0, \kappa^T(t)\right]^T. \tag{14}$$

Because $W^{\varpi(t)} \otimes I_n$ is a nonsingular matrix, we can see from the above analysis that $\Gamma_c(t)$ and $\Gamma_z(t)$ are linearly independent. Hence, we can use $v_1(t)/\sqrt{N}$ to determine the explicit formulation of the formation center function.

According to Definition 2 and equation (12), we have

$$\lim_{t \longrightarrow +\infty} \left(x_l(t) - z_l(t) - \frac{1}{\sqrt{N}} v_1(t) \right) = 0. \tag{15}$$

From equation (16), we have

$$v_1(0) = \left(\left(\frac{1}{\sqrt{N}} 1_N^T \right) \otimes I_n \right) \varphi(0). \tag{16}$$

Then, we can derive that

$$\begin{aligned} &\int_0^t e^{A(t-\tau)} \left(\left(\frac{1}{\sqrt{N}} 1_N^T \right) \otimes I_n \right) \dot{z}(\tau) \mathrm{d}\tau \\ &= \left(\left(\frac{1}{\sqrt{N}} 1_N^T \right) \otimes I_n \right) z(t) - e^{At} \left(\left(\frac{1}{\sqrt{N}} 1_N^T \right) \otimes I_n \right) z(0) \\ &\quad + \int_0^t e^{A(t-\tau)} \left(\left(\frac{1}{\sqrt{N}} 1_N^T \right) \otimes A \right) z(\tau) \mathrm{d}\tau. \end{aligned} \tag{17}$$

From equations (6), (16), and (17), we can obtain the conclusion of Theorem 1. □

Remark 3. The explicit formulation of the formation center function in Theorem 1 describes the formation movement trajectory of the networked agents as a whole. Note that when the time-varying formation is reached, all the networked agents will keep a formation shape and move along with the trajectory determined by the formation center function, which contains two parts. The first part $h_c(t)$ is associated with the dynamics and the initial states of each agent, which describe the influence mechanism of the consensus states on the formation movement. The second part $h_z(t)$ is related to the time-varying formation instruction. From the formulation of $h(t)$, we can find that the discontinuous communication does not impact the formation movement trajectory of the whole networked agents.

According to the analysis from (10) to (14), we can conclude that network (1) reaches time-varying formation if and only if $\lim_{t \longrightarrow +\infty} \kappa(t) = 0$. Based on this fact, we give the time-varying formation design criterion in the following theorem.

Theorem 2. *Network (1) is leaderless time-varying formation reachable by protocol (2) with $K = 0.5\underline{\lambda}^{-1}\vartheta B^T \tilde{Q}^{-1}$ if the following conditions hold simultaneously:*

(i) The formation feasibility condition $\dot{z}_l(t) = Az_l(t)$ $(l = 1, 2, \ldots, N)$ holds.

(ii) The discontinuous communication condition $\alpha(1 - \sigma_{\max}) > \mu\sigma_{\max}$ is satisfied, where α and μ are positive constants given previously.

(iii) There exist $\vartheta > 0$ and $\tilde{Q} = \tilde{Q}^T > 0$ such that

$$\begin{aligned} A\tilde{Q} + \tilde{Q}A^T - \mu\tilde{Q} < 0, \\ A\tilde{Q} + \tilde{Q}A^T + \alpha\tilde{Q} - \vartheta BB^T < 0. \end{aligned} \tag{18}$$

Proof. Construct the Lyapunov function as follows:

$$V(t) = \kappa^T(t)\left(I_{N-1} \otimes \tilde{Q}^{-1}\right)\kappa(t). \tag{19}$$

For $t \in [T_s, \tilde{T}_s)$ with $\forall s \in \aleph$, we can obtain the time derivative of $V(t)$ according to the dynamics of subnetwork (7) that

$$\begin{aligned} \dot{V}(t) = \kappa^T(t)\Big(& I_{N-1} \otimes \left(\tilde{Q}^{-1}A + A^T\tilde{Q}^{-1}\right) \\ & - \Omega^{\varpi(t)} \otimes \left(\tilde{Q}^{-1}BK + K^TB^T\tilde{Q}^{-1}\right)\Big)\kappa(t) \\ & + 2\kappa^T(t)\left(\left(\tilde{W}^{\varpi(t)}\right)^T \otimes \tilde{Q}^{-1}A\right)z(t) \\ & - 2\kappa^T(t)\left(\left(\tilde{W}^{\varpi(t)}\right)^T \otimes \tilde{Q}^{-1}\right)\dot{z}(t). \end{aligned} \tag{20}$$

Since the formation feasibility condition $\dot{z}_l(t) = Az_l(t)$ $(l = 1, 2, \ldots, N)$ holds, we can deduce that $((\tilde{W}^{\varpi(t)})^T \otimes \tilde{Q}^{-1})\dot{z}(t) = ((\tilde{W}^{\varpi(t)})^T \otimes \tilde{Q}^{-1}A)z(t)$. Then, we can see from (20) that

$$\begin{aligned} \dot{V}(t) = \kappa^T(t)\Big(& I_{N-1} \otimes \left(\tilde{Q}^{-1}A + A^T\tilde{Q}^{-1}\right) \\ & - \Omega^{\varpi(t)} \otimes \left(\tilde{Q}^{-1}BK + K^TB^T\tilde{Q}^{-1}\right)\Big)\kappa(t). \end{aligned} \tag{21}$$

Substituting $K = 0.5\underline{\lambda}\vartheta B^T\tilde{Q}^{-1}$ into equation (21) yields

$$\dot{V}(t)+\alpha V(t)=\sum_{l=2}^{N}\kappa_l^T(t)\Big(\widetilde{Q}^{-1}A+A^T\widetilde{Q}^{-1}+\alpha\widetilde{Q}^{-1} -\lambda_l^{\varpi(t)}\underline{\lambda}^{-1}\vartheta\widetilde{Q}^{-1}BB^T\widetilde{Q}^{-1}\Big)\kappa_l(t). \tag{22}$$

By pre- and postmultiplying $A\widetilde{Q}+\widetilde{Q}A^T+\alpha\widetilde{Q}-\vartheta BB^T<0$ with $\widetilde{Q}^{-1}$ and $-\lambda_l^{\varpi(t)}\underline{\lambda}^{-1}\leq 1$ $(l=2,3,\ldots,N)$, we can obtain from (22) that

$$\dot{V}(t)<-\alpha V(t). \tag{23}$$

For $t\in[\widetilde{T}_s,T_{s+1})$ with $\forall s\in\aleph$, we can obtain the time derivative of $V(t)$ along the dynamics of subnetwork (7) that

$$\dot{V}(t)=\kappa^T(t)\Big(I_{N-1}\otimes\Big(\widetilde{Q}^{-1}A+A^T\widetilde{Q}^{-1}\Big)\Big)\kappa(t). \tag{24}$$

Then, it follows that

$$\dot{V}(t)-\mu V(t)=\sum_{l=2}^{N}\kappa_l^T(t)\Big(\widetilde{Q}^{-1}A+A^T\widetilde{Q}^{-1}-\mu\widetilde{Q}^{-1}\Big)\kappa_l(t). \tag{25}$$

According to $A\widetilde{Q}+\widetilde{Q}A^T-\mu\widetilde{Q}<0$, we have

$$\dot{V}(t)<\mu V(t). \tag{26}$$

In the sequel, we discuss the convergency of $V(t)$ along the time unit sequence $[T_s,T_{s+1})$, $\forall s\in\aleph$. Firstly, for $t\in[T_0,T_1)$, we can show that

$$\begin{aligned}V(T_1)&<e^{\mu\left(T_1-\widetilde{T}_0\right)}V\left(\widetilde{T}_0\right)<e^{\mu\left(T_1-\widetilde{T}_0\right)}e^{-\alpha\left(\widetilde{T}_0-T_0\right)}V(T_0)\\&=e^{-\beta_0}V(0),\end{aligned} \tag{27}$$

where $\beta_0=(\alpha-(\alpha+\mu)\sigma_0)T_0^*$. By the discontinuous communication condition $\alpha(1-\sigma_{\max})>\mu\sigma_{\max}$, we can see that $\beta_0>0$. Then, we have for any positive integer s that

$$V(T_{s+1})<V(0)e^{-\sum_{q=0}^{s}\beta_q}, \tag{28}$$

where $\beta_q=(\alpha-(\alpha+\mu)\sigma_q)T_q^*$, $q=1,2,\ldots,s$. For any $t>0$, we can see that an integer $i\geq 1$ exists such that $T_i<t\leq T_{i+1}$. Hence, we have

$$\begin{aligned}V(t)&\leq e^{\mu T_{\max}^*}V(T_i)\leq e^{\mu T_{\max}^*}V(0)e^{-\sum_{j=0}^{i-1}\beta_j}\leq e^{\mu T_{\max}^*}\\&\quad\cdot V(0)e^{-i\left(\alpha-(\alpha+\mu)\sigma_{\max}\right)T_{\min}^*}\\&\leq e^{\mu T_{\max}^*}V(0)e^{-\left(\left(\left(\alpha-(\alpha+\mu)\sigma_{\max}\right)T_{\min}^*\right)/T_{\max}^*\right)t}.\end{aligned} \tag{29}$$

According to equation (29), we can conclude that $\lim_{t\longrightarrow+\infty}\kappa(t)=0$, which means that network (1) reaches time-varying formation exponentially. The proof of Theorem 2 is finished.

The time-varying formation design criterion in Theorem 2 provides an approach to design the gain matrix of protocol (2). However, if the gain matrix is given previously, then it is interesting to check that the given gain matrix is feasible or not. Let $Q=\widetilde{Q}^{-1}$, then the following corollary gives the time-varying formation analysis criterion. □

Corollary 1. *For any given gain matrix K, network (1) with protocol (2) reaches leaderless time-varying formation if the following conditions hold simultaneously:*

(i) The formation feasibility condition $\dot{z}_l(t)=Az_l(t)$ $(l=1,2,\ldots,N)$ holds.

(ii) The discontinuous communication condition $\alpha(1-\sigma_{\max})>\mu\sigma_{\max}$ is satisfied, where $\alpha>0$ and $\mu>0$.

(iii) There exist $Q=Q^T>0$ such that

$$\begin{aligned}&QA+A^TQ-\mu Q<0,\\&QA+A^TQ+\alpha Q-\underline{\lambda}\left(QBK+K^TB^TQ\right)<0,\\&QA+A^TQ+\alpha Q-\overline{\lambda}\left(QBK+K^TB^TQ\right)<0.\end{aligned} \tag{30}$$

Remark 4. The formation feasibility condition in Theorem 2 and Corollary 1 indicates that not all the desired formation instructions can be reached by the networked agents. We can find intuitively that the formation structure to be formed is restrained by the dynamics of each agent. For example, due to the constraint of the maneuvering characteristics, a team of unmanned aerial vehicles cannot perform some specific actions and thus cannot form the corresponding formation shape. For the time-varying formation, some formation feasibility conditions similar to the condition $\dot{z}_l(t)=Az_l(t)$ $(l=1,2,\ldots,N)$ can be found in [28–30]. Note that if the derivative of the time-varying formation instruction is zero; i.e., $\dot{z}_l(t)\equiv 0$, $(l=1,2,\ldots,N)$, then the formation feasibility condition becomes $Az_l(t)=0$. In this case, the formation instruction is time-invariant.

Remark 5. Since the intermittent control strategy leads to the piecewise continuous right-hand side of the closed-loop systems, the stability property of the closed-loop systems should be analyzed in the communication time units and the noncommunication time units. It should be pointed out that the states of the closed-loop systems may be divergent in the noncommunication time units. To ensure the whole convergency of the Lyapunov function, the discontinuous communication condition $\alpha(1-\sigma_{\max})>\mu\sigma_{\max}$ is provided, which can establish the relationship between the convergency factor α and the divergent factor μ via the maximum noncommunication rate $\sigma_{\max}$. As the result, the Lyapunov function $V(t)$ can be convergent with the rate faster than $(\alpha-(\alpha+\mu)\sigma_{\max})T_{\min}^*/T_{\max}^*$ in virtue of inequality (29).

4. Extensions to Leader-Follower Cases

This section extends the main results of the leaderless time-varying formation design and analysis with discontinuous communications to the leader-follower cases.

The dynamics of networked agents with leader-follower structures are modeled as

$$\begin{cases} \dot{x}_N(t) = Ax_N(t), \\ \dot{x}_l(t) = Ax_l(t) + Bu_l(t), \end{cases} \tag{31}$$

where $l = 1, 2, \ldots, N-1$ are the labels of $N-1$ followers, and the leader is labeled by the subscript N.

Assumption 2. It is assumed that the communication topology among followers is represented by the connected undirected graph.

For leader-follower structures, we propose the following time-varying formation control protocol:

$$u_l(t) = \begin{cases} K_{lf} \sum_{m \in \mathcal{N}_l^{\varpi(t)}} b_{lm}^{\varpi(t)} \left(x_m(t) - z_m(t) - x_l(t) + z_l(t)\right), & t \in \left[T_s, \widetilde{T}_s\right), \\ +K_{lf} b_{lN}^{\varpi(t)} \left(x_N(t) - x_l(t) + z_l(t)\right), & \\ 0, & t \in \left[\widetilde{T}_s, T_{s+1}\right), \end{cases} \tag{32}$$

where $l = 1, 2, \ldots, N-1$, $K_{lf} \in \mathfrak{R}^{d \times n}$ is the gain matrix. Then, the definitions of the leader-follower time-varying formation design and analysis are given as follows.

Definition 4. (*leader-follower time-varying formation analysis*). For any given gain matrix $K_{lf} \in R^{d \times n}$ and bounded initial states $x_l(0) - z_l(0)$, $l = 1, 2, \ldots, N-1$, if $\lim_{t \longrightarrow +\infty} (x_l(t) - z_l(t) - x_N(t)) = 0$, $l = 1, 2, \ldots, N-1$, then it is said that network (1) with protocol (2) reaches leader-follower time-varying formation.

Definition 5. (*leader-follower time-varying formation design*). If there exists a gain matrix K_{lf} such that network (1) with protocol (2) reaches leader-follower time-varying formation, then it is said to be leader-follower time-varying formation reachable.

Form (31) and (32), we can obtain that

$$\dot{\varphi}(t) = \begin{cases} (I_N \otimes A)(\varphi(t) + z(t)) - \left(L_{lf}^{\varpi(t)} \otimes BK_{lf}\right)\varphi(t) - (I_N \otimes I_n)\dot{z}(t), & t \in \left[T_s, \widetilde{T}_s\right), \\ (I_N \otimes A)(\varphi(t) + z(t)) - (I_N \otimes I_n)\dot{z}(t), & t \in \left[\widetilde{T}_s, T_{s+1}\right), \end{cases} \tag{33}$$

where $L_{lf}^{\varpi(t)}$ satisfies that

$$\begin{aligned} L_{lf}^{\varpi(t)} &= \begin{bmatrix} L_f^{\varpi(t)} + \Delta_l^{\varpi(t)} & -l_l^{\varpi(t)} \\ 0 & 0 \end{bmatrix}, \\ \Delta_l^{\varpi(t)} &= \operatorname{diag}\left\{b_{1N}^{\varpi(t)}, b_{2N}^{\varpi(t)}, \ldots, b_{(N-1)N}^{\varpi(t)}\right\}, \\ l_l^{\varpi(t)} &= \left[b_{1N}^{\varpi(t)}, b_{2N}^{\varpi(t)}, \ldots, b_{(N-1)N}^{\varpi(t)}\right]^T. \end{aligned} \tag{34}$$

$L_f^{\varpi(t)}$ is the Laplacian matrix of the topology among followers.

Construct the following nonsingular matrix:

$$Y^{\varpi(t)} = \begin{bmatrix} I_{N-1} & 1_{N-1} \\ 0 & 1 \end{bmatrix}. \tag{35}$$

Let $\widetilde{\varphi}_l(t) = \varphi_l(t) - x_N(t)$ $(l = 1, 2, \ldots, N-1)$, then we have

$$\left(\left(Y^{\varpi(t)}\right)^{-1} \otimes I_n\right)\varphi(t) = \left[\widetilde{\varphi}_2^T(t), \ldots, \widetilde{\varphi}_N^T(t), x_1^T(t)\right]^T. \tag{36}$$

Then, it follows from $\Delta_l^{\varpi(t)} 1_{N-1} = l_l^{\varpi(t)}$ that

$$\left(Y^{\varpi(t)}\right)^{-1} L^{\varpi(t)} Y^{\varpi(t)} = \begin{bmatrix} L_f^{\varpi(t)} + \Delta_l^{\varpi(t)} & 0 \\ 0 & 0 \end{bmatrix}. \tag{37}$$

Since Assumptions 1 and 2 hold, we can find that $L_f^{\varpi(t)} + \Delta_l^{\varpi(t)}$ is positive definite and symmetric. Hence, there exists an orthonormal matrix $\widetilde{Y}^{\varpi(t)} \in \mathfrak{R}^{(N-1)\times(N-1)}$ such that

$$\left(\widetilde{Y}^{\varpi(t)}\right)^T \left(L_f^{\varpi(t)} + \Delta_l^{\varpi(t)}\right)\widetilde{Y}^{\varpi(t)} = \Omega_f^{\varpi(t)} = \operatorname{diag}\left\{\widetilde{\lambda}_1^{\varpi(t)}, \widetilde{\lambda}_2^{\varpi(t)}, \ldots, \widetilde{\lambda}_{N-1}^{\varpi(t)}\right\}, \tag{38}$$

where $0 < \widetilde{\lambda}_1^{\varpi(t)} \leq \widetilde{\lambda}_2^{\varpi(t)} \leq \cdots \leq \widetilde{\lambda}_{N-1}^{\varpi(t)}$ are the eigenvalues of $L_f^{\varpi(t)}$. Let $\widetilde{\varphi}(t) = [\widetilde{\varphi}_1^T(t), \widetilde{\varphi}_2^T(t), \ldots, \widetilde{\varphi}_{N-1}^T(t)]^T$ and $((Y^{\varpi(t)})^{-1} \otimes I_n)\widetilde{\varphi}(t) = \eta(t) = [\eta_1^T(t), \eta_2^T(t), \ldots, \eta_{N-1}^T(t)]^T$, then network (33) is converted to the following two subnetworks:

$$\dot{x}_N(t) = Ax_N(t),$$
$$\dot{\eta}(t) = \begin{cases} \left(I_{N-1}\otimes A - \Omega_f^{\varpi(t)}\otimes BK_{lf}\right)\eta(t) + \left(\left(\widetilde{Y}^{\varpi(t)}\right)^T\otimes A\right)z(t) - \left(\left(\widetilde{Y}^{\varpi(t)}\right)^T\otimes I_n\right)\dot{z}(t), & t\in\left[T_s,\widetilde{T}_s\right), \\ \left(I_{N-1}\otimes A\right)\eta(t) + \left(\left(\widetilde{Y}^{\varpi(t)}\right)^T\otimes A\right)z(t) - \left(\left(\widetilde{Y}^{\varpi(t)}\right)^T\otimes I_n\right)\dot{z}(t), & t\in\left[\widetilde{T}_s,T_{s+1}\right). \end{cases} \tag{39}$$

Because $Y^{\varpi(t)}$ and $\widetilde{Y}^{\varpi(t)}$ are nonsingular, we can find that network (31) reaches leader-follower time-varying formation if and only if $\lim_{t\longrightarrow+\infty}\eta(t) = 0$. Let $\underline{\lambda}_{lf} = \min\left\{\widetilde{\lambda}_1^i\colon\ \forall i\in\{1,2,\ldots,k\}\right\}$ and $\overline{\lambda}_{lf} = \max\left\{\widetilde{\lambda}_1^i\colon\ \forall i\in\{1,2,\ldots,k\}\right\}$, then we give the following theorem to show the sufficient conditions of the leader-follower time-varying formation design.

Theorem 3. *Network (31) is leader-follower time-varying formation reachable by protocol (32) with $K_{lf} = 0.5\underline{\lambda}_{lf}^{-1}\xi B^T\widetilde{R}^{-1}$ if the following conditions hold simultaneously:*

(i) The formation feasibility condition $\dot{z}_l(t) = Az_l(t)$ $(l = 1,2,\ldots,N)$ holds.

(ii) The discontinuous communication condition $\alpha(1-\sigma_{\max}) > \mu\sigma_{\max}$ is satisfied, where $\alpha > 0$ and $\mu > 0$.

(iii) There exist $\xi > 0$ and $\widetilde{R} = \widetilde{R}^T > 0$ such that

$$\begin{aligned} A\widetilde{R} + \widetilde{R}A^T - \mu\widetilde{R} < 0, \\ A\widetilde{R} + \widetilde{R}A^T + \alpha\widetilde{R} - \xi BB^T < 0. \end{aligned} \tag{40}$$

Corollary 2. *For any given gain matrix K_{lf}, network (31) with protocol (32) reaches leader-follower time-varying formation if the following conditions hold simultaneously:*

(i) The formation feasibility condition $\dot{z}_l(t) = Az_l(t)$ $(l = 1,2,\ldots,N)$ holds.

(ii) The discontinuous communication condition $\alpha(1-\sigma_{\max}) > \mu\sigma_{\max}$ is satisfied, where $\alpha > 0$ and $\mu > 0$.

(iii) There exists $R = R^T > 0$ such that

$$\begin{aligned} RA + A^TR - \mu R < 0, \\ RA + A^TR + \alpha R - \underline{\lambda}_{lf}\left(RBK_{lf} + K_{lf}^TB^TR\right) < 0, \\ RA + A^TR + \alpha R - \overline{\lambda}_{lf}\left(RBK_{lf} + K_{lf}^TB^TR\right) < 0. \end{aligned} \tag{41}$$

Remark 6. The leader-follower time-varying formation can be regarded as an extension of the leaderless time-varying formation since their main conclusions are similar. In this case, the control gains of these two cases can be designed via solving the linear matrix inequalities in the similar form. However, the topology structures of these two cases are different, which can be reflected on the difference of the eigenvalues of the Laplacian matrices. Moreover, the formation trajectories of the leaderless time-varying formation and the leader-follower time-varying formation are different. For the leaderless case, the formation trajectory is determined by the states of all agents and the formation instruction, which is described by the formation center function. For the leader-follower case, the formation trajectory is determined by the leader.

5. Numerical Simulations

In this section, we provide two numerical simulation examples to demonstrate the effectiveness of the proposed theorems regarding the leaderless and leader-following time-varying formation design and analysis with discontinuous communications.

Example 1. (*leaderless topologies*). Consider a group of networked agents labeled by 1–6, which are of third order as follows:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & -1 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}. \tag{42}$$

The initial states of the agents are set as

$$\begin{aligned} x_1(0) &= [3.5,-5.1,-2.8]^T, \\ x_2(0) &= [-2.3,-7.5,1.6]^T, \\ x_3(0) &= [2.8,-4.6,1.3]^T, \\ x_4(0) &= [2.7,-2.1,2.1]^T, \\ x_5(0) &= [3.7,-1.4,-5.9]^T, \\ x_6(0) &= [7.2,-2.9,4.3]^T. \end{aligned} \tag{43}$$

The switching topologies are given in Figure 1, where the dwell time is set as 0.3 s.

The communication time unit and noncommunication time unit are set to be $t\in[s,s+0.8)s$ and $t\in[s+0.8,s+1)s$, respectively, which are period time units for better operability of the simulation. In this case, the maximum noncommunication rate $\sigma_{\max} = 0.2$. Choose $\alpha = 1.6$ and $\mu = 6$, and then, we can find that the discontinuous communication condition $\alpha(1-\sigma_{\max}) > \mu\sigma_{\max}$ is satisfied. The time-varying formation instruction is chosen as follows:

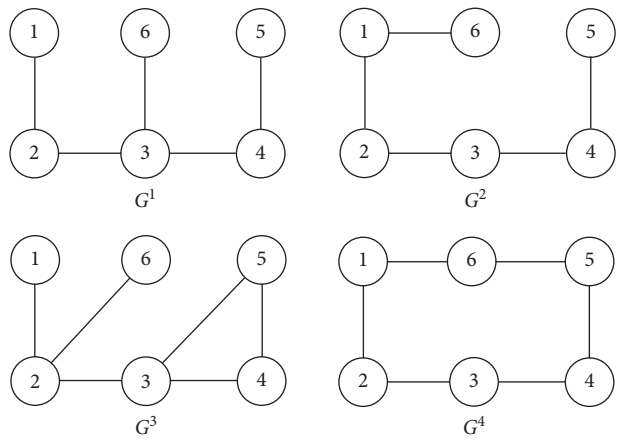


Figure 1: Leaderless switching topologies.

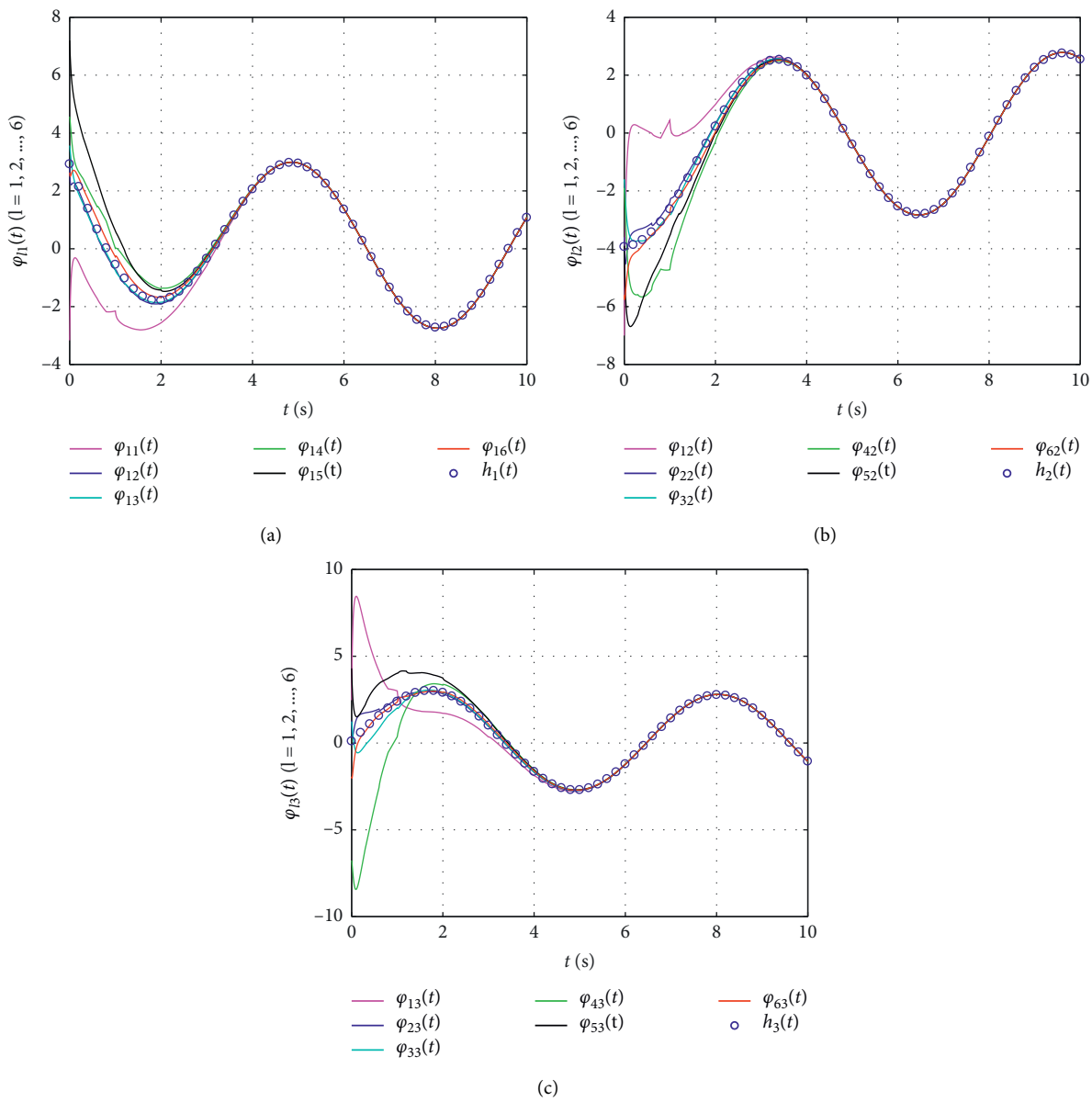


Figure 2: Trajectories of φ_l $(l = 1, 2, \ldots, 6)$.

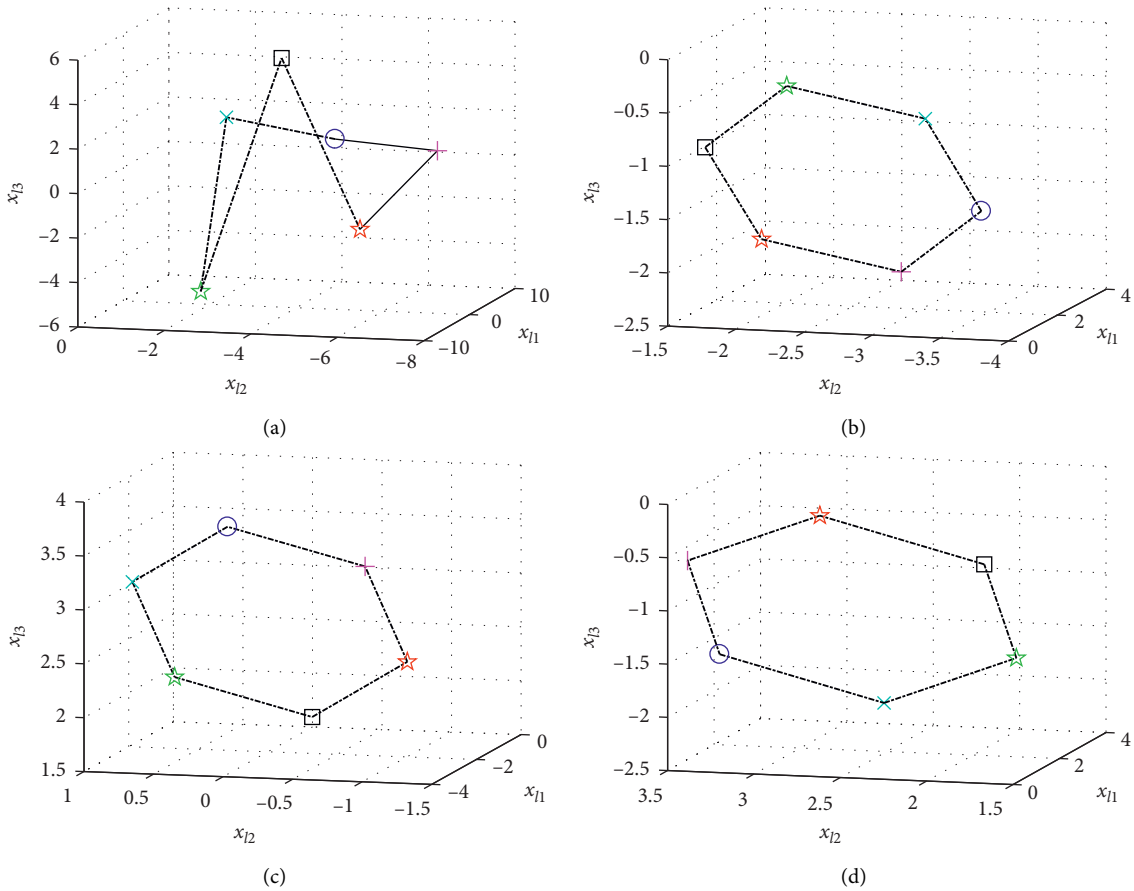


Figure 3: State snapshots of six agents at different times: (a) $t = 0$ s; (b) $t = 8$ s; (c) $t = 9$ s; (d) $t = 10$ s.

$$z_l(t) = \begin{bmatrix} \sin\left(t + \dfrac{(l-1)\pi}{3}\right) \\ \cos\left(t + \dfrac{(l-1)\pi}{3}\right) \\ -\sin\left(t + \dfrac{(l-1)\pi}{3}\right) \end{bmatrix}, \quad l = 1, 2, \ldots, 6. \tag{44}$$

Based on Theorem 2, the matrix variable and the gain matrix are calculated as

$$\begin{aligned} \beta &= 8.5573, \\ \tilde{Q} &= \begin{bmatrix} 4.1772 & -2.2475 & 4.2758 \\ -2.2475 & 7.3766 & 1.4573 \\ 4.2758 & 1.4573 & 20.4770 \end{bmatrix}, \\ K &= [8.2490, 6.9729, -0.6591]. \end{aligned} \tag{45}$$

Figure 2 shows the trajectories of φ_l $(l = 1, 2, \ldots, 6)$ for the networked agents, where the full curves with different colors denote the φ_l $(l = 1, 2, \ldots, 6)$ for the six agents and the sequence of blue circles represents the formation center function. We can see from Figure 2 that the curve of each agent converges to that of the formation center function; that is, they reach the consensus with the states of the formation center functions.

Figure 3 depicts the state snapshots of six agents at $t = 0$ s, $t = 8$ s, $t = 9$ s, and $t = 10$ s, where six agents are depicted by blue circles, pink plusses, red hexagrams, black squares, green hexagrams, and indigo x-marks. We can find from Figure 3 that the six agents reach a time-varying regular hexagon around the formation center at different times, which can keep rotating. The above simulation results indicate that network (1) with protocol (2) can reach the leaderless time-varying formation with discontinuous communications.

Example 2. (*leader-follower topologies*). Consider a leader labeled by 6 and five followers labeled by 1–5, whose dynamics and initial states are the same as the leaderless cases. The switching topologies are given in Figure 4, where the dwell time is set as 0.3 s. The communication time unit and noncommunication time unit are $t \in [s, s + 0.8)s$ and $t \in [s + 0.8, s + 1)s$, respectively, and the corresponding parameters are chosen as the same as the leaderless cases.

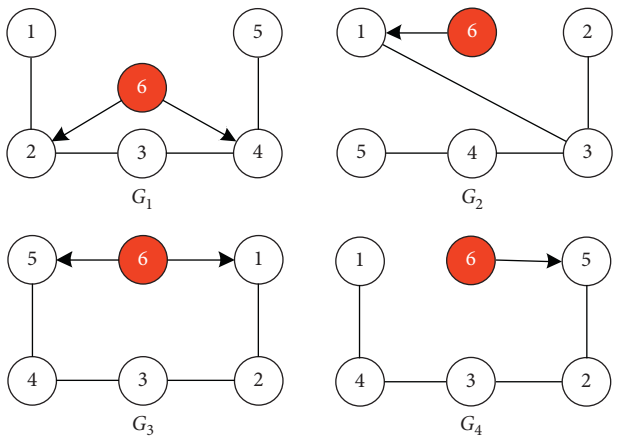


FIGURE 4: Leader-follower switching topologies.

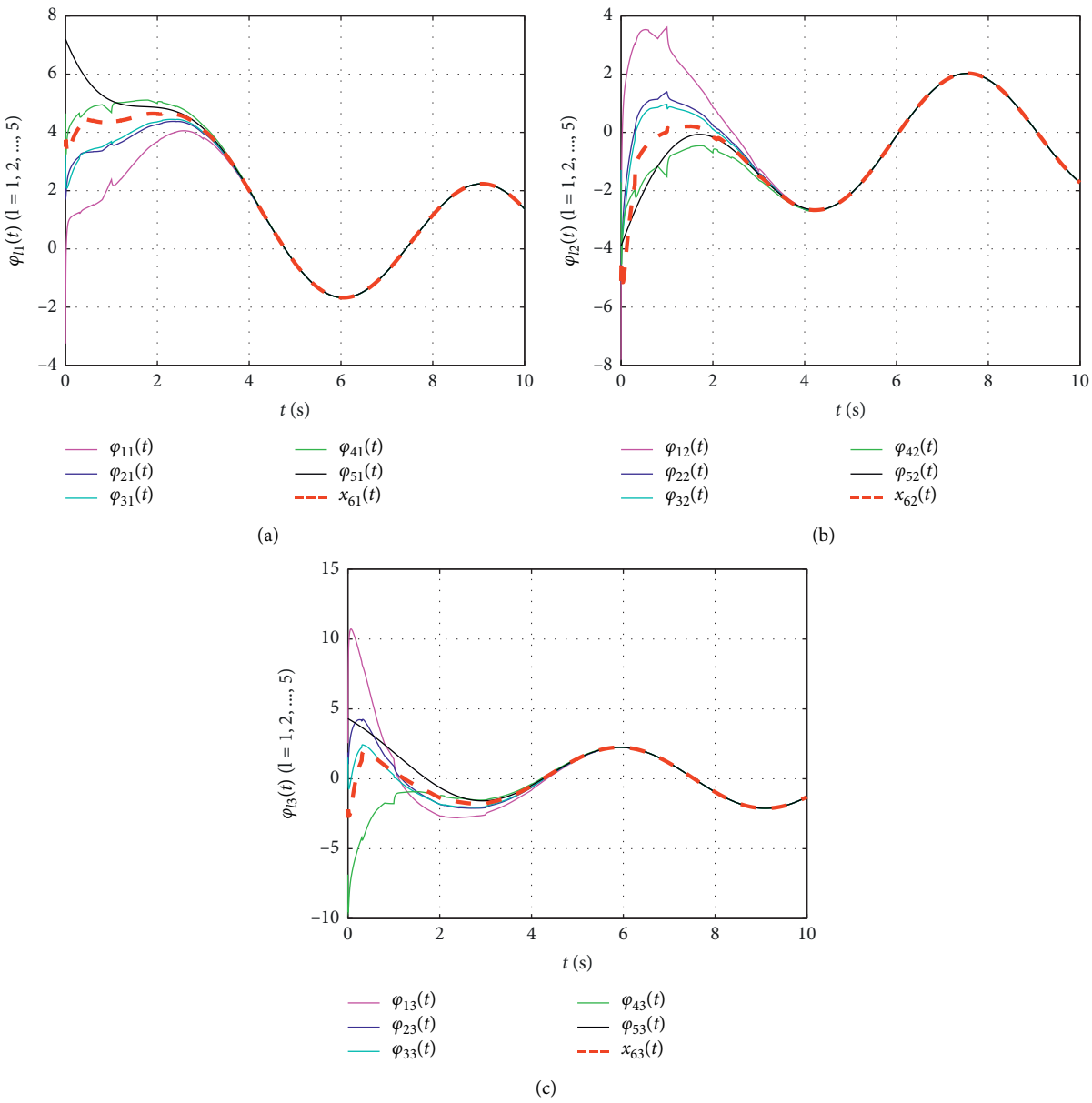


FIGURE 5: Trajectories of φ_l $(l = 1, 2, \ldots, 5)$ and that of the leader.

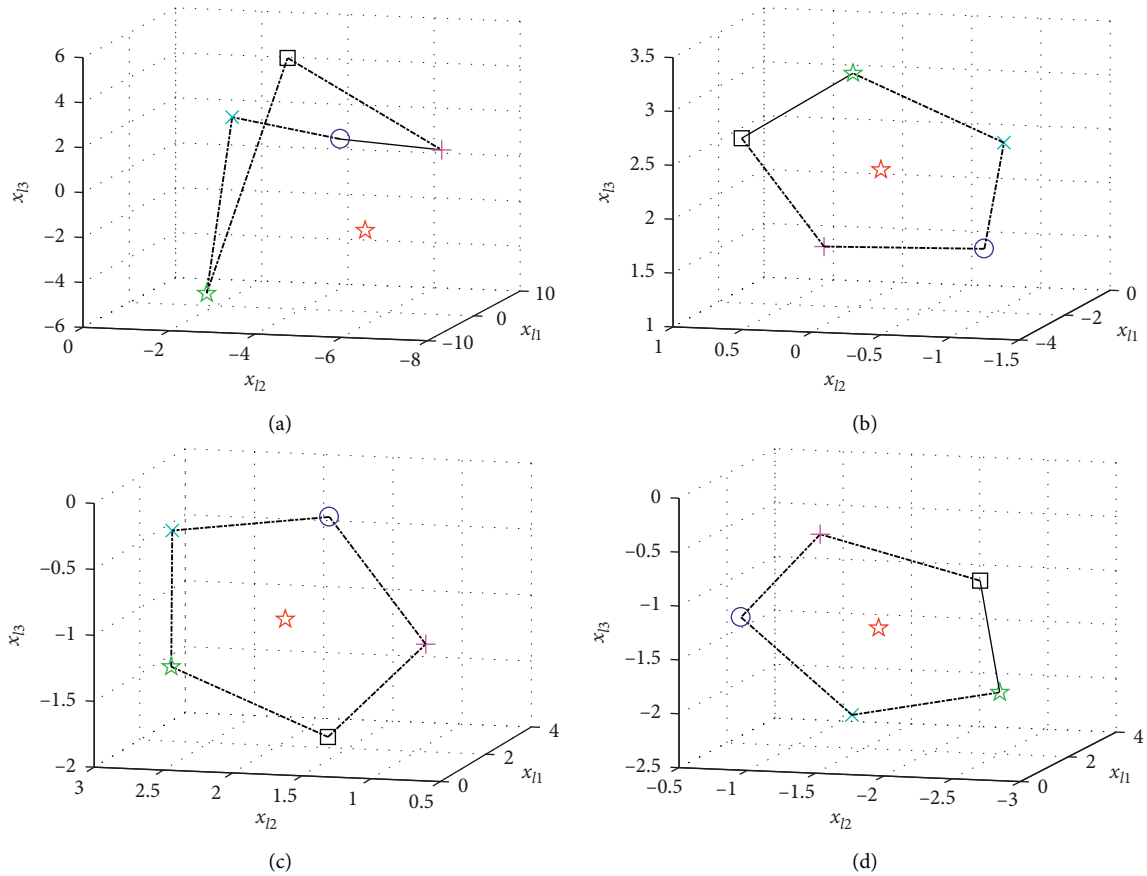


Figure 6: State snapshots of five followers and the leader at different times: (a) $t = 0$ s; (b) $t = 8$ s; (c) $t = 9$ s; (d) $t = 10$ s.

The time-varying formation instruction is chosen as

$$z_l(t) = \begin{bmatrix} \sin\left(t + \dfrac{2(l-1)\pi}{5}\right) \\ \cos\left(t + \dfrac{2(l-1)\pi}{5}\right) \\ -\sin\left(t + \dfrac{2(l-1)\pi}{5}\right) \end{bmatrix}, \quad l = 1, 2, \ldots, 5. \tag{46}$$

According to Theorem 3, the matrix variable and the gain matrix are calculated as

$$\begin{aligned} \xi &= 8.5573, \\ \widetilde{R} &= \begin{bmatrix} 4.1772 & -2.2475 & 4.2758 \\ -2.2475 & 7.3766 & 1.4573 \\ 4.2758 & 1.4573 & 20.4770 \end{bmatrix}, \\ K_{lf} &= [27.2831, 23.0624, -2.1799]. \end{aligned} \tag{47}$$

Figure 5 shows the trajectories of φ_l $(l = 1, 2, \ldots, 5)$ for five followers and that of the leader, where the full curves with different colors stand for the φ_l $(l = 1, 2, \ldots, 5)$ for the five followers and the sequence of red imaginary curves denotes that of the leader. It can be found from Figure 5 that the curves of all followers converge to that of the leader; that is, they can track the trajectory of the leader.

Figure 6 depicts the state snapshots of five followers and the leader at $t = 0$ s, $t = 8$ s, $t = 9$ s, and $t = 10$ s, where they are depicted by blue circles, pink plusses, red hexagrams, black squares, green hexagrams, and indigo x-marks. We can find from Figure 3 that the six agents reach a time-varying regular pentagon and keep rotating around the leader, which means that network (26) with protocol (27) can reach the leader-follower time-varying formation with discontinuous communications.

6. Conclusions

The leaderless and leader-follower time-varying formation design and analysis for networked agents with discontinuous communications were studied. The leaderless time-varying formation control protocol was proposed via the intermittent control strategy, which contains both the communication time unit and the noncommunication time unit. An explicit formulation of the formation center function was determined, which can describe the formation movement trajectory of the networked agents as a whole. Leaderless time-varying formation design and analysis with discontinuous communications were given, where the formation

feasibility conditions and discontinuous communication conditions were constructed to ensure the stability of the closed-loop subnetworks. Moreover, the main results of the leaderless cases were extended to the leader-follower cases, where the trajectory of the formation movement was determined by the state of the leader.

Data Availability

The data used to support this study are included within this article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61867005, 61763040, and 61703411).

References

[1] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.

[2] J. Qu, Z. Ji, C. Lin, and H. Yu, "Fast consensus seeking on networks with antagonistic interactions," *Complexity*, vol. 2018, Article ID 7831317, 15 pages, 2018.

[3] J. Xi, Z. Fan, H. Liu, and T. Zheng, "Guaranteed-cost consensus for multiagent networks with Lipschitz nonlinear dynamics and switching topologies," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 7, pp. 2841–2852, 2018.

[4] J. Sun, Z. Geng, Z. Li, and Z. Ding, "Distributed adaptive consensus disturbance rejection for multi-agent systems on directed graphs," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 202–212, 2018.

[5] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4585–4598, 2020.

[6] Y. Zhang, H. Li, J. Sun, and W. He, "Cooperative adaptive event-triggered control for multiagent systems with actuator failures," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1759–1768, 2019.

[7] L. Consolini, F. Morbidi, D. Prattichizzo, and M. Tosques, "Leader-follower formation control of nonholonomic mobile robots with input constraints," *Automatica*, vol. 44, no. 5, pp. 1343–1349, 2008.

[8] K.-K. Oh and H.-S. Ahn, "Formation control of mobile agents based on distributed position estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 737–742, 2013.

[9] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 2442–2454, 2020.

[10] H. Liu, T. Ma, F. L. Lewis, and Y. Wan, "Robust formation control for multiple quadrotors with nonlinearities and disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1362–1371, 2020.

[11] X. Yang, G. Lin, Y. Liu, F. Nie, and L. Lin, "Fast spectral embedded clustering based on structured graph learning for large-scale hyperspectral image," *IEEE Geoscience and Remote Sensing Letters*, p. 1, 2020 to be published.

[12] R. Lu, X. Yang, X. Jing et al., "Infrared small target detection based on local hypergraph dissimilarity measure," *IEEE Geoscience and Remote Sensing Letters*, p. 1, 2020 to be published.

[13] R. Lu, X. Yang, W. Li, J. Fan, D. Li, and X. Jing, "Robust infrared small target detection via multidirectional derivative-based weighted contrast measure," *IEEE Geoscience and Remote Sensing Letters*, p. 1, 2020 to be published.

[14] N. Cai, M. He, Q. Wu, and M. J. Khan, "On almost controllability of dynamical complex networks with noises," *Journal of Systems Science and Complexity*, vol. 32, no. 4, pp. 1125–1139, 2019.

[15] Z.-Y. Tan, N. Cai, J. Zhou, and S.-G. Zhang, "On performance of peer review for academic journals: analysis based on distributed parallel system," *IEEE Access*, vol. 7, pp. 19024–19032, 2019.

[16] W. Ren, "Consensus strategies for cooperative control of vehicle formations," *IET Control Theory & Applications*, vol. 1, no. 2, pp. 505–512, 2007.

[17] J. Xi, C. Wang, H. Liu, and L. Wang, "Completely distributed guaranteed-performance consensualization for high-order multiagent systems with switching topologies," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1338–1348, 2019.

[18] L. Wang, J. Xi, Z. Yu, and X. Liu, "Limited-budget finite-time average consensus design for multi-agent systems," *IET Control Theory & Applications*, vol. 14, no. 15, pp. 2197–2204, 2020.

[19] J. Xi, M. He, H. Liu, and J. Zheng, "Admissible output consensualization control for singular multi-agent systems with time delays," *Journal of the Franklin Institute*, vol. 353, no. 16, pp. 4074–4090, 2016.

[20] W. Zou, P. Shi, Z. Xiang, and Y. Shi, "Finite-time consensus of second-order switched nonlinear multi-agent systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1757–1762, 2020.

[21] W. Zou, C. K. Ahn, and Z. Xiang, "Fuzzy-approximation-based distributed fault-tolerant consensus for heterogeneous switched nonlinear multiagent systems," *IEEE Transactions on Fuzzy Systems*, p. 1, 2020.

[22] W. Zou, P. Shi, Z. Xiang, and Y. Shi, "Consensus tracking control of switched stochastic nonlinear multiagent systems via event-triggered strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 1036–1045, 2020.

[23] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1465–1476, 2004.

[24] M. Jafarian, E. Vos, C. De Persis, J. Scherpen, and A. van der Schaft, "Disturbance rejection in formation keeping control of nonholonomic wheeled robots," *International Journal of Robust and Nonlinear Control*, vol. 26, no. 15, pp. 3344–3362, 2016.

[25] H. Du, G. Wen, Y. Cheng, Y. He, and R. Jia, "Distributed finite-time cooperative control of multiple high-order nonholonomic mobile robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2998–3006, 2017.

[26] L. Brinon-Arranz, A. Seuret, and C. Canudas-de-Wit, "Cooperative control design for time-varying formations of multi-

agent systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2283–2288, 2014.

[27] R. Rahimi, F. Abdollahi, and K. Naqshi, "Time-varying formation control of a collaborative heterogeneous multi agent system," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1799–1805, 2014.

[28] X. Dong, Y. Zhou, Z. Ren, and Y. S. Zhong, "Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5014–5024, 2016.

[29] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[30] X. Dong and G. Hu, "Time-varying formation tracking for linear multiagent systems with multiple leaders," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3658–3664, 2017.

[31] J. Shao, W. X. Zheng, T.-Z. Huang, and A. N. Bishop, "On leader-follower consensus with switching topologies: an analysis inspired by pigeon hierarchies," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3588–3593, 2018.

[32] R. Wang, "Adaptive output-feedback time-varying formation tracking control for multi-agent systems with switching directed networks," *Journal of the Franklin Institute*, vol. 357, no. 1, pp. 551–568, 2020.

[33] L. Wang, J. Xi, B. Hou, and G. Liu, "Limited-budget consensus design and analysis for multiagent systems with switching topologies and intermittent communications," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, 2021.

[34] J. Sun and Z. Wang, "Consensus of multi-agent systems with intermittent communications via sampling time unit approach," *Neurocomputing*, vol. 397, pp. 149–159, 2020.

[35] W. Qin, Z. Liu, and Z. Chen, "A novel observer-based formation for nonlinear multi-agent systems with time delay and intermittent communication," *Nonlinear Dynamics*, vol. 79, no. 3, pp. 1651–1664, 2015.

[36] C. Godsil and G. Royal, *Algebraic Graph Theory*, Springer, New York, NY, USA, 2001.

Hindawi
Complexity
Volume 2021, Article ID 6625141, 9 pages
https://doi.org/10.1155/2021/6625141

Research Article

Analytical Comparison of Two Emotion Classification Models Based on Convolutional Neural Networks

Huiping Jiang, Demeng Wu, Rui Jiao, and Zongnan Wang

Brain Cognitive Computing Lab, School of Information and Engineering, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Huiping Jiang; jianghp@muc.edu.cn

Received 30 December 2020; Revised 31 January 2021; Accepted 11 February 2021; Published 25 February 2021

Academic Editor: Ning Cai

Electroencephalography (EEG) is the measurement of neuronal activity in different areas of the brain through the use of electrodes. As EEG signal technology has matured over the years, it has been applied in various methods to EEG emotion recognition, most significantly including the use of convolutional neural network (CNN). However, these methods are still not ideal, and shortcomings have been found in the results of some models of EEG feature extraction and classification. In this study, two CNN models were selected for the extraction and classification of preprocessed data, namely, common spatial patterns- (CSP-) CNN and wavelet transform- (WT-) CNN. Using the CSP-CNN, we first used the common space model to reduce dimensionality and then applied the CNN directly to extract and classify the features of the EEG; while, with the WT-CNN model, we used the wavelet transform to extract EEG features, thereafter applying the CNN for classification. The EEG classification results of these two classification models were subsequently analyzed and compared, with the average classification accuracy of the CSP-CNN model found to be 80.56%, and the average classification accuracy of the WT-CNN model measured to 86.90%. Thus, the findings of this study show that the average classification accuracy of the WT-CNN model was 6.34% higher than that of the CSP-CNN.

1. Introduction

An electroencephalogram (EEG) is a record of changes registered on a human or animal scalp, which indicate the electrophysiological activity of brain nerve cells on the cerebral cortex or scalp surface [1]. An EEG captures the spontaneous bioelectric activity of brain cell groups (also known as brain waves) through electrodes and uses potential as the vertical axis and time as the horizontal axis to display the EEG in the form of a curve [2, 3].

The process of emotion recognition based on EEG encompasses the following key steps: emotion induction, EEG signal acquisition, EEG signal preprocessing, EEG feature extraction, emotion pattern learning, and classification [4, 5]. Of these, feature extraction and classification are of particular importance and are the focus of this study.

Feature extraction selects certain feature signals to be used as classification parameters to form characteristic feature vectors [6]. It is a relatively mature technique in machine learning and has developed to include time-domain features, such as mean value, standard deviation, skewness, peak amplitude, variance, skewness, and kurtosis of the EEG signals and frequency-domain features, which transform the time-domain signal into the frequency domain and then extract relevant parameters for analysis. Common features are extracted by Fourier transform, parameter model methods (such as autoregressive (AR), moving average model (MA), autoregressive-moving-average (ARMA), and harmonic signal models). Extraction methods for time-frequency domain features include short-time Fourier transform (STFT) and wavelet transform, while those for nonlinear dynamic characteristics include those based on chaos theory methods, such as Lorenz scatter plot, maximum Lyapunov exponent, correlation dimension, and Hurst exponent. Methods based on information theory include permutation entropy, singular value decomposition entropy, LZC complexity, approximate entropy, and sample entropy [7, 8]. For the extraction of statistical features, statistical methods commonly used in EEG analysis include

probability random analysis, independent component analysis, and principal component analysis, among others [8].

With the rise of artificial intelligence, the convolutional neural network (CNN) has been able to achieve increasingly better results in image and speech, but they are still rarely used in the feature extraction and classification of EEG. Some scholars have tried to directly use the CNN with EEG; however, the accuracy achieved by the two-class classification is only about 50%, and classification effects remain unsatisfactory. When Jie et al. [9] used the CNN alone to image EEG signal classification, for example, their accuracy rate was just 45%. In another study, in which EEG classification of addiction craving was based on the CNN [10], a new matrix was formed for each electrode and then sent to the convolutional neural network to detect craving for addiction, and the accuracy was improved to approximately 70%. This accuracy is slightly higher than that of the classification effect with the mean value as the feature, but it does not compare with the frequency domain feature, and this method varies greatly among different subjects. Reference [9] also tried to use common spatial patterns (CSP) for dimensionality reduction, selecting standardized covariance as a feature to classify the data of motor imagination, and the accuracy rate reached 91.46% [11]. Furthermore, another study, in which the wavelet transform- (WT-) CNN model was proposed and applied to competitive sports thinking data, the accuracy rate reached 88.1%, which is 8.2% higher than traditional WT or support vector machine [12]. In this study, the effect of the CNN on emotion classification is explored by applying the WT-CNN model to the classification of emotion recognition. Moreover, a new CSP-CNN model is proposed, and a comparative analysis of these two methods is performed.

The main research content of this study is, thus, the application of a CNN in EEG emotion classification. After the collection of the EEG data, it was preprocessed by removing ocular and other artifacts and filtering. Thereafter, a CNN was used directly to extract and classify the EEG data after either dimensionality reduction or wavelet transformation. It should be noted that the CSP-CNN in this study is different from that previously published as "Multi-class motor imaging EEG signal classification based on CSP and convolutional neural network algorithms" [11]. Moreover, in the CNN-CSP model in this study, no feature extraction work is performed between the CNN and CSP, such as the identification of standardized covariance or energy. Both of the two emotion recognition models presented in this study were designed and developed by the authors.

2. Related Work

2.1. Cospace Mode and Wavelet Transform. A spatial filter is highly suitable for the collection and processing of EEG signals such as multidimensional signals and data. It can simultaneously utilize the spatial correlation of EEG signals, eliminate signal noise, and realize local cortical nerve activity. Spatial-domain filtering effectively combines time-domain and frequency-domain features, through which better processing results can be achieved [13, 14]. At present, the commonly used spatial filtering techniques in EEG-BCI research include common average reference (CAR), Laplace transform, principal component analysis (PCA), independent component analysis (ICA), and common spatial pattern (CSP), the most widely used approach. The application process of CSP is shown in Figure 1. This spatial filter features an extraction algorithm for two classification tasks, which can extract the spatial distribution components of each category from multichannel brain-computer interface data [15].

A more recently developed transform analysis method, WT, inherited the concept of localization of STFT; however, at the same time, it provides a "time-frequency" window that can change according to frequencies, and is, thus, an ideal tool for signal time-frequency analysis and processing [16].

2.2. Convolutional Neural Network. The CNN has been widely used in the classification of speech and images and has achieved good results. However, there are relatively few studies on their application in EEG, with only minimal reports on their recognition of emotions based on EEG, such as [17], in which the CNN was introduced to EEG emotion recognition, and its application was explored [17]. Since the EEG signal is relatively weak and the extracted feature may not be sufficiently clear for the classification of emotions, we introduced a CNN to develop the feature vector of the EEG signal. Secondary processing and classification are designed to improve the accuracy and robustness of classification [17, 18]. At the same time, we also used the CNN directly after dimensionality reduction in order to improve the EEG characteristics [19], after which the classification results were evaluated.

3. Method

3.1. Feature Extraction Based on CSP. As mentioned, CSP is a commonly used EEG dimensionality reduction method in EEG feature extraction. Its basic principle is to, first, find a space transformation matrix and then transform the EEG to obtain a new matrix [20]. We represented the EEG signals used for classification with a matrix E of $N * T$, where N represents the number of channels for collecting EEG, T represents the number of samples per EEG signal, and T is greater than or equal to N. The normalized covariance matrix is

$$R = \frac{\mathrm{EE}^T}{\mathrm{trace}\left(\mathrm{EE}^T\right)}, \tag{1}$$

where T is the transpose operation, and trace indicates the trace of the matrix during operation. $\overline{R}_1$ and $\overline{R}_2$ are used to represent the spatial covariance matrix of positive and negative emotions, respectively, which is obtained by calculating the mean value of the covariance matrix [21]. Thereafter, the composite matrix of the two covariance matrices can be expressed as

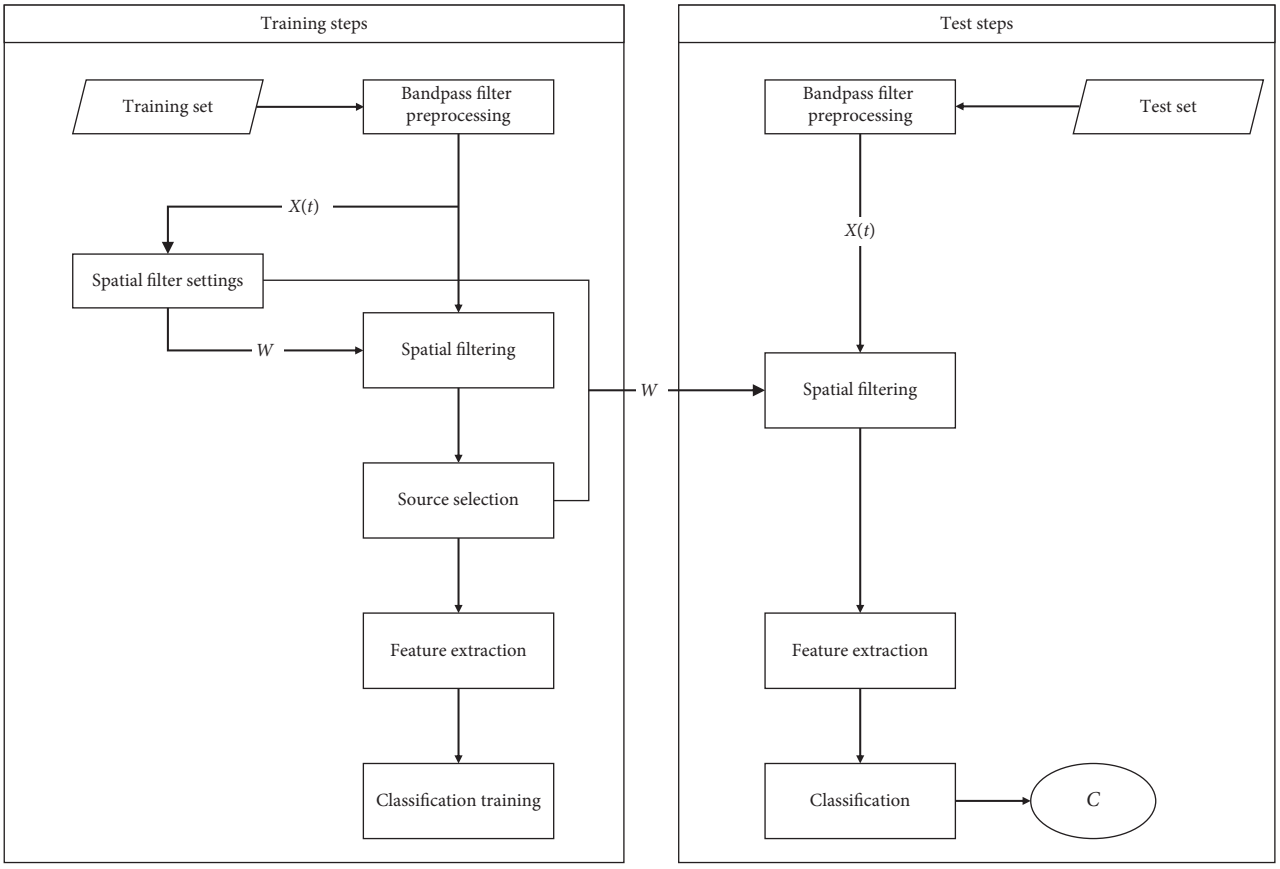


Figure 1: Design and implementation of EEG-BCI based on spatial filtering.

$$R_z = \overline{R}_1 + \overline{R}_2. \tag{2}$$

R_z can be decomposed into

$$R_z = U_z \lambda_z U_Z^T. \tag{3}$$

In the above formula, U_z is the eigenvector of R_z, and λ_z is the diagonal matrix formed by the eigenvalues of R_z. The whitening matrix was calculated as follows

$$P = \lambda_z^{(-1/2)} U_0^T. \tag{4}$$

Thereafter, the calculated whitening matrix was used to transform the average covariance matrix, using the following formula:

$$\begin{aligned} S_1 &= P\overline{R}_1 P^T, \\ S_2 &= P\overline{R}_1 P^T, \end{aligned} \tag{5}$$

S_1 and S_2 had the same feature vectors, namely,

$$\begin{aligned} S_1 &= B\overline{\lambda}_1 B^T, \\ S_2 &= B\overline{\lambda}_2 B^T. \end{aligned} \tag{6}$$

In the above two formulae, λ_1 and λ_2 satisfy $\lambda_1 + \lambda_2 = I$. That is, the largest eigenvalue S_1 corresponds to the smallest eigenvalue S_2. The eigenvalues λ_1 were sorted from large to small, and the eigenvector B was also sorted accordingly to get B^{sort}. The whitened matrix was used to obtain the optimal separation covariance matrix [20, 21], with the first m rows and the last m rows of the transformation matrix B^{sort} used to form a new matrix, B^{2m}. The projection matrix for transforming the original signal is

$$F = \left(B^{\text{sort}}\right)^T P. \tag{7}$$

The transformed matrix is

$$Z = FE. \tag{8}$$

In this study, the data collected after the CSP had reduced the dimensionality of the EEG was changed from the feature value to form the feature vector, as follows:

$$f_p = \frac{\text{var}\left(Z_p\right)}{\sum_1^{2m} \text{var}\left(Z_i\right)}. \tag{9}$$

3.2. Feature Extraction Based on Wavelet Transform (WT). As WT has been widely introduced in numerous reference literatures, this study will only briefly explain the principle of this approach. Every WT has a "mother" wavelet and a "father" wavelet [22], or "parent" wavelet, also termed the "scaling function." Suppose $\psi(t)$ is a square-integrable function, which is $\psi(t) \in L^2(t)$. If the Fourier transform

$\psi(t)$ satisfies the condition (10), then $\psi(t)$ can be used as the mother wavelet.

$$\int_{-\infty}^{+\infty} \frac{|\psi(t)|^2}{|\omega|} \mathrm{d}\omega < +\infty. \tag{10}$$

All of the wavelet series of WT can be obtained by translation scaling of the parent wavelet and mother wavelet. The scaling factor is an integer power of 2, and the magnitude of the translation is related to the scaling factor [22, 23]. The wavelet series are orthonormal, which means that they are not only pairwise orthogonal but also must be normalized. The wavelet series can be expressed as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-a}{a}\right). \tag{11}$$

The expansion formula of the complete wavelet transform is

$$f(t) = \sum_{-\infty}^{+\infty} c_k \varphi(t-k) + \sum_{k=-\infty}^{+\infty} \sum_{j=0}^{+\infty} \mathrm{d}_{j,k} \psi\left(2^j t - k\right). \tag{12}$$

In the above formula, $\varphi(x)$ is the parent wavelet and $\psi(x)$ is the mother wavelet; therefore, c and d can be calculated by selecting the appropriate parent wavelet and mother wavelet, respectively. The approximate formula for wavelet expansion is

$$f(t) = \sum_k \sum_j a_{k,j} \psi_{j,k}^*(t). \tag{13}$$

WT is performed on the signal, which is then decomposed into a sequence of wavelet bases and scale functions. The solution formula is

$$WT_f(a,b) = \frac{1}{\sqrt{a}} \int_R (t) \psi^* f\left(\frac{t-b}{a}\right) \mathrm{d}t, \quad a > 0, \tag{14}$$

where $f(t) \in L^2(R)$. According to the Nyquist sampling theorem, when the sampling frequency fs.max is greater than twice the highest frequency fmax in the signal (fs.max > 2fmax), the digital signal after sampling can completely retain the information contained in the original signal. The collection frequency of electricity, for example, is 250 HZ, so the highest frequency of information retained in the original signal is 125 HZ. In this study, we performed a five-scale WT on the downsampled data (as shown in Figure 2), with each layer decomposing the low-frequency band.

3.3. Feature Classification Based on the CNN. In the base layer of the volume, the size of the filter, that is, the size of the convolution kernel, is usually a $3 * 3$ or $5 * 5$ square matrix. We used $w_{x,y}$ to represent the weight of the filter, b to indicate the bias term of the filter, and f to activate the function. The output of the filter was as follows:

$$g = f\left(\sum_x \sum_y a_{x,y} \times w_{x,y} + b\right). \tag{15}$$

The above formula was used for the forward propagation process of the roll base structure to move from the upper left corner of the current layer of the neural network to the lower right corner through the filter. Each corresponding unit matrix was calculated in the moving process [24]. A pooling layer is often added between the volume base layers, which can effectively reduce both the matrix size and the parameters in the subsequent volume base pooling layer and the fully connected layer. This study uses the maximum pooling layer, the formula for which is

$$g = \mathrm{Max}\left(a_{x,y}\right). \tag{16}$$

Each node of the fully connected layer is connected to all the nodes of the previous layer and is used to integrate the features extracted from the front and to act as a "classifier" in the entire network [25]. In this study, the dropout layer was added after the fully connected layer. The addition of the dropout layer not only reduces error in the training model each time and accelerates the training speed but also effectively prevents the occurrence of overfitting. The last layer of the CNN is the Softmax layer. Its function is to turn the original output result of the neural network into a probability distribution, thus contributing to normalization. Assuming that the output of the original neural network is $y_1, y_2, y_3, \cdots\cdots, y_n$, the output after Softmax regression processing is

$$\mathrm{Softmax}(y)_i = y_i' = \frac{e^{yj}}{\sum_{i=1}^n e^{yj}}. \tag{17}$$

In addition, cross-entropy verification, a method used to describe the distance between two probability distributions, was applied in this study. Given two probability distributions as p and q, the formula for expressing the cross-entropy of p by q is

$$H(p,q) = -\sum_x p(x) \log q(x). \tag{18}$$

Error backpropagation is based on the principle of gradient descent, in which it is only necessary to update in the direction of the negative gradient. Suppose J is the cost function, then the iterative process of each $w_{i,j}$, b_{ij} is

$$\begin{aligned} w_{i,j}^{l+1} &= w_{i,j}^l - \alpha \frac{\partial J}{w_{i,j}^l}, \\ b_{i,j}^{l+1} &= b_{i,j}^l - \alpha \frac{\partial J}{b_{i,j}^l}, \end{aligned} \tag{19}$$

among which α is the learning rate, and $(\partial J / w_{i,j}^l)$ and $(\partial J / b_{i,j}^l)$ and are the partial derivatives of the error.

4. Experiment

4.1. Selection and Design of Stimulus Materials. A total of 210 images were used for the stimulus file, of which 105 were intended to induce positive emotions and the other 105 intended to induce negative emotions. The experiment

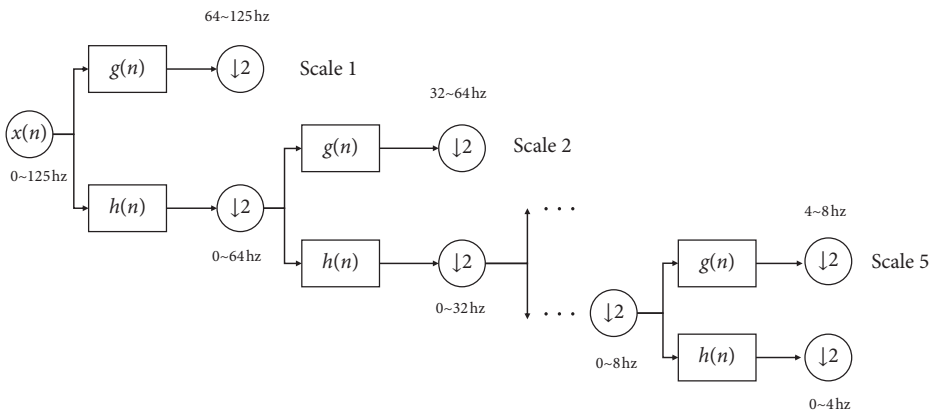


Figure 2: Schematic diagram of wavelet decomposition.

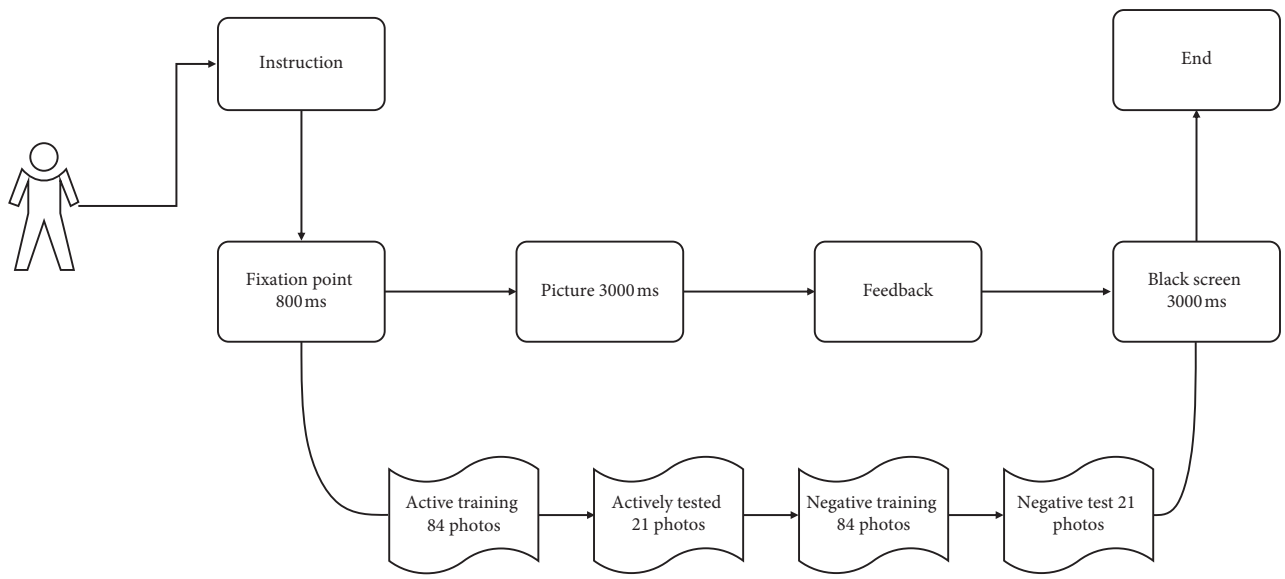


Figure 3: Flow chart of stimulus file.

process is illustrated in Figure 3. At the outset of the experiment, subjects were requested to read the instructions on the screen carefully in order to fully understand the experiment process and details. Once the experiment had been completed, the EEG data samples, comprising the training set of positive and negative emotions, were mixed for the model training, and the EEG data samples of the positive and negative emotion test set were mixed for classification.

4.2. Selection of Mother Wavelet. There are many types of mother wavelets, and therefore, it is essential to select one that is most suitable for the effective extraction of EEG features. The engineering realization of the WT in this study was completed by Matlab. Matlab can complete 15 kinds of female wavelets based on Haar, Daubechies, Biorthogonal, Coiflet, Symlet, Morlet, Mexican hat, Meyer, Gaus, Demeyer, ReverseBior, Cgau Cmor, Fbsp, and Shan, amongst others. At present, there is no unified standard for the selection of wavelet bases, and it is based mainly on the accuracy of classification. In an emotion classification experiment, one of many such experiments previously conducted in our laboratory, the Symlets 8 wavelet (sym8), was found most effective in reducing the original signal, and based on "Video Stimulus EEG Signal Feature Research" [10], its comparative effects were better than other mother wavelets. Therefore, sym8 was selected as the mother wavelet in this assay.

4.3. Acquisition and Preprocessing of EEG Signals. Six students from the Minzu University of China were chosen to be the subjects for the EEG collection. The subjects were aged between 22 and 26, all of them right-handed, healthy, with good sleeping patterns and no brain damage or the history of mental illness. Preprocessing is performed mainly to remove any noise components in the EEG signal and to provide a guarantee for the analysis of the EEG signal characteristics and extraction of the emotional characteristics of the signal. In this study, preprocessing was performed using Scan 4.5 software to remove obstructive artifacts and for digital filtering.

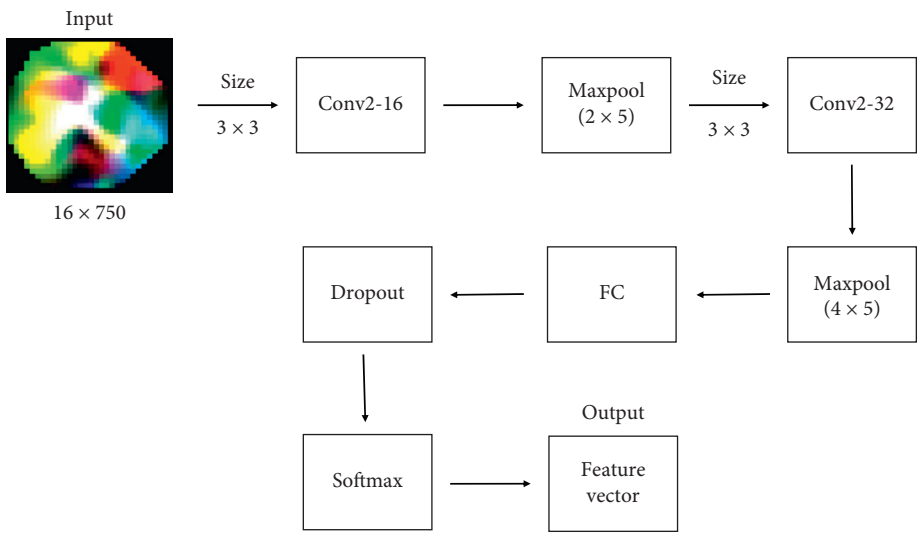


Figure 4: CNN in the CSP-CNN model.

4.4. Training. Two classification models, the WT-CNN and CSP-CNN, were used in this assay to analyze the preprocessed brain. Electric data were used for emotion recognition; whereafter, the classification results were compared and analyzed. The individual differences of an EEG are obvious; therefore, all EEG classifications in study were based on single-person EEG classification.

4.4.1. CSP-CNN. The CSP-CNN was used directly in this study to perform feature extraction on the EEG data after dimensionality reduction. That is to say, the CNN was used to directly perform convolution operations on a $16 * 750$ matrix. After continuous improvement, the CNN model was established, as shown in Figure 4.

This model consists of two volume base layers, two pooling layers, a fully connected layer, a dropout layer, and a Softmax layer. The size of the two base layer convolution kernels in the network is $3 * 3$, the first base layer has 16 convolution kernels, and the second base layer has 32 convolution kernels. The size of the first pooling layer filter is $2 * 5$, while the size of the second pooling layer filter is $4 * 5$, and both pooling layers are the largest pooling layer.

In Table 1, it can be seen that, although the sample dimension was very large since the main parameters of the pooling layer were 12,484, the addition of the pooling layer effectively reduced the number of training parameters and sped up the operation and training of the network. Moreover, this model was able to keep the value of cross-entropy to mostly below 0.01 after training within 40,000 steps. The smaller value of the loss function indicates that the convolutional neural network became more convergent after training. To ensure that the training of the network had reached a stable state, in this model, we used 50,000 steps to mark the final result of the classification. The accuracy of the emotion recognition of the six subjects after the training and classification of this model is presented in Table 2.

The CNN was employed to directly extract and classify the data of the public space model after dimensionality reduction, achieving an average accuracy rate of 80.56%. This result shows that a CSP-CNN can be used to effectively extract features from EEG data.

Table 1: Main parameters of the CNN model used to classify data after CSP dimensionality reduction.

Layer	Output	Parameter
Conv2d_1	(16, 16, 750)	160
MaxPool_1	(16, 8, 150)	0
Conv2d_2	(32, 8, 150)	4640
MaxPool_2	(32, 2, 30)	0
Dense_1	(4)	7684
Sum		12484

Table 2: Classification accuracy of the CSP-CNN model.

Subject	Accuracy
aw	78.57
Lll	90.48
sc	85.71
xcl	80.95
xtc	76.19
dst	71.43

4.4.2. WT-CNN. When building a CNN model, its structure is determined by a variety of parameters. It is necessary to select the appropriate number of layers and to determine the number and size of each layer of the convolution kernel. After constant debugging, the WT-CNN model shown in Figure 5 was established for the classification of wavelet entropy features.

This model is comprised of two volume base layers, a fully connected layer, a dropout layer, and a Softmax layer. No pooling layer was used for dimensionality reduction, in order to optimize information retention and classification accuracy. The parameters of each layer of the WT-CNN model are shown in Table 3, in which it is evident that, while

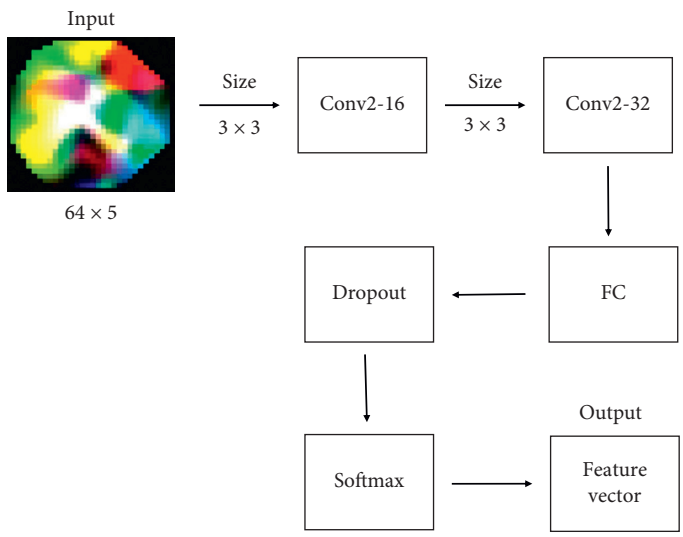


Figure 5: CNN in the WT-CNN model.

Table 3: Main parameters of the WT-CNN model.

Layer	Output	Parameter
Conv2d_1	(16, 64, 5)	160
Conv2d_2	(32, 64, 5)	4640
Dense_1	(4)	40964
Sum		45764

there is no pooling layer, the main training parameters are fewer than 50,000 due to the small training samples, the small number of convolution kernels, and minimal parameters in the fully connected layer. Therefore, the single-step training speed of this model is relatively fast, with training generally completed within 100,000 steps, and the value of the loss function remains below 0.02.

A smaller loss function value indicates that a network is more convergent after training. In order to ensure that the network in this WT-CNN model remained stable after training, 110,000 steps were completed, and the classification result at that point was selected as final. The classification results of the WT-CNN model on the six groups of individual EEG data are shown in Table 4.

The average accuracy of the WT-CNN model was 86.90%. At this stage, a wavelet transform was used for feature extraction. A support vector machine is a more mature method based on EEG emotion recognition. The WT-CNN model's slight improvement in the classification indicates that it is a feasible approach for EEG-based emotion recognition.

Table 5 presents a comparison of the classification results of the two classification models, CSP-CNN and WT-CNN, in which it is evident that a CNN can be used for emotion feature classification, as the classification results were relatively accurate. Furthermore, of the two emotion recognition models, WT was found to be an excellent method for extracting emotional features, and the classification effect achieved by the WT-CNN model was also best.

Table 4: WT-CNN model classification results.

Subject	Accuracy (%)
aw	80.95
ll	92.86
sc	97.62
xcl	73.81
xtc	85.71
dst	90.48

Table 5: Comparison of CSP-CNN and WT-CNN classification results.

Model	Average accuracy (%)	Variance of accuracy
CSP-CNN	80.56	0.00462963
WT-CNN	86.90	0.00742630

5. Conclusion

The study presents research and comparative analysis of the application of two CNN models in EEG-based emotion classification of processed samples.

As the results of previously used methods for feature classification, such standardized variance among others, are generally not sufficiently accurate, and because CNNs are widely used in image feature extraction, it was hypothesized that this approach could be used to achieve more effective outcomes. First, a CSP-CNN model was established in which the CNN extracts and classifies the data after the dimensionality reduction of a cospace model. The average classification accuracy of the CSP-CNN model was 80.56%, and its classification effect was good. In addition, because wavelet transform is known to be an excellent method for extracting emotional features, we established the WT-CNN model. Its average classification accuracy was 86.90%, realizing an improvement of 6.34% compared with the results of the CSP-CNN model. These experiments, thus, showed the feasibility of using wavelet entropy as an effective method for feature extraction.

Analytical comparison of the two approaches shows that the WT-CNN achieves better results than the CSP-CNN for the following reasons: first, wavelet variance is an effective feature quantity based on multiresolution analysis. It can characterize the signal characteristics of different scales, and it does not directly process a large number of wavelet coefficients, but instead mines the data to obtain coimplemented information. Furthermore, wavelet variance has the characteristics of clarity, simple calculation and is not sensitive to noise. Finally, the wavelet transform can greatly reduce or even remove correlations between the different extracted features by selecting the appropriate filter, thereby reducing the difficulty and speed of calculations and improving accuracy.

In the following research, we can try to (1) study the application of the convolutional neural network in multitype emotion recognition, (2) use the convolutional neural network in emotion recognition of large sample EEG data, and (3) investigate whether EEG contains emotional features or look for the timepoint when emotional features appear.

Data Availability

The EEG data used to support the findings of this study are supplied by the National Nature Science Foundation of China under license and so cannot be made freely available. The data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Al-Nafjan, A. Al-Wabil, and M. Hosny, "Classification of human emotions from electroencephalogram (EEG) signal using deep neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 9, 2017.

[2] V. Lafuente, J. M. Gorriz, J. Ramirez, E. Gonzalez, and E. Gonzalez, "P300 brainwave extraction from EEG signals: an unsupervised approach," *Expert Systems with Applications*, vol. 74, pp. 1–10, 2017.

[3] S. SimKok and L. Z. You, "Fast fourier analysis and EEG classification brainwave controlled wheelchair," in *Proceedings Of 2016 2nd International Conference on Control Science and Systems Engineering (ICCSSE),* pp. 20–23, Singapore, July 2016.

[4] Z. Wen, R. Xu, and J. Du, "A novel convolutional neural networks for emotion recognition based on EEG signal," in *Proceedings of the 2017 International Conference On Security, Pattern Analysis, And Cybernetics (Spac),* pp. 672–677, Guangzhou, China, December 2017.

[5] S. Alhagry, F. Aly Aly, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 9, pp. 355–358, 2017.

[6] J. Wang and Y. Li, "Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy," *Applied Energy*, vol. 230, pp. 429–443, 2018.

[7] G. Chen, W. Xie, T. D. Bui, and A. Krzyżak, "Automatic epileptic seizure detection in EEG using nonsubsampled wavelet-fourier features," *Journal of Medical and Biological Engineering*, vol. 37, no. 1, pp. 123–131, 2017.

[8] L. Yuan and J. Cao, "Patients' EEG data analysis via spectrogram image with a convolution neural network," *Intelligent Decision Technologies 2017*, vol. 72, pp. 13–21, 2018.

[9] X. Jie, R. Cao, and L. Li, "Emotion recognition based on the sample entropy of EEG," *Bio-medical Materials and Engineering*, vol. 24, no. 1, pp. 1185–1192, 2014.

[10] J. Liao, Q. Zhong, Y. Zhu, and D. Cai, "Multimodal physiological signal emotion recognition based on convolutional recurrent neural network," *IOP Conference Series: Materials Science and Engineering*, vol. 782, 2020.

[11] N. Wu, *Research on Emotion Classification Based on EEG Signal*, M. S. dissertation, Minzu University of China, Beijing, China, 2013.

[12] B. Zhang, H. Jiang, and L. Dong, "Classification of EEG signal by WT-CNN model in emotion recognition system," in *Proceedings of the IEEE 16th International Conference on Cognitive Informatics and Cognitive Computing (ICCI∗CC)*, New York, NY, USA, July 2017.

[13] A. Ghoshroy, W. Adams, X. Zhang, and D. Ö Güney, "Enhanced superlens imaging with loss-compensating hyperbolic near-field spatial filter," *Optics Letters*, vol. 43, no. 8, pp. 1810–1813, 2018.

[14] Y. Park and W. Chung, "Optimal channel selection using correlation coefficient for CSP based EEG classification," *Institute of Electrical and Electronics Engineers Access*, vol. 8, pp. 111514–111521, 2020.

[15] L. Yi, Y. Huang, X. Zheng, and J. Cheng, "Seismic time-frequency analysis based on entropy-optimized Paul wavelet transform," *Institute of Electrical and Electronics Engineers Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 342–346, 2020.

[16] R. Tibor Schirrmeister, L. Gemein, K. Eggensperger, F. Hutter, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, pp. 5391–5420, 2017.

[17] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *Institute of Electrical and Electronics Engineers Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2020.

[18] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[19] P. J. GarcÃ-a-Laencina, G. RodrÃ-guez-Bermudez, and J. Roca-Dorda, "Exploring dimensionality reduction of EEG features in motor imagery task classification," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5285–5295, 2014.

[20] J. Olias, R. Martin-Clemente, M. A. Sarmiento-Vega, and S. Cruces, "EEG signal processing in MI-BCI applications with improved covariance matrix estimators," *Institute of Electrical and Electronics Engineers Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 895–904, 2019.

[21] M. Mohammad and A. Trounev, "Implicit Riesz wavelets based-method for solving singular fractional integro-differential equations with applications to hematopoietic stem cell modeling," *Chaos, Solitons & Fractals*, vol. 138, p. 109991, 2020.

[22] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Fourier-Bessel series expansion based empirical wavelet transform for

analysis of non-stationary signals," *Digital Signal Processing*, vol. 78, pp. 185–196, 2018.

[23] V. J. Lawhern, A. J Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, pp. 1741–2552, 2018.

[24] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96–107, 2019.

[25] W. Zaperty, T. Kozacki, and M. Kujawińska, "Multi-SLM color holographic 3D display based on RGB spatial filter," *Journal of Display Technology*, vol. 12, no. 12, pp. 1724–1731, 2016.

Hindawi
Complexity
Volume 2021, Article ID 6633878, 10 pages
https://doi.org/10.1155/2021/6633878

Research Article

NPQ-RRT*: An Improved RRT* Approach to Hybrid Path Planning

Zihan Yu and Linying Xiang

School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

Correspondence should be addressed to Linying Xiang; xianglinying@neuq.edu.cn

Received 6 December 2020; Accepted 3 February 2021; Published 17 February 2021

Academic Editor: Jianxiang Xi

In recent years, the path planning of robot has been a hot research direction, and multirobot formation has practical application prospect in our life. This article proposes a hybrid path planning algorithm applied to robot formation. The improved Rapidly Exploring Random Trees algorithm PQ-RRT* with new distance evaluation function is used as a global planning algorithm to generate the initial global path. The determined parent nodes and child nodes are used as the starting points and target points of the local planning algorithm, respectively. The dynamic window approach is used as the local planning algorithm to avoid dynamic obstacles. At the same time, the algorithm restricts the movement of robots inside the formation to avoid internal collisions. The local optimal path is selected by the evaluation function containing the possibility of formation collision. Therefore, multiple mobile robots can quickly and safely reach the global target point in a complex environment with dynamic and static obstacles through the hybrid path planning algorithm. Numerical simulations are given to verify the effectiveness and superiority of the proposed hybrid path planning algorithm.

1. Introduction

With the continuous deepening of network applications, especially the rapid development of the Internet of Things, big data, cloud computing, and edge computing, the integration of information and physical systems has become increasingly close. Also, the connection between the network and human society has become much more closer. Cyber-physical-social system (CPSS) includes system engineering such as embedded environment perception, dynamic analysis of human organization behavior, network communication, and network control. Such CPSSs have functions of computing, communication, precise control, remote collaboration, and autonomy. Technologies such as artificial intelligence, multiagent, and machine learning have been more widely used in the CPSS area [1–4]. As a typical representative of agents, multiple autonomous robots, including unmanned aerial vehicles (UAVs), automatic ground vehicles (AGVs), and unmanned underwater vehicles (UUVs), play an important role in military and civil fields [5–13]. To ensure robots carry out related work in our life efficiently and safely, path planning, which means finding a feasible path without collisions from the starting state to the target state, has been a hot research point in the field of mobile robot applications [14–18]. At present, path planning algorithms can be divided into global path planning algorithms and local path planning algorithms, which mainly include geometric algorithms, artificial potential field algorithms, grid-based search algorithms, and sampling-based algorithms. Among them, the sampling-based algorithm has received extensive attention because of its superior performance in high-dimensional state space. Furthermore, the probability of finding a feasible path in space approaches 1 as the sampling number approaches infinity.

Rapidly Exploring Random Tree (RRT) [19], as a representative of sampling-based path planning, has attracted wide attention of the research community. A large number of improved algorithms for RRT have emerged in the past decade. RRT-connect [20] is a dual-tree RRT algorithm. Two random trees are generated from the start point and the end

point, respectively. However, neither RRT nor RRT-connect considers the path cost; therefore, the optimality of the algorithm cannot be guaranteed. Based on the previous algorithms, the RRT* algorithm [21] was proposed, in which the cost of the path is covered. Also, the steps of selection and rewiring are added. This algorithm obtains progressive optimality, which has become a breakthrough in the development of the Rapidly Exploring Random Trees algorithm. However, the convergence speed of the algorithm has become a new problem. One of the fundamental reasons for the slow convergence speed of RRT* is its global exploration, which does not have a specific direction. To solve this problem, P-RRT* was proposed in [22]. This algorithm incorporates APF into RRT*, and the addition of APF [23] provides a direction for exploration, making P-RRT* converge faster than RRT*. In addition, another algorithm named Quick-RRT* [24] was proposed which uses the triangle inequality to optimize the process of selecting the parent node and connection. Compared with RRT*, it has a faster convergence speed. PQ-RRT* [25] combines P-RRT* and Quick-RRT*, which makes the algorithm generate a better initial solution and can quickly converge to obtain a relatively optimal solution. However, the dynamic obstacles are not considered in PQ-RRT*. Therefore, it can be only used for static path planning and still has some limitations.

In the local path planning algorithm part, the dynamic window approach (DWA) [26] and other algorithms plan the path of the mobile robot through the surrounding information collected by the sensor. However, these algorithms usually do not consider the global map information. Tang et al. proposed a high-speed USV local response obstacle avoidance based on the DWA method [27]. However, it fails to consider the global map information. It is difficult to find the optimal path in the global range using only this kind of algorithm. Based on this kind of situation, various hybrid planning algorithms have been proposed [28–30].

Motivated by the above discussions, we improve the traditional PQ-RRT* algorithm and propose a hybrid planning algorithm—New Potential Quick-RRT* (NPQ-RRT*), which takes the attitude adjustment angle of the robot into consideration and adds the DWA local planning algorithm. Moreover, the algorithm is extended and applied to the path planning problem of multirobot.

The remainder of the paper is organized as follows. The problem description is addressed in Section 2, and the traditional PQ-RRT* algorithm is explained in Section 3. Then, the hybrid planning algorithm NPQ-RRT* is presented in Section 4. In Section 5, simulation results are provided to show the effectiveness of the proposed approach. Finally, the paper is concluded in Section 6.

2. Problem Description

Throughout the paper, R denotes the set of real numbers, N denotes the set of natural numbers, and R^d denotes the space of real d-vectors.

We consider an n-robot system, where each robot moves in the region. Let $X_{\text{obs}} \subset X$ be the obstacle area and the unobstructed area $X_{\text{free}} = (X/X_{\text{obs}})$. The start position and the target position of the robot i are x^i_{init} and x^i_{goal}, respectively.

A trajectory of robot i $(i = 1, 2, \ldots, n)$ is defined as follows: k_i: $[0, \tau_i] \longrightarrow X$, where τ_i is the duration of the trajectory. In addition, $k_i(0) = x^i_{\text{init}}$ and $k_i(\tau_i) = x^i_{\text{goal}}$. The trajectory is obstacle-free if $k_i(t) \in X_{\text{free}}$ for all $t \in [0, \tau_i]$. The cost function $c(\cdot)$ finds the path length in terms of Euclidean distance function.

Trajectory k_i is said to be conflict-free if it is obstacle-free and also keeps robot i at a safety distance $d_s > 0$ from all other robots. $\|x_i(t) - x_j(t)\| > d_s$, where $x_i(t)$ and $x_j(t)$ represent the positions of robots i and j, $t \in [0, \tau]$, $i, j = 1, 2, \ldots, n$, $i \neq j$, and $\tau = \min(\tau_i, \tau_j)$. The total cost of the trajectories is defined as $\sum_{i=1}^{n} c(k_i)$.

Our goal is to find the trajectory set $K = \{k_1, k_2, \ldots, k_n\}$ in which k_i is conflict-free.

3. Related Work

The Rapidly Exploring Random Trees is a sampling-based planning method that builds an undirected graph on a known map through sampling and then finds a relatively optimal path through a search method. The PQ-RRT* is an improved version after adding the target attraction function RGD and the deep parent node search function ***Ancestry***, which can generate a better initial solution and quickly converge to obtain the optimal solution. The pseudocode of the specific algorithm flow is shown in Algorithm 1 [25], where $G = (V, E)$ represents the generated graph.

The related functions are defined as follows:

SampleFree: randomly pick points in the global map. Here, the return value is random point x_{rand}.

RGD: the adjustment function that adjusts the random point x_{rand} under the gravitational force of the target point. Here, the return value is the improved sample x_{prand}. *NearestObstacle* function calculates the distance from x_{prand} to the obstacle space X_{obs}. The parameter z represents the number of iterations. d_{obs} represents the safety distance, and λ represents the step size. The specific pseudocode is shown in Algorithm 2.

Nearest: distance evaluation function. The Euclidean distance function is selected in this situation, which returns the node closest to x_{prand} in the graph $G = (V, E)$ and defines the node as x_{nearest}.

Steer: this function connects two given points. The return value of the function is the segment k between the two points.

CollisionFree: this function detects whether there is a collision with a static obstacle.

Near: given a graph $G = (V, E)$, it returns a set X_{near}, which contains the nodes in the range with x_{prand} as the center and r as the radius.

Ancestry: this function deeply searches the parent node of each point in X_{near}. And it returns the parent node set X_{sparent} of X_{near}. The specific process is as follows: in

Input: $V \longleftarrow x_{\text{init}}$; $E \longleftarrow \varnothing$
Output: $G = (V, E)$
for $i = 1$ to n **do**
  $x_{\text{rand}} \longleftarrow \text{SampleFree}(i)$;
  $x_{\text{prand}} \longleftarrow \text{RGD}(x_{\text{rand}})$;
  $x_{\text{nearest}} \longleftarrow \text{Nearest}(V, x_{\text{prand}})$;
  $k \longleftarrow \text{steer}(x_{\text{nearest}}, x_{\text{prand}})$;
  if CollisionFree(k) **then**
    $X_{\text{near}} \longleftarrow \text{Near}(V, x_{\text{prand}}, r)$;
    $X_{\text{sparent}} \longleftarrow \text{Ancestry}(G, X_{\text{near}})$;
    $(x_{\text{parent}}, k_{\text{parent}}) \longleftarrow \text{ChooseParent}(X_{\text{near}} \bigcup X_{\text{sparent}}, x_{\text{nearest}}, k)$;
    $V \longleftarrow V \bigcup \{x_{\text{prand}}\}$;
    $E \longleftarrow E \bigcup \{x_{\text{parent}}, x_{\text{prand}}\}$;
    $G \longleftarrow \text{Rewire} - \text{PQ} - \text{RRT} * (G, x_{\text{prand}}, X_{\text{near}})$;
  end
end

Algorithm 1: PQ-RRT* [25].

Input: random point x_{rand}, target point x_{goal}
Output: improved point x_{prand}
for $n = 1$ *to* z **do**
  $\overrightarrow{F} = (x_{\text{goal}} - x_{\text{rand}})$
  $d_{\min} \longleftarrow \text{NearestObstacle}(X_{\text{obs}}, x_{\text{rand}})$;
  if$d_{\min} \leq d_{\text{obs}}$**then**
    returnx_{prand};
  else
    $x_{\text{prand}} \longleftarrow x_{\text{rand}} + \lambda(\overrightarrow{F}/|\overrightarrow{F}|)$;
  end
end

Algorithm 2: RGD(x_{rand}).

a given graph $G = (V, E)$, for a node V_1, a natural number $p \in N$, if the depth $p = 0$, it returns $\varnothing$ and otherwise returns the pth parent of V_1.

ChooseParent: compare the cost of each path and then determine the parent node x_{parent} and the path k_{parent}.

*Rewire-PQ-RRT**: a function to generate the final path diagram.

4. An Improved Algorithm NPQ-RRT* for Multirobot

4.1. Overall Ideas. For the robot formation, we divide it into a leader and several followers. When planning the formation path, we first select the leader as the research object and generate a global path through the global planning algorithm. The path is taken as the target path of the robot formation. Each node in the global path is taken as the local starting point and the local target point. Subsequently, for the movement between the local starting point and the local target point of the robot formation, a local path is generated by the local planning algorithm. Compared with the traditional RRT, RRT*, and other algorithms, the global planning algorithm PQ-RRT* has excellent global search capabilities, but it still has limitations: all these algorithms mentioned above discuss fast random expansion search while ignoring the characteristics of the robot to find a feasible path in the global state. However, when applying this algorithm to practical path planning, the attitude adjustment angle of the robot will have an impact on the operation of the algorithm, as shown in Figure 1.

According to the steps in the traditional RRT* series algorithms, the closest point to Position 1 is evaluated according to the Euclidean distance. It is concluded that Position 2 is the closest point to Position 1. However, for a mobile robot, there is an attitude adjustment angle β to Position 2. First, the robot needs to adjust the direction that it faces and then goes to Position 2. For Position 3, it only needs to travel along a straight line. Therefore, the attitude adjustment angle needs to be incorporated into the process of calculating the closest point to Position 1 so as to balance the problems of algorithm convergence time and path smoothness caused by the cumulative rotation angle in the practical application.

At the same time, there are dynamic obstacles in the real environment, which increases the safety risk of mobile robots that perform path planning. The local planning algorithms can better solve these problems and optimize the local path when taking into account dynamic obstacle

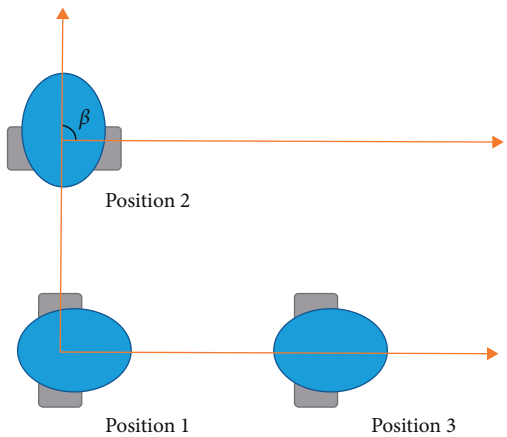


Figure 1: The influence of attitude adjustment angle on the selection of the nearest node.

avoidance. Based on the abovementioned requirements, this paper proposes a hybrid path planning algorithm that considers the attitude adjustment angle.

The local path planner generates a local path for each robot in the formation to follow the local target and avoid obstacles on the local map. By using local targets, the local path planner and the global path planner are combined. The global planner plans the path to the target point in a relatively long period of time for the leader, and the followers follow the leader's path. The local planner updates the trajectory in real time to avoid dynamic and static obstacles. The main idea is shown in Figure 2.

4.2. Specific Steps

4.2.1. The Generation of Global Target Path. The hybrid planning algorithm NPQ-RRT* proposed in this paper improves the problems of the attitude adjustment and dynamic obstacle avoidance. The leader generates the formation target path through the global planning part of the algorithm.

After applying the new distance evaluation function and local programming algorithm to the PQ-RRT* algorithm, the pseudocode of NPQ-RRT* is shown in Algorithm 3. In the pseudocode, most of the function operations are consistent with the operations of related functions in Algorithm 1. The specific operations of the newly proposed distance evaluation function *NewNearest* are given as follows.

The function incorporates the attitude adjustment angle as a new influencing factor into the distance consideration range:

$$q = \varepsilon_v l + \zeta_w \varphi, \tag{1}$$

where ε_v and ζ_w are the evaluation weights of velocity and angular velocity, respectively. Since the local starting point is close to the local target point, l is the Euclidean distance between the nth node and the sample point x_{prand}. φ is the attitude adjustment angle, as shown in Figure 3.

l and φ are defined as follows:

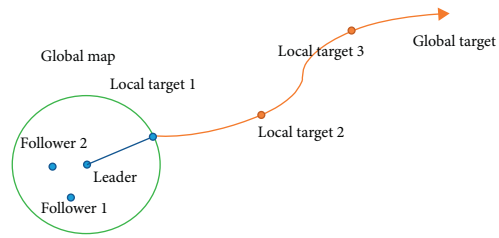


Figure 2: The intersection of global path planner and local path planner.

$$l = \sqrt{(x_v - x_r)^2 + (y_v - y_r)^2},$$
$$\varphi = \phi - \frac{180^\circ \cdot \arctan((y_r - y_v / x_r - x_v))}{\pi}, \tag{2}$$

where (x_v, y_v) and (x_r, y_r) are the coordinates of the node and the modified sample x_{prand}. ϕ is the angle between the orientation of the mobile robot and the X-axis of the global coordinate system, as shown in Figure 3.

Since l and ϕ have different units in the new distance evaluation function, it is of no practical significance to directly add and subtract numerically. The selection of the evaluation weights ε_v and ζ_w becomes the core of the evaluation function. Under the circumstances, the time to reach the destination can be used to measure the length of the distance. When measuring the distance between two points, we assume that the forward speed and the attitude adjustment angular velocity are constant. Therefore, the new distance evaluation function will select the time as the evaluation index. The smaller the value of the function, the shorter the distance between the two points. The evaluation weights ε_v and ζ_w are numerically equal to the reciprocal of the robot's speed and angular velocity at the current moment. Therefore, the new distance evaluation function returns the node that minimizes the function value in the graph $G = (V, E)$ and then defines this node as $x_{\text{new\_nearest}}$.

After the leader completes the steps in the pseudocode, the graph $G = (V, E)$ is obtained which is regarded as the global goal path of the formation.

4.2.2. The Generation of Local Path. For each robot in the formation, take the selected x_{parent} and x_{prand} as the starting point and target point of the local planning, respectively. Using the function *Local* to obtain the LocalPath, the specific operations are given as follows:

Step 1: generate velocity space [26]. Define the ith robot's velocity set V_{mi} as follows:

$$V_{mi} = \left\{ \begin{array}{l} v_i \in [v_{\min}, v_{\max}], \\ \omega_i \in [\omega_{\min}, \omega_{\max}] \end{array} \right\}, \tag{3}$$

where $v_{\min}$ and $v_{\max}$ represent the minimum and maximum velocities that the robot can reach, respectively. $\omega_{\min}$ and $\omega_{\max}$ represent the minimum and maximum angular velocities that the robot can reach, respectively.

Input: $V \longleftarrow x_{\text{init}}; E \longleftarrow \varnothing$
Output: $G = (V, E)$
for $i = 1$ *to* n **do**
  $x_{\text{rand}} \longleftarrow \text{SampleFree}(i)$;
  $x_{\text{prand}} \longleftarrow \text{RGD}(x_{\text{rand}})$;
  $x_{\text{nearest}} \longleftarrow \text{NewNearest}(V, x_{\text{prand}})$;
  $k \longleftarrow \text{steer}(x_{\text{new\_nearest}}, x_{\text{prand}})$;
  if *CollisionFree(k)* **then**
    $X_{\text{near}} \longleftarrow \text{Near}(V, x_{\text{prand}}, r)$;
    $X_{\text{sparent}} \longleftarrow \text{Ancestry}(G, X_{\text{near}})$;
    $(x_{\text{parent}}, k_{\text{parent}}) \longleftarrow \text{ChooseParent}(X_{\text{near}} \bigcup X_{\text{sparent}}, x_{\text{nearest}}, k)$
    LocalPath $\longleftarrow \text{Local}(x_{\text{parent}}, x_{\text{prand}})$
    $V \longleftarrow V \bigcup \{x_{\text{prand}}\}$;
    $E \longleftarrow E \bigcup \{x_{\text{parent}}, x_{\text{prand}}\}$;
    $G \longleftarrow \text{Rewire} - \text{NPQ} - \text{RRT} * (G, x_{\text{prand}}, X_{\text{near}})$;
  end
end

Algorithm 3: NPQ-RRT*.

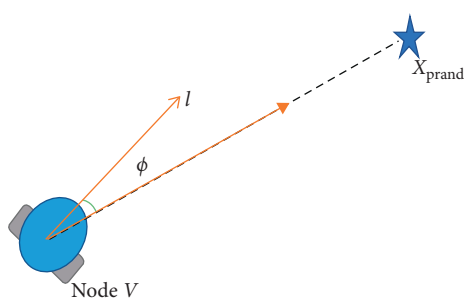


Figure 3: The attitude adjustment of the mobile robot.

Due to the limited torque of the motor, there are maximum acceleration and deceleration limits. The achievable velocity set V_{di} affected by the motor performance is defined as

$$V_{di} = \{(v_i, \omega_i) | v_i \in [v_{ci} - a_b \Delta t, v_{ci} + a_m \Delta t] \cap \omega_i \in \cdot [\omega_{ci} - \alpha_b \Delta t, \omega_{ci} + \alpha_m \Delta t]\}, \tag{4}$$

where v_{ci} and ω_{ci} are the current velocity and angular velocity of the ith mobile robot, respectively. a_b and a_m correspond to the maximum acceleration during deceleration and the maximum acceleration during acceleration. α_b and α_m correspond to the maximum angular acceleration during deceleration and the maximum angular acceleration during acceleration. The opposite direction of the original movement direction is defined as the deceleration direction.

When the robot decelerates with the maximum acceleration at the current velocity, it can be guaranteed to stop before encountering an obstacle, then the velocity is safe. The safe velocity set V_{si} is defined as follows:

$$V_{si} = \left\{(v_i, \omega_i) | v_i \le \sqrt{2 \cdot d(v_i, \omega_i) \cdot a_b} \cap \omega_i \le \sqrt{2 \cdot d(v_i, \omega_i) \cdot \alpha_b}\right\}, \tag{5}$$

where the function $d(v_i, \omega_i)$ represents the distance between the ith robot and the nearest obstacle on the current trajectory.

The final definition of the ith robot's feasible velocity space set is

$$V_{ai} = V_{mi} \cup V_{di} \cup V_{si}. \tag{6}$$

Step 2: obtain the predicted trajectory corresponding to each velocity and avoid collisions within the formation.

First of all, we need to build a model of the robot. Assume that the robot only has two movement modes: forward and rotating. At the current moment, the robot has velocities $v(t)$ and $\omega(t)$. Consider two adjacent moments, as shown in Figure 4. Since the robot's adjacent moment Δt (usually measured by the code disc sampling period in ms) is relatively short, the motion at the two adjacent moments can be regarded as uniform motion. The trajectory of the motion can be regarded as a straight line. Within Δt, the robot moves $v(t)\Delta t$ in the current direction. $\phi(t)$ is the angle between the direction of the mobile robot and the X-axis of the global coordinate system.

In the real coordinate system, the displacement Δx of the robot moving in the X-axis direction of the global coordinate system and the displacement Δy of the Y-axis movement in the global coordinate system are defined as follows:

$$\begin{aligned} \Delta x &= v(t)\Delta t \cos\phi(t), \\ \Delta y &= v(t)\Delta t \sin\phi(t). \end{aligned} \tag{7}$$

As for the trajectory of the robot, t represents the previous moment and $t+1$ represents the current moment. $x(t+1)$, $y(t+1)$, and $\phi(t+1)$ represent the robot's position information and orientation angle information, respectively, which are defined as

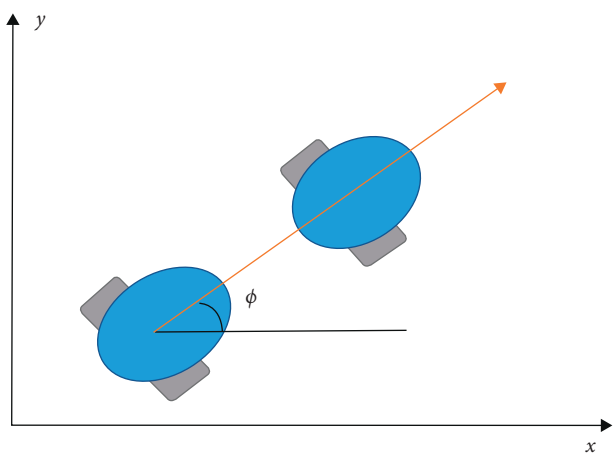


Figure 4: The model of the mobile robot.

$$x(t+1) = x(t) + v(t)\Delta t\cos\phi(t), \tag{8}$$

$$y(t+1) = y(t) + v(t)\Delta t\sin\phi(t), \tag{9}$$

$$\phi(t+1) = \phi(t) + \omega(t)\Delta t. \tag{10}$$

Substituting the velocity space V_{ai} of the robot obtained in the first step into equations (8)–(10), we can obtain the corresponding trajectory expression.

In addition, due to the possibility of collisions between the robots in the robot formation, we make further constraints. For the obtained local predicted trajectories of two adjacent robots, when the predicted trajectories of the two robots have no intersection, they will not collide. On the contrary, when the two predicted trajectories have intersections, they may collide, as shown in Figure 5.

Define the positions of the two robots as $p_i(t)$ and $p_j(t)$, respectively. The meeting point is $m(t)$ and the velocities of the two robots are $v_i(t)$ and $v_j(t)$. The distances Δl_1 and Δl_2 are defined as

$$\begin{aligned} \Delta l_1 &= \left\| p_i(t) - m(t) \right\|, \\ \Delta l_2 &= \left\| p_j(t) - m(t) \right\|. \end{aligned} \tag{11}$$

According to the distance, we can constrain the velocities of the two robots: if $\Delta l_1 \geq \Delta l_2$, then

$$\begin{aligned} v_i(t+1) &= v_i(t) - a_b \cdot \epsilon, \\ v_j(t+1) &= v_j(t) + a_m \cdot \epsilon. \end{aligned} \tag{12}$$

Else, if $\Delta l_1 \leq \Delta l_2$, then

$$\begin{aligned} v_i(t+1) &= v_i(t) + a_m \cdot \epsilon, \\ v_j(t+1) &= v_j(t) - a_b \cdot \epsilon, \end{aligned} \tag{13}$$

where ϵ is a constant that can be set according to the relationship between the velocity and the maximum acceleration.

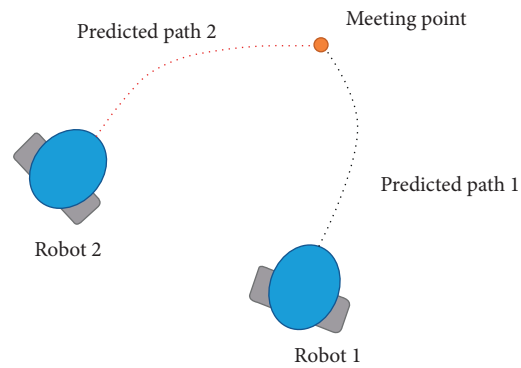


Figure 5: The model of multirobots.

Step 3: select the locally optimal path through the evaluation function. Define the evaluation function of the ith robot's local path as

$$\begin{aligned} G(v_i, \omega_i) = {} & \alpha \cdot m\_\text{head}(v_i, \omega_i) + \delta \cdot m\_d(v_i, \omega_i) \\ & + \gamma \cdot m\_\text{vel}(v_i, \omega_i) - \rho \cdot m\_\text{meet}(v_i, \omega_i), \end{aligned} \tag{14}$$

where $(v_i, \omega_i) \in V_{ai}$ and the variables α, δ, γ and ρ are the initial weights of the function. m is the total number of all trajectories sampled. The function $m\_\text{head}(v_i, \omega_i)$ is used to evaluate the heading score, which is defined as follows:

$$m\_\text{head}(v_i, \omega_i) = \frac{\text{head}(v_i, \omega_i)}{\sum_{c=1}^{m} \text{head}(v_i, \omega_i)}. \tag{15}$$

The function $\text{head}(v_i, \omega_i)$ is defined as

$$\text{head}(v_i, \omega_i) = 180^\circ - \theta_i, \tag{16}$$

where θ_i is the angle between the current direction that the mobile robot is facing and the direction when it reaches the local target point. It can be defined as

$$\theta_i = \phi - \frac{180^\circ \cdot \arctan\left(\left(y_b - y_p / x_b - x_p\right)\right)}{\pi}, \tag{17}$$

where (x_b, y_b) is the coordinate of the local target point in the global map and (x_p, y_p) is the coordinate of the predicted position of the ith robot in the global map.

The function $m\_d(v_i, \omega_i)$ is used to evaluate the safety distance score, which is defined as follows:

$$m\_d(v_i, \omega_i) = \frac{d(v_i, \omega_i)}{\sum_{c=1}^{m} d(v_i, \omega_i)}. \tag{18}$$

The function $d(v_i, \omega_i)$ can be defined as

$$d(v_i, \omega_i) = \min\left\{\text{dist}\left(x_i, s_j\right)\right\}, \tag{19}$$

where x_i is a random point on the trajectory, s_j represents the corresponding obstacle, and *dist* is a function to calculate Euclidean distance. If there are no obstacles on this trajectory, set $d(v_i, \omega_i)$ as a constant.

The function $m\_\text{vel}(v_i, \omega_i)$ is used to evaluate the speed score of the current trajectory. The function $\text{vel}(v_i, \omega_i)$

corresponds to the velocity value of the current trajectory. $m\_\mathrm{vel}(v_i, \omega_i)$ is defined as follows:

$$m\_\mathrm{vel}(v_i, \omega_i) = \frac{\mathrm{vel}(v_i, \omega_i)}{\sum_{c=1}^{m} \mathrm{vel}(v_i, \omega_i)}. \tag{20}$$

The function $m\_\mathrm{meet}(v_i, \omega_i)$ is used to evaluate the score of internal collision probability. The function $\mathrm{meet}(v_i, \omega_i)$ corresponds to the number of intersections with the predicted trajectories of other robots in the formation. $m\_\mathrm{meet}(v_i, \omega_i)$ is defined as follows:

$$m\_\mathrm{meet}(v_i, \omega_i) = \frac{\mathrm{meet}(v_i, \omega_i)}{\sum_{c=1}^{m} \mathrm{meet}(v_i, \omega_i)}. \tag{21}$$

The path with the highest overall score obtained by the evaluation function evaluation is the optimal path. This path is regarded as the local path and is combined with the global planning path to obtain the final path.

5. Simulation

In this section, we discuss the practical effects of the proposed hybrid path planning algorithm NPQ-RRT* in complex environments with dynamic and static obstacles. We perform a comparative simulation experiment through MATLAB on an Intel Core i5 4-core, 8 GB RAM computer.

5.1. Build a Simulation Environment and Set Relevant Parameters. We first construct a 1000×1000 map environment. We set (0,0) as the starting point for global planning and (1000,1000) as the target point for global planning. There are 5 static obstacles, which are represented by green rectangles. The obstacles are distributed in this environment, as shown in Figure 6.

After the command to start path planning is issued, additional obstacles (not overlapping with the current mobile robot position) will be randomly generated as dynamic obstacles in the environment at any given time, as shown in Figure 7.

In this simulation, the relevant parameters of the NPQ-RRT* hybrid path planning algorithm are set as follows: in the *RGD* function, $\lambda = d_{\mathrm{obs}} = 1$ and $z = 80$. In the *ChooseParent*function, $d = 2$. In the *Local* function, the initial weights α, δ, γ, and ρ of the evaluation function are set to 0.05, 0.2, 0.1, and 0.1, respectively. The maximum speed of the mobile robot is 1 m/s, the maximum angular speed is 0.5 rad/s, and the acceleration range is $[-1, 1]$. The angular acceleration range is $[-0.5, 0.5]$, the velocity resolution is 5 m/s, and the angular velocity resolution is 6 rad/s.

5.2. The Simulation Design and Results

5.2.1. Group 1. To verify the performance of using the improved distance evaluation function in the path planning process of mobile robots, we conduct 10 sets of comparative tests without considering dynamic obstacles. The original PQ-RRT* algorithm and our proposed NPQ-RRT* perform the same path planning missions. The sum of the changes in

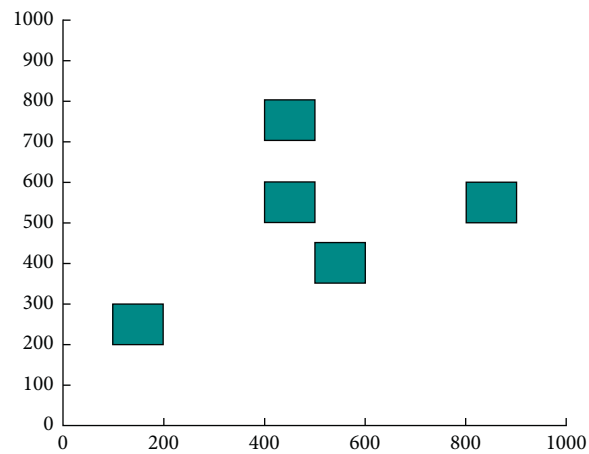


Figure 6: The global environment of simulation.

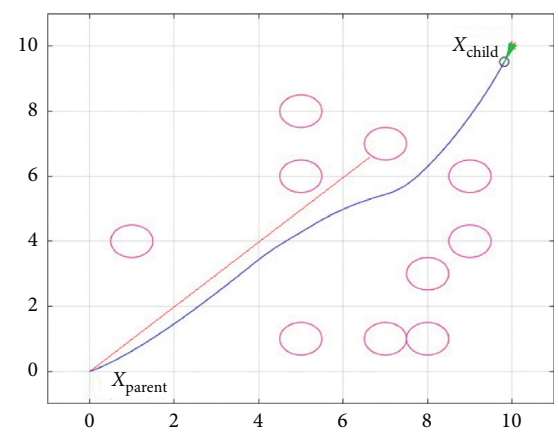


Figure 7: The intersection of global path planner and local path planner.

the heading angle between the nodes makes up the entire adjustment angles. The adjustment angular velocity $\omega = 0.5$ rad/s is used for obtaining the posture adjustment time. After conducting 10 groups of calculations, the average results in Table 1 are obtained.

Figures 8 and 9 show the results of path planning for NPQ-RRT* and PQ-RRT*, respectively. It can be seen that the improved algorithm's robot attitude adjustment time is shortened, the algorithm's running speed is sharply accelerated, and the resulting path is smoother than the original PQ-RRT*.

5.2.2. Group 2. To verify that the hybrid path planning algorithm NPQ-RRT* after adding local planning has better dynamic obstacle avoidance performance than the original PQ-RRT* algorithm, we carry out the following comparative test in consideration of obstacle dynamic obstacles, as shown in Figure 7.

The confirmed x_{prand} and x_{parent} are used as the target point and starting point of local planning. The original algorithm PQ-RRT* lacks a local path planning algorithm; that is, it moves along the line of x_{child} and x_{parent}

Table 1: Comparing the time of attitude adjustment.

Number	PQ-RRT*	NPQ-RRT*
1	65.971	59.311
2	56.567	50.889
3	57.288	45.499
4	63.755	47.139
5	73.209	62.714
6	67.423	41.650
7	59.422	64.298
8	64.478	52.669
9	71.492	49.184
10	68.424	51.770
Average	64.703	52.512

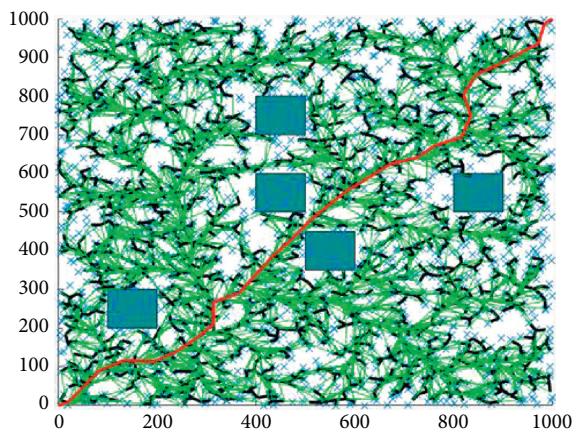


Figure 8: The path planned by NPQ-RRT*.

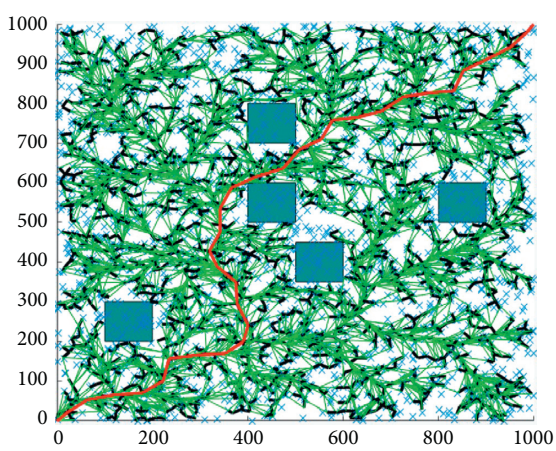


Figure 9: The path planned by PQ-RRT*.

(represented by the red line in Figure 7). It is unable to reach the target due to the existence of dynamic obstacles, causing path planning to fail. After using NPQ-RRT* with local planning algorithm, the mobile robot successfully avoids obstacles and reaches the local target point (the blue line in Figure 7).

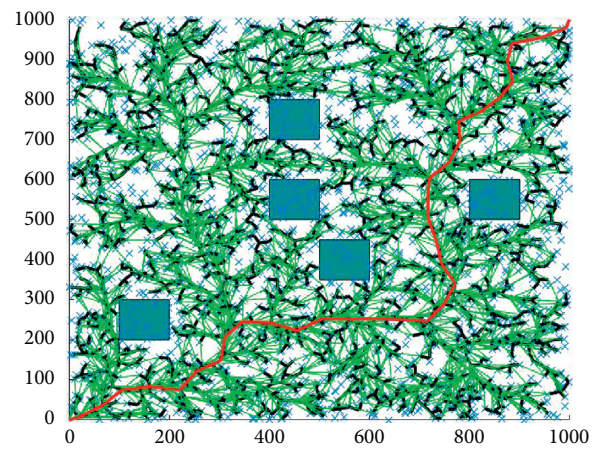


Figure 10: The path planned by hybrid RRT.

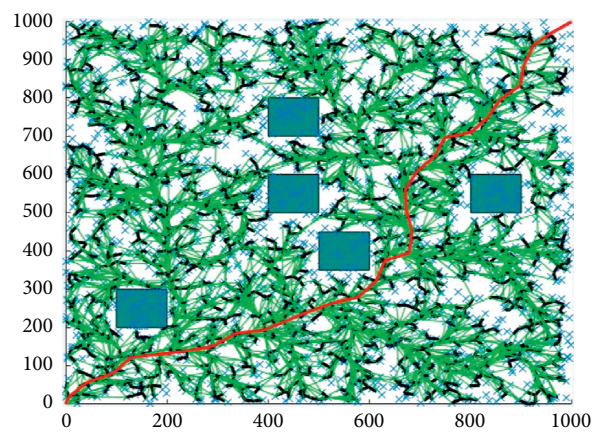


Figure 11: The path planned by NPQ-RRT*.

5.2.3. Group 3. To compare the performance of the common RRT mixed planning algorithm and NPQ-RRT*, we perform experiments in the same obstacle environment. The results are obtained, as shown in Figures 10 and 11.

It can be seen that NPQ-RRT* can obtain a smoother path with a shorter overall length and better overall performance compared to the ordinary RRT mixed planning algorithm.

5.2.4. Group 4. To test the performance of the algorithm when applied to multirobot path planning, we take three robots as an example to perform the following test. Among them, the small black circle represents the robot. The green area represents the detection range of the robotic lidar. The big red circle represents obstacles randomly generated on the map. The blue lines represent the local paths generated by each robot. We use confirmed x_{prand} and x_{parent} as the target point and starting point of the formation planning. Obstacles are randomly generated on the map, and the robot formation moves from the starting point to the target point at the same time. For randomly generated obstacles, the three robots use local planning algorithms to plan their paths

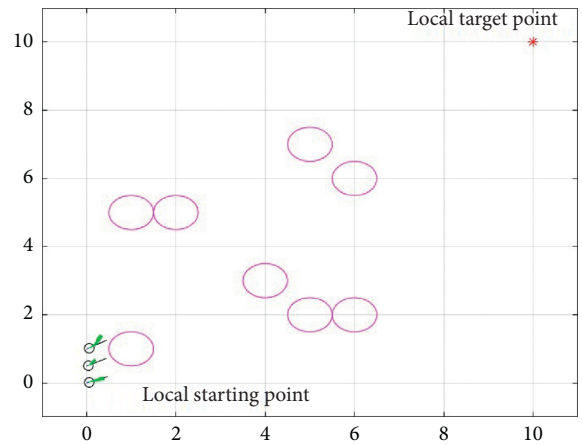


Figure 12: The initial states of the robots.

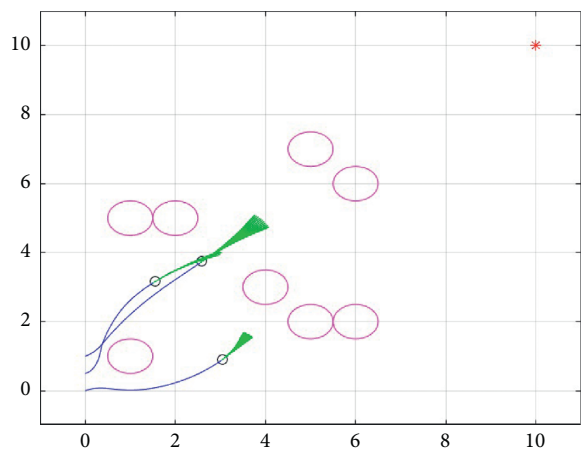


Figure 13: The positions of the robots: $t = 10$s.

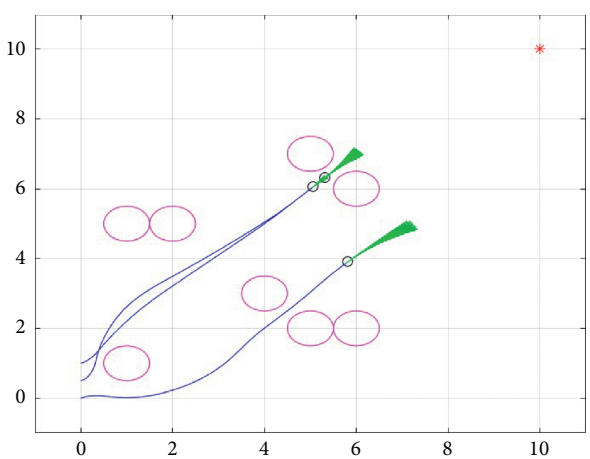


Figure 14: The positions of the robots: $t = 15$s.

to generate obstacle-free paths. At the same time, each robot in the robot formation adopts an avoidance strategy to maintain a safe distance to avoid collisions in the formation. The result is shown in Figures 12–15.

Through the simulation results, we can find that the robot formation completes the obstacle avoidance to the target point, which verifies the feasibility of the proposed algorithm applied to the robot formation.

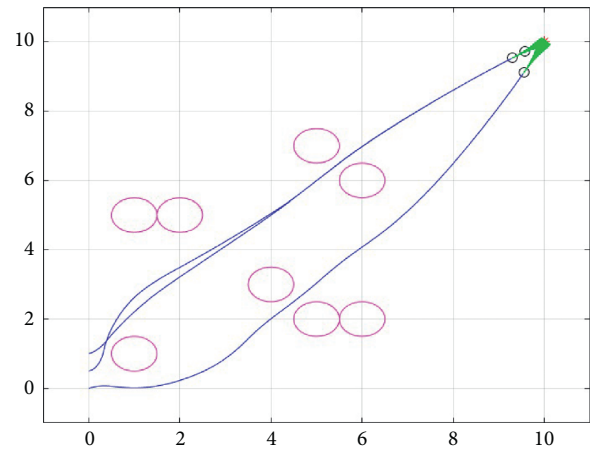


Figure 15: The positions of the robots: $t = 20$s.

Based on the above simulation results, it can be found that NPQ-RRT* has better dynamic obstacle avoidance ability and can effectively shorten the attitude adjustment time and get a smoother path. In addition, this algorithm can also obtain ideal results when it is applied to the robot formation path planning.

6. Conclusions

This paper proposes a hybrid path planning algorithm NPQ-RRT*, which studies the path planning of multi-robot in an environment with dynamic and static obstacles. NPQ-RRT* chooses the improved version of the Rapidly Exploring Random Trees algorithm PQ-RRT* as the global planning algorithm. Combined with the attitude adjustment angle of the mobile robot, we propose a new distance evaluation function, which optimizes the selection of the nearest node. After the parent node and the child node are identified, the local planning step is added. The parent node and child node are used as the local starting point and the local target point to generate a local path avoiding dynamic obstacles. The global path is obtained by tracking the local target point. At the same time, the algorithm optimizes the potential collisions within the robot formation to ensure the safety of the robots. Also, the potential collision possibility in the formation is added as a new evaluation index into the evaluation function to select the optimal path. The simulation results show that compared with PQ-RRT*, the hybrid path planning algorithm NPQ-RRT* has better dynamic obstacle avoidance ability. Furthermore, it can get a relatively better path compared with the ordinary RRT hybrid planning algorithm. When applied to the path planning of robot formation, it can effectively shorten the attitude adjustment time and obtain a smoother path.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grant no. 61973064, the Natural Science Foundation of Hebei Province of China under grant no. F2019501126, the Natural Science Foundation of Liaoning Province of China under grant no. 2020-KF-11-03, and the Fundamental Research Funds for the Central Universities under grant no. N182304013.

References

[1] S. M. M. Rahman, "Cyber-physical-social system between a humanoid robot and a virtual human through a shared platform for adaptive agent ecology," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, 2017.

[2] L. Cheng, T. Yu, X. Zhang, and B. Yang, "Parallel cyber-physical-social systems based smart energy robotic dispatcher and knowledge automation: concepts, architectures and challenges," *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 54–64, 2018.

[3] Z. Zhu, Y. Wen, Z. Zhang, Z. Yan, S. Huang, and X. Xu, "Accurate position estimation of mobile robot based on cyber-physical-social systems (CPSS)," *IEEE Access*, vol. 8, pp. 56359–56370, 2020.

[4] H. Tang, L. Li, and N. Xiao, "Smooth sensor motion planning for robotic cyber physical social sensing (CPSS)," *Sensors*, vol. 17, no. 2, p. 393, 2017.

[5] C. Yang, G. Ganesh, S. Haddadin, S. Parusel, A. Albu-Schaeffer, and E. Burdet, "Human-like adaptation of force and impedance in stable and unstable interactions," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 918–930, 2011.

[6] F. Chen, X. Chen, L. Xiang, and W. Ren, "Distributed economic dispatch via a predictive scheme: heterogeneous delays and privacy preservation," *Automatica*, vol. 123, 2020.

[7] F. Chen and W. Ren, "Sign projected gradient flow: a continuous-time approach to convex optimization with linear equality constraints," *Automatica*, vol. 120, 2020.

[8] Z. Li, P. Y. Tao, S. S. Ge, M. Adams, and W. S. Wijesoma, "Robust adaptive control of cooperating mobile manipulators with relative motion," *IEEE Transactions on Systems Man & Cybernetics Part B*, vol. 39, no. 1, pp. 103–116, 2009.

[9] M. Chen, S. S. Ge, and B. Ren, "Robust attitude control of helicopters with actuator dynamics using neural networks," *IET Control Theory & Applications*, vol. 4, no. 12, pp. 2837–2854, 2010.

[10] Z. Li, X. Cao, and N. Ding, "Adaptive fuzzy control for synchronization of nonlinear teleoperators with stochastic time-varying communication delays," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 4, pp. 745–757, 2011.

[11] R. Cui, J. Guo, and B. Gao, "Game theory-based negotiation for multiple robots task allocation," *Robotica*, vol. 31, no. 6, pp. 923–934, 2013.

[12] F. Chen and J. Chen, "Minimum-energy distributed consensus control of multiagent systems: a network approximation approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 1144–1159, 2020.

[13] L. Xiang, F. Chen, W. Ren, and G. Chen, "Advances in network controllability," *IEEE Circuits & Systems Magazine*, vol. 19, no. 2, pp. 8–32, 2019.

[14] Y. Huang, Z. Li, Y. Jiang, and L. Cheng, "Cooperative path planning for multiple mobile robots via hafsa and an expansion logic strategy," *Applied Sciences*, vol. 9, no. 4, 2019.

[15] A. Majeed and S. Lee, "A new coverage flight path planning algorithm based on footprint sweep fitting for unmanned aerial vehicle navigation in urban environments," *Applied Sciences*, vol. 9, no. 7, p. 1470, 2019.

[16] H. Y. Lee, H. Shin, and J. Chae, "Path planning for mobile agents using a genetic algorithm with a direction guided factor," *Electronics*, vol. 7, no. 10, 2018.

[17] J. Wang, J. Xu, and R. Tai, "A bi-level probabilistic path planning algorithm for multiple robots with motion uncertainty," *Complexity*, vol. 2020, Article ID 9207324, , 2020.

[18] P. Cui, W. Yan, R. Cui, and J. Yu, "Smooth path planning for robot docking in unknown environment with obstacles," *Complexity*, vol. 2018, Article ID 4359036, 17 pages, 2018.

[19] S. M. LaValle, "Rapidly-exploring random trees: a new tool for path planning," 1998.

[20] J. J. K. Jr and S. M. Lavalle, "RRT-Connect: An efficient approach to single-query path planning," in *Proceedings of the 2000 IEEE International Conference on Robotics and Automation, ICRA 2000*, San Francisco, CA, USA, April 2000.

[21] A. T. Perez, S. Karaman, A. C. Shkolnik, E. Frazzoli, and M. R. Walter, "Asymptotically-optimal path planning for manipulation using incremental sampling-based algorithms," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems*, Tokyo, Japan, November 2013.

[22] A. H. Qureshi and Y. Ayaz, "Potential functions based sampling heuristic for optimal path planning," *Autonomous Robots*, vol. 40, no. 6, pp. 1079–1093, 2016.

[23] A. Azzabi and K. Nouri, "An advanced potential field method proposed for mobile robot path planning," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 11, pp. 3132–3144, 2019.

[24] I.-B. Jeong, S.-J. Lee, and J.-H. Kim, "Quick-RRT*: triangular inequality-based implementation of RRT* with improved initial solution and convergence rate," *Expert Systems with Applications*, vol. 123, pp. 82–90, 2019.

[25] Y. Li, W. Wei, Y. Gao, D. Wang, and Z. Fan, "PQ-RRT*: an improved path planning algorithm for mobile robots," *Expert Systems with Applications*, vol. 45, Article ID 113425, 2020.

[26] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 2002.

[27] P. Tang, R. Zhang, D. Liu, L. Huang, G. Liu, and T. Deng, "Local reactive obstacle avoidance approach for high-speed unmanned surface vehicle," *Ocean Engineering*, vol. 106, pp. 128–140, 2015.

[28] H. Yang, J. Qi, Y. Miao, H. Sun, and J. Li, "A new robot navigation algorithm based on a double-layer ant algorithm and trajectory optimization," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, pp. 8557–8566, 2019.

[29] Z. Chen, Y. Zhang, Y. Zhang, Y. Nie, J. Tang, and S. Zhu, "A hybrid path planning algorithm for unmanned surface vehicles in complex environment with dynamic obstacles," *IEEE Access*, vol. 7, pp. 126439–126449, 2019.

[30] G. Huskić, S. Buck, and A. Zell, "Gerona: generic robot navigation," *Journal of Intelligent & Robotic Systems*, vol. 95, no. 2, 2018.

Hindawi
Complexity
Volume 2021, Article ID 6680509, 10 pages
https://doi.org/10.1155/2021/6680509

Research Article

Dynamic Warping Network for Semantic Video Segmentation

Jiangyun Li,[1,2] **Yikai Zhao**,[1] **Xingjian He**,[3,4] **Xinxin Zhu**,[3,4] **and Jing Liu**[4]

[1]*School of Automation & Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[2]*Shunde Graduate School of University of Science and Technology Beijing, Foshan 528300, China*
[3]*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100083, China*
[4]*School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100083, China*

Correspondence should be addressed to Jiangyun Li; leejy@ustb.edu.cn

Received 6 December 2020; Revised 3 January 2021; Accepted 24 January 2021; Published 8 February 2021

Academic Editor: Ning Cai

A major challenge for semantic video segmentation is how to exploit the spatiotemporal information and produce consistent results for a video sequence. Many previous works utilize the precomputed optical flow to warp the feature maps across adjacent frames. However, the imprecise optical flow and the warping operation without any learnable parameters may not achieve accurate feature warping and only bring a slight improvement. In this paper, we propose a novel framework named Dynamic Warping Network (DWNet) to adaptively warp the interframe features for improving the accuracy of warping-based models. Firstly, we design a flow refinement module (FRM) to optimize the precomputed optical flow. Then, we propose a flow-guided convolution (FG-Conv) to achieve the adaptive feature warping based on the refined optical flow. Furthermore, we introduce the temporal consistency loss including the feature consistency loss and prediction consistency loss to explicitly supervise the warped features instead of simple feature propagation and fusion, which guarantees the temporal consistency of video segmentation. Note that our DWNet adopts extra constraints to improve the temporal consistency in the training phase, while no additional calculation and postprocessing are required during inference. Extensive experiments show that our DWNet can achieve consistent improvement over various strong baselines and achieves state-of-the-art accuracy on the Cityscapes and CamVid benchmark datasets.

1. Introduction

Semantic segmentation aims to assign a specific semantic label to each pixel for a given image. In recent years, the models based on deep learning [1–5] have brought the performance of the task to a new level. However, most existing methods are only designed for parsing images and may produce inconsistent results to video frames, due to lack of temporal information.

To address the problem, many methods tend to incorporate temporal information of the video to improve the accuracy of video segmentation. And optical flow, which encodes the temporal consistency across frames in the video, has been widely used for semantic video segmentation. Gaddel et al. [6] propose to combine the features wrapped from previous frames with optical flow and those from the current frame to enhance the features. Studies of [7–9] use feature warping for acceleration.

However, there are two main problems with existing warping-based methods. Firstly, the optical flow obtained by the traditional algorithms or optical flow estimation networks [10–12] cannot accurately estimate the motion of all pixels across adjacent frames. Second, the warping operation adopted by previous methods [6, 7, 13] is implemented with standard bilinear interpolation and does not contain any learnable parameters. Therefore, warping features relying on the imprecise optical flow may result in feature misalignment between the warped features and expected ones. TWNet [9] introduces a correction stage after warping to refine the warped features. However, the method has some limitations, because it needs motion vectors and residuals in the compressed video according to a specific compression standard.

In this paper, we propose a novel framework named Dynamic Warping Network (DWNet) to adaptively warp the interframe features for improving the accuracy of

warping-based models. First, we design a flow refinement module (FRM) to optimize the precomputed optical flow and produce more accurate pixel displacement for every pixel location. Besides, we propose a flow-guided convolution (FG-Conv) to achieve the adaptive feature alignment based on the refined optical flow instead of the original warping operation. Furthermore, we introduce the temporal consistency loss including the feature consistency loss and prediction consistency loss to explicitly supervise the warped features and guarantee the temporal consistency of video segmentation, as shown in Figure 1. Our DWNet adopts extra constraints to improve the temporal consistency instead of simple feature fusion and feature propagation [6, 7], which makes the network explicitly model the temporal consistency of the video in the training phase. And, in the inference phase, the optical flow network, the flow refinement module, and the flow-guided convolution can be removed. Hence, the final network can be regarded as a semantic image segmentation network with no post-processing during inference.

We evaluate our DWNet on two semantic video segmentation benchmarks: Cityscapes and CamVid. Extensive experiments show that our DWNet can significantly outperform existing warping-based methods and achieve state-of-the-art accuracy on the two benchmark datasets. In particular, our DWNet can achieve consistent improvement over various strong baselines, which demonstrates the generalization ability of our method.

To conclude, our main contributions are five-fold:

(i) We propose a novel framework named Dynamic Warping Network (DWNet) to adaptively warp the interframe features

(ii) We design a flow refinement module (FRM) to optimize the optical flow and propose a flow-guided convolution (FG-Conv) to adaptively align features across adjacent frames according to the refined optical flow

(iii) We explicitly model the temporal consistency of the video and introduce the temporal consistency loss to supervise the warped features

(iv) Our DWNet needs no additional parameters and calculation during inference because the optical flow network, the flow refinement module, and the flow-guided convolution can be removed in the inference phase

(v) The experimental results demonstrate that our DWNet can outperform previous warping-based methods and achieve state-of-the-art accuracy on the Cityscapes and CamVid datasets

2. Related Work

2.1. Semantic Video Segmentation. Semantic video segmentation aims to carry out dense labeling for all pixels in each frame of a video sequence. Compared with semantic image segmentation, semantic video segmentation needs to focus more on the temporal consistency of consecutive frames and produces a more consistent interframe prediction. Therefore, many works tend to incorporate temporal information of the video to improve the video segmentation accuracy, including optical flow-based feature warping [6, 8, 9, 13–17], propagation-based [18, 19], LSTM-based [15, 20], 3D CNN-based method [21], and the weakly supervised method [22]. And optical flow, which encodes the temporal consistency across frames in the video, has been most widely used for semantic video segmentation. The optical flow-based methods first compute the optical flow between the current frame and the previous frame and then enhance the features of the current frame by warping the features of the previous frame or utilize the warped features from the keyframe as the features of the current frame for acceleration. Despite its relative strength, the optical flow-based feature warping contains two main problems as discussed above. TWNet [9] and DMNet [23] propose to correct the warped features by utilizing the postprocessing, which only brings a slight improvement. To our best knowledge, we are the first to directly optimize the warping operation and propose the learnable dynamic warping operation instead of the original one.

2.2. Dynamic Convolution. The study [24] proposes dynamic filters or kernels to generate context-aware filters which are adaptive to the input and are predicted by the network. Many works [25, 26] have adopted the predicted dynamic filters to obtain better feature representations. Deformable convolution [27, 28] utilizes the input features to generate different offsets and weights for each sample position. Motivated by deformable convolution, we observe that the optical flow can be regarded as the offset and we can utilize the offset to adaptively align interframe features. Different from the deformable convolution whose offsets are generated by the input features, we utilize the flow refinement module to optimize the optical flow and obtain more accurate pixel displacement. Furthermore, we propose a flow-guide convolution to dynamically warp the features based on the refined optical flow and achieve better feature warping.

3. Methods

In the section, we first give an overview of our DWNet framework and then describe each of its components in detail. Finally, we describe how to optimize the whole network for improving semantic video segmentation.

3.1. Overview. The overall structure of our DWNet framework is illustrated in Figure 2. The inputs of our DWNet are a pair of RGB images I_t and I_{t+k}, where I_t represents the labeled frame and I_{t+k} represents the unlabeled frame randomly selected from the near-by frames of I_t with $k \in [-5, 5]$. The two images are first sent to the shared segmentation network to extract the semantic features F_t and F_{t+k}. Meanwhile, the two images are also sent to the optical flow estimation network to predict the coarse optical flow $O_{t+k \longrightarrow t}$. Then, we utilize the flow refinement module to optimize the optical flow and produce more accurate optical

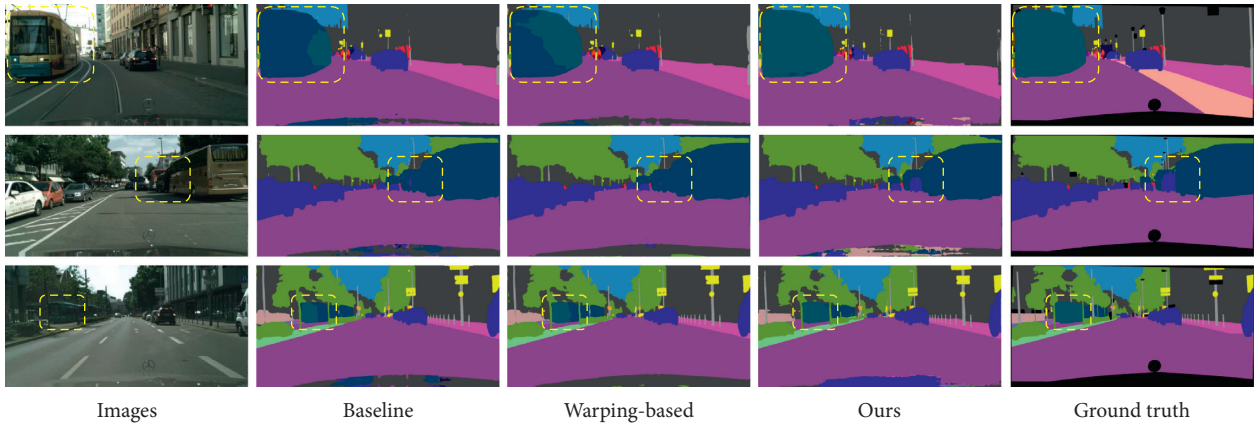


Figure 1: Qualitative results from the Cityscapes dataset. Baseline method: training the model on single frames and inferring the segmentation maps on single frames. Warping-based method: adopting the original warping operation implemented with standard bilinear interpolation to propagate and fuse the features brings a slight improvement. Our method: utilizing the flow-guided convolution to adaptively warp the interframe features and introducing temporal consistency loss to explicitly supervise the warped features instead of simple feature propagation and fusion, hence producing more accurate segmentation results.

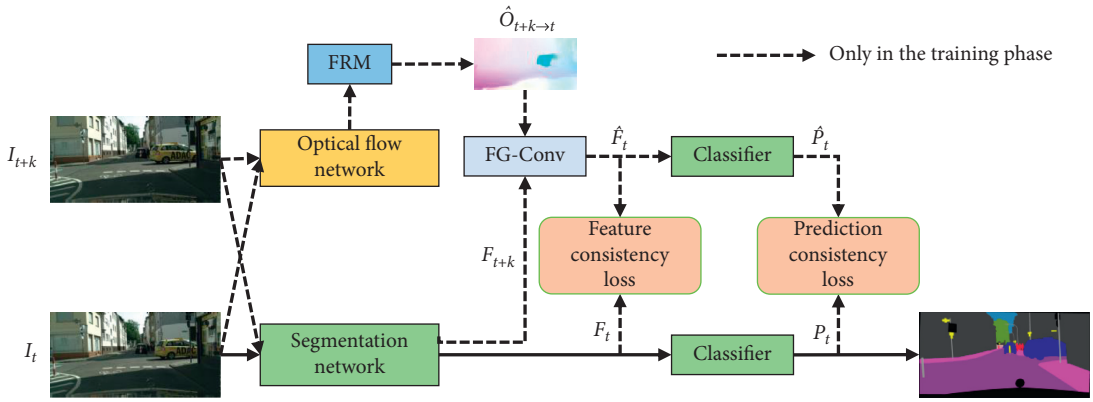


Figure 2: The overall structure of our DWNet framework. FRM denotes the flow refinement module. FG-Conv denotes the flow-guided convolution. Feature consistency loss and prediction consistency loss are both the temporal consistency loss, which improves the temporal consistency of video segmentation. The dotted lines denote that the components are only used in the training phase and will be removed in the inference phase.

flow $\widehat{O}_{t+k\longrightarrow t}$ for every pixel position. After that, we adopt the flow-guided convolution to dynamically warp F_{t+k} to $\widehat{F}_t$ according to the refined optical flow $\widehat{O}_{t+k\longrightarrow t}$. Finally, F_t and $\widehat{F}_t$ are sent to the shared classifier to produce the segmentation map P_t and $\widehat{P}_t$ respectively, and we introduce two kinds of temporal consistency losses as extra constraints to supervise the warped features $\widehat{F}_t$ and $\widehat{P}_t$, respectively. In the following, we will introduce each key component of our DWNet in detail.

3.2. Flow Refinement Module. We first utilize the existing optical flow estimation network to obtain the optical flow $O_{t+k\longrightarrow t}$. The optical flow network computes the pixel displacement $(\Delta x, \Delta y)$ for every pixel location (x, y) in I_t to the spatial locations (x', y') in I_{t+k}, which means that $(x', y') = (x + \Delta x, y + \Delta y)$. And Δx and Δy are floating point numbers and denote pixel displacements in horizontal and vertical directions, respectively [6]. However, the optical flow estimated by the optical flow network may not be enough accurate due to occlusion and new objects. Therefore, we propose the flow refinement module to optimize the coarse optical flow. We concatenate the two input images, the difference of the two images, and the coarse optical flow, resulting in an 11 channel tensor as the input to the flow refinement module. The flow refinement module consists of 4 convolution layers. The first 3 layers are made up of 3×3 kernels with stride 2 following BatchNorm and ReLU, and the number of the output channels is set to 64, 128, and 256, respectively. The output of the third layer is then passed on to the last 3×3 convolution layer with $2s^2$ output channels to attain the refined optical flow $\widehat{O}_{t+k\longrightarrow t}$, whose spatial size is corresponding to the feature F_t and F_{t+k}. s represents the kernel size of the flow-guided convolution which will be discussed in Section 3.3 and is set to 1 as default. We visualize the original optical flow and the refined optical flow, as shown in Figure 3. The refined optical flow has sharper motion boundaries for moving objects and semantics, such as humans and cars, which demonstrates the effectiveness of the flow refinement module. Next, we will introduce how to

Figure 3: Visual comparison on the Cityscapes dataset for the original optical flow. The first column denotes the input frame. The middle column denotes the coarse optical flow produced by the optical flow network. The last column denotes the refined optical flow optimized by the flow refinement module. The refined optical flow has sharper motion boundaries than the original optical flow.

use the refined optical flow to achieve better features warping.

3.3. Flow-Guided Convolution. The flow refinement module utilizes the original optical flow and images to produce more precise optical flow estimation. Given the optical flow, previous methods utilize the warping operation to transform the feature F_{t+k} to the feature of the current frame $\widehat{F}_t$:

$$\widehat{F}_t = \text{warp}\left(F_{t+k}, O_{t+k\longrightarrow t}\right). \tag{1}$$

However, it cannot accurately align the warped feature and the feature of the current frame due to the imprecise optical flow and the original warping operation without any learnable parameters. Hence, we firstly utilize the flow refinement module to optimize the optical flow as discussed in Section 3.2. Besides, we propose the flow-guided convolution to adaptively warp the interframe features. The standard convolution samples the input feature map at fixed locations, and the DCNv1 [27] adds 2D offsets to the regular grid sampling locations to enable free form deformation of the sampling grid. Motivated by this work, we observe that the optical flow which encodes the pixel displacement across frames can be regarded as a specific offset, and we can utilize the optical flow to dynamically warp the interframe features. Formally, the standard 2D convolution can be written as

$$y[i] = \sum_{p}^{P} w[p] \cdot x[i+p], \tag{2}$$

where y denotes the output after the convolution, i denotes the location, x denotes the input features, w denotes the convolution filters with a length of P, and p enumerates P. p is usually the regular sampling locations in a $s \times s$ kernel, and we propose the flow-guided convolution by adding the location offsets into p as follows:

$$y[i] = \sum_{p}^{P} w[p] \cdot x[i+p-\Delta p], \tag{3}$$

where $\Delta p \in \widehat{O}_{t+k\longrightarrow t}$. The refined optical flow is regarded as the offsets for the flow-guided convolution to adaptively sample more corresponding pixel locations between interframe features. The kernel size s is the key parameter for the flow-guided convolution, and we will discuss the parameter in 4.2.2. Compared with the DCNv1 [27], we obtain the

offsets from the flow refinement module instead of applying a convolution layer to the input feature. Hence, we can attain more accurate offsets and achieve better feature warping.

3.4. Temporal Consistency Loss. The flow-guided convolution can dynamically warp the feature F_{t+k} and produce the estimated feature $\widehat{F}_t$ of the current frame. Previous methods concatenate or do the weighted sum of the warped feature $\widehat{F}_t$ and the original feature F_t to achieve feature fusion and propagation. However, we argue that the warped feature $\widehat{F}_t$ is expected to be consistent with the original feature F_t, and the two features should be the same ideally. Hence, we propose the temporal consistency loss to explicitly supervise the feature $\widehat{F}_t$ and the segmentation map $\widehat{P}_t$ respectively. Compared with the previous methods using feature fusion or fusion propagation, we utilize extra constraints to improve the temporal consistency of video segmentation, which is more reasonable and does not introduce additional calculation or postprocessing in the inference phase. The temporal consistency loss contains the feature consistency loss and the prediction consistency loss, which are related to the feature $\widehat{F}_t$ and the segmentation map $\widehat{P}_t$, respectively.

3.4.1. Feature Consistency Loss. We attempt to constraint both features of F_t and $\widehat{F}_t$ to be similar enough by designing a feature consistency loss. Instead of per-pixel similarity calculation, we measure the similarity between the self-attention maps A_t and $\widehat{A}_t$ of both features. Since the self-attention maps present high-order relationships among pixels, such a similarity measurement is more robust than the typical per-pixel one. Let $a_{i,j}$ denote the similarity between the ith pixel and the jth pixel of the original feature F_t, and let $\widehat{a}_{i,j}$ denote the similarity between the ith pixel and the jth pixel of the warped feature $\widehat{F}_t$, where $a_{i,j} \in A_t$ and $\widehat{a}_{i,j} \in \widehat{A}_t$. The $a_{i,j}$ is computed from the feature $F_{t,i}$ and $F_{t,j}$ as

$$a_{i,j} = \frac{F_{t,i}^T F_{t,j}}{\left(\left\|F_{t,i}\right\|_2 \left\|F_{t,j}\right\|_2\right)}. \tag{4}$$

And, we adopt the squared difference to formulate the feature consistency loss:

$$\ell_{\mathrm{fc}}\left(F_t, \widehat{F}_t\right) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(a_{ij} - \widehat{a}_{ij}\right)^2, \tag{5}$$

where N denotes the total number of the pixels. The warped feature and the original feature should produce a similar attention map that encodes the pixel correlations. Hence, this loss can strengthen the feature consistency by explicitly supervising the attention maps.

3.4.2. Prediction Consistency Loss. The segmentation map $\widehat{P}_t$ produced by the feature $\widehat{F}_t$ should be also consistent with the segmentation map P_t of the current frame. Hence, we introduce the prediction consistency loss [17] to improve the temporal consistency of video segmentation as follows:

$$\ell_{\mathrm{pc}}\left(I_t, I_{t+k}\right) = \frac{1}{N} \sum_{i=1}^{N} M_{t+k\longrightarrow t,i} \left\|P_{t,i} - \widehat{P}_{t+k\longrightarrow t,i}\right\|_2^2. \tag{6}$$

Due to the occlusion and new objects across frames, we predict a mask $M_{t+k\longrightarrow t}$ to assign different weights to each pixel according to the warping error $E_{t+k\longrightarrow t}$, where $E_{t+k\longrightarrow t} = |I_t - \widehat{I}_t|$ and $\widehat{I}_t$ denotes the warped input frame from I_{t+k}. Then, $M_{t+k\longrightarrow t}$ is denoted as

$$M_{t+k\longrightarrow t} = \exp\left(-\frac{E_{t+k\longrightarrow t}}{\delta}\right), \tag{7}$$

where δ is a hyperparameter which controls the amplitude of the difference between high error and low error. The pixels with higher warping errors are assigned to lower weights and vice versa, because higher warping error represents that the optical flow and the warped feature are more inaccurate. $M_{t+k\longrightarrow t}$ can speed up the convergence of the prediction consistency loss and improve the accuracy of video segmentation by considering the pixels with more precise optical flow and ignoring the noise produced by occlusion and new objects.

3.5. Optimization. The loss of our DWNet consists of the conventional cross-entropy loss ℓ_{ce} and the temporal consistency loss including the feature consistency loss ℓ_{fc} and the prediction consistency loss ℓ_{pc}. Hence, our final objective function is

$$\ell = \ell_{\mathrm{ce}} + \lambda_1 \ell_{\mathrm{fc}} + \lambda_2 \ell_{\mathrm{pc}}, \tag{8}$$

where λ_1 and λ_2 denote the weights for multiple losses. As illustrated in Figure 2, our DWNet can be trained in an end-to-end fashion. And in the inference phase, the optical flow network, the flow refinement module, and the flow-guided convolution in the dotted line can be removed. Hence, the final network can be regarded as a semantic image segmentation network with no additional calculation or postprocessing during inference.

4. Experiments

4.1. Experimental Setup

4.1.1. Datasets. We evaluate our proposed DWNet on two semantic video segmentation benchmarks datasets Cityscapes [29] and CamVid [30].

Cityscapes is an urban scene dataset and contains 5000 video snippets collected from 50 cities in different seasons. Each snippet contains 30 frames and only the 20th frame is pixel-level finely annotated, leading to the dataset containing 5000 images which are divided into 2975, 500, and 1525 images for training, validation, and testing respectively. Besides, the dataset also contains 20000 coarsely annotated images, but we do not utilize these data in all experiments except otherwise stated.

CamVid is composed of 701 densely annotated images from five video sequences. The images are labeled every 30 frames with 11 semantic classes. Following the previous

work [6], the dataset is split into 367 training, 101 validation, and 233 testing images.

4.1.2. Models. To validate the effectiveness of our proposed method, we conduct extensive experiments with different network configurations. We adopt the ResNet50 [31], ResNet101 [31], and MobileNetv2 [32] as the backbone to extract the high-level features. And we choose the PSPNet [33], DeeplabV3+ [3], and DANet [5] as the segmentation model. The segmentation network is combined with different backbones and segmentation models. We conduct the ablation experiments on ResNet50 with the structure of PSPNet, namely, PSPNet50. Because the optical flow network can be removed in the inference phase, we adopt the more powerful optical flow estimation network FlowNetV2 [11] to extract the more accurate optical flow, though it is slower and with more parameters during training compared with the lightweight FlowNet, like [10, 12].

4.1.3. Implementation Details. We implement our method based on PyTorch. We employ an SGD optimizer and a poly learning rate policy, where the initial learning rate is multiplied by $(1 - (\text{epoch}/\text{max\_epoch}))^{\text{power}}$ with power = 0.9 after each iteration. The base learning rate is set to 0.01 for both datasets. Momentum and weight decay are set to 0.9 and 0.0001, respectively. We utilize the synchronized batch normalization [4] with a batch size of 8 for both datasets. For data augmentation, we apply random scaling of the input images (from 0.5 to 2.2 on Cityscapes, from 0.5 to 2.0 on CamVid), random cropping (768×768 for Cityscapes, 384×384 for CamVid), and random left-right flipping during training. Note that the optical flow network FlowNetV2 is also joint optimized with the base learning rate 0.00001. We employ the standard pixel-wise cross-entropy loss function as the main loss to train the whole network with 8 cards of NVIDIA TITAN RTX. The loss weights are set to be $\lambda_1 = 10$ and $\lambda_2 = 0.1$ for all experiments. After training, we utilize the original images to inference unless otherwise stated. Following the previous works [6, 8], we apply mean intersection-over-union (mIoU) as the evaluation metric to validate our method.

4.2. Ablation Study. We build the DWNet based on the single-frame segmentation model. And, we adopt the PSPNet50 as the single-frame model to conduct all the ablation experiments on the Cityscapes dataset.

4.2.1. Effectiveness of the Proposed Method. In this section, we evaluate the different components of our DWNet with different settings, and the results are shown in Table 1. The baseline model is the PSPNet50 with single-frame training and inference. When we utilize the original warping operation and adopt the feature consistency loss as a constraint, the performance is only improved by 0.55%. However, when we replace the original warping operation with our proposed flow-guided convolution, it brings a further improvement by 0.57%, which demonstrates that the dynamic warping is better than the original warping operation. Besides, the flow refinement module and the prediction consistency loss can improve the performance by 0.47% and 0.38%, respectively. And introducing the two components simultaneously can further improve the accuracy to 75.62%. We also verify whether the two components are beneficial to the warping-based method, and the results show that the accuracy can be improved from 74.3% to 74.76%, whose improvement is lower than our proposed method (from 74.87% to 75.62%).

Table 1: Ablation study of our DWNet on the Cityscapes validation set.

Warp	ℓ_{fc}	FG-Conv	FRM	ℓ_{pc}	mIoU %
					73.75
√	√				74.30
	√	√			74.87
	√	√	√		75.34
	√	√		√	75.25
√	√		√	√	74.76
	√	√	√	√	**75.62**

Warp denotes the original warping operation. ℓ_{fc} and ℓ_{pc} denote the feature consistency loss and prediction consistency loss, respectively. FG-Conv denotes the flow-guided convolution. FRM denotes flow refinement module. The bold values denote our method can achieve the best accuracy compared with other methods.

4.2.2. Flow-Guided Convolution. The flow-guided convolution is the core operation of our DWNet, which utilizes the refined optical flow to adaptively warp the interframe features. The kernel size s is the key parameter for the flow-guided convolution. According to the original warping operation, each pixel corresponds to a specific offset, and we can utilize the offset to warp each pixel independently. However, we argue that we can consider more adjacent pixels to judge the warped result of each pixel. Hence, we can adjust s to achieve more precise feature warping. When s is equal to 1, the flow-guided convolution is similar to the original warping operation which treats each pixel independently. However, our flow-guided convolution contains the learnable parameters and can adaptively adjust the warped features. As shown in Table 2, when we set s to 3, the flow-guided convolution yields the best performance. Besides, the flow-guided convolution with different values of s all outperforms the original warping operation, which demonstrates that our proposed method can achieve better feature warping. When s is set to 5, the accuracy gets worse. We think that the larger s may bring more noise and influence the stable training of the whole model.

4.2.3. Prediction Consistency Loss. The prediction consistency loss aims to improve segmentation stability. We calculate the occlusion mask to speed up the convergence and improve the accuracy of video segmentation by considering the pixels with more precise optical flow and ignoring the noise produced by occlusion and new objects. And the δ is a hyperparameter that controls the amplitude of the difference between high error and low error. Hence, we provide a discussion about the δ, and the results are shown in

Table 2: Ablation study of the flow-guided convolution on the Cityscapes validation set.

Method	mIoU %
Baseline	73.75
Baseline + warp	74.30
Baseline + FG-Conv ($s = 1$)	75.02
Baseline + FG-Conv ($s = 3$)	**75.34**
Baseline + FG-Conv ($s = 5$)	75.16

s denotes the kernel size of the flow-guided convolution. The bold values denote our method can achieve the best accuracy when s is set to 3.

Table 3. We first try the prediction consistency loss without the occlusion mask, and we find the performance decrease by 0.22% compared with the baseline, which demonstrates the importance of the occlusion mask. If we treat all pixels equally, the pixels with high warping errors will seriously affect the training and the final segmentation accuracy. And when we introduce the mask and set δ to 2, it can obtain the best performance.

In fact, the first designs for both temporal consistency losses consider the occlusion and new objects. However, the impact on the feature consistency loss is slight (from 74.87% to 74.89%). The occlusion and new objects usually reflect some small and local changes across different frames, and the feature consistency loss aims to model the long-range and high-order relationships and is more robust to such small changes, while the prediction consistency loss aims to model the per-pixel similarity and is susceptible to the occlusion and new objects. Hence, we only add the occlusion mask in the prediction consistency loss.

4.2.4. Feature Fusion and Propagation. To mask the use of the warped features, previous methods try to do weighted sum or concatenate the warped features and the original features for feature fusion and propagation. We compare the previous methods with our proposed method in Table 4. The results show that our proposed method is obviously better than the previous methods, which demonstrates our conjecture to the warped feature reuse.

4.3. Comparative Results on Cityscapes Dataset

4.3.1. Effectiveness of Different Network Structures. To validate the effectiveness of our DWNet, we apply different network configurations. The results are shown in Table 5. SWarp (Static Warping) denotes the original warping operation and DWarp (Dynamic Warping) denotes our proposed DWNet. The results demonstrate that our DWNet has a strong generalization ability for different network structures and can significantly improve the accuracy compared with the SWarp.

4.3.2. Comparison with State-of-the-Art. We compare our DWNet with existing methods on the Cityscapes test dataset. The results are shown in Table 6, and our DWNet can outperform the existing methods with a significant advantage. In particular, with the PSPNet as the backbone, our method with the only fine set for the train can improve the mIoU score by 0.9%, which is superior to previous methods with both fine and coarse sets for the train, like [6, 13, 15]. And when we also utilize both fine and coarse images for the train, our method can bring a further improvement by 0.7%, which demonstrates the effectiveness of our method. Besides, we utilize the DANet as the segmentation network and the accuracy is improved to 82.1%, which shows that our method has a strong generalization for different segmentation networks.

4.3.3. Qualitative Results. The qualitative comparison is shown in Figure 4. Existing warping-based methods adopt the standard bilinear interpolation without any learnable parameters to warp the interframe features based on imprecise precomputed optical flow and produce the negative

Table 3: Ablation study of the prediction consistency loss on the Cityscapes validation set.

Method	mIoU %
Baseline	75.34
Baseline + ℓ_{pc} + w/o mask	75.13
Baseline + ℓ_{pc} + w/mask ($\delta = 1$)	75.46
Baseline + ℓ_{pc} + w/mask ($\delta = 2$)	**75.62**
Baseline + ℓ_{pc} + w/mask ($\delta = 5$)	75.44

δ denotes the amplitude of the difference between high warping error and low warping error. The bold values denote our method can achieve the best accuracy when delta is set to 2.

Table 4: Ablation study of feature fusion and propagation on the Cityscapes validation set.

Method	mIoU %
Baseline	73.75
Baseline + sum	74.30
Baseline + concatenate	74.25
Baseline + TCLoss (ℓ_{fc})	74.87
Baseline + TCLoss ($\ell_{fc} + \ell_{pc}$)	**75.25**

Sum and Concatenate denote the weighted sum and concatenation of the warped features and the original features for feature fusion, respectively. TCLoss denotes the temporal consistency loss, including feature consistency loss and prediction consistency loss. The bold values denote our method can achieve the best accuracy using both the feature consistency loss and prediction consistency loss.

Table 5: Comparison of our DWNet with different network structures on the Cityscapes validation set.

Method	Backbone	SWarp	DWarp	mIoU %
PSPNet	MNV2			72.34
PSPNet	MNV2	√		73.52
PSPNet	MNV2		√	**74.46**
PSPNet	ResNet101			78.90
PSPNet	ResNet101	√		79.32
PSPNet	ResNet101		√	**79.85**
DeeplabV3+	ResNet101			80.15
DeeplabV3+	ResNet101	√		80.32
DeeplabV3+	ResNet101		√	**80.78**
DANet	ResNet101			79.94
DANet	ResNet101	√		80.21
DANet	ResNet101		√	**80.67**

SWarp (Static Warping) denotes the original warping operation. DWarp (Dynamic Warping) denotes our proposed DWNet. MNV2 denotes the MobileNetV2. The bold values denote our method can achieve the higher accuracy than the static warp with different baseline models.

Table 6: Comparison of state-of-the-art semantic video segmentation models on the Cityscapes test set.

Method	Source	mIoU %
Clockwork [34]	ECCV2016	66.4
PEARL [35]	ICCV2017	75.4
LLVSS [18]	CVPR2018	76.8
Accel [8]	CVPR2019	75.5
TDNet [19]	CVPR2020	79.9
ESVS [17]	ECCV2020	76.6
PSPNet [33]	CVPR2017	80.2
PSPNet + NetWarp ‡ [6]	ICCV2017	80.5
PSPNet + GRFP ‡ [15]	CVPR2018	80.6
PSPNet + EFC ‡ [13]	AAAI2020	81.0
PSPNet + ours		**81.1**
PSPNet + ours‡		**81.8**
DANet [5]	CVPR2019	81.5
DANet + ours		**82.1**

Methods trained using both fine and coarse sets are marked with "‡." The bold values denote our method can achieve the best accuracy compared with other state-of-the-art methods.

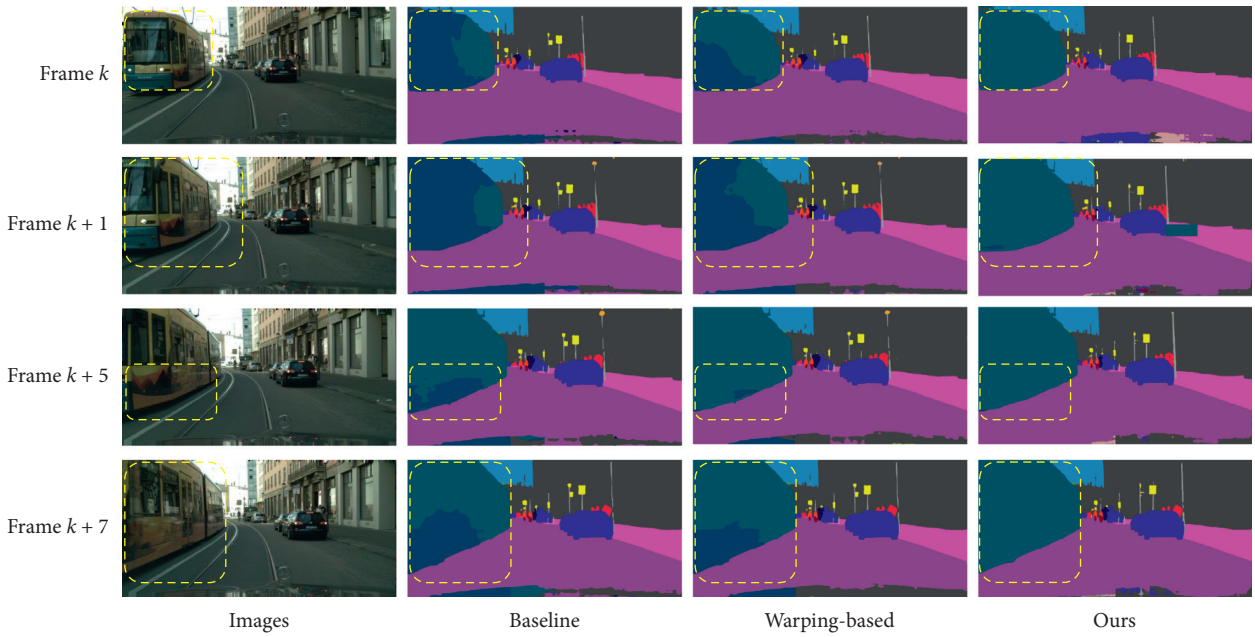


Figure 4: Qualitative results of consecutive frames on the Cityscapes dataset. Baseline methods: training and inferring on single frames. Warping-based method: adopting the original warping operation to enhance the feature. Our method: utilizing the flow-guided convolution to adaptively warp the interframe features. Compared with the baseline, the warping-based method brings a slight improvement in the moving objects, and our method can produce more accurate and consistent segmentation results.

Table 7: Comparison of state-of-the-art semantic video segmentation models on the CamVid test set.

Method	Source	mIoU %
STFCN [20]	arXiv2016	65.9
DFF [7]	CVPR2017	66.0
NetWarp [6]	ICCV2017	70.3
GRFP [15]	CVPR2018	66.1
Accel [8]	CVPR2019	69.3
EFC [13]	AAAI2020	67.4
TDNet [19]	CVPR2020	76.0
ESVS [17]	ECCV2020	76.3
PSPNet [33]	CVPR2017	75.4
PSPNet + ours		**76.5**

The bold values denote our method can achieve the best accuracy compared with other methods on the CamVid test set.

results in the highlighted regions. Compared with the existing warping-based methods, our method adopts the dynamic warping operation to achieve more precise feature alignment based on the refined optical flow and improve temporal consistency of video segmentation.

4.4. Comparative Results on CamVid Dataset. To evaluate the generalization of our method on different datasets, we conduct experiments on the CamVid dataset. We use the ResNet101 as the backbone with the architecture of PSPNet. The results are shown in Table 7, and our method outperforms the current state-of-the-art methods, which demonstrates the generalization for different datasets.

5. Conclusion

In this paper, we propose a novel framework named DWNet to adaptively warp the interframe features. We design the flow refinement module to optimize the optical flow and propose the flow-guide convolution to achieve adaptive feature alignment. Besides, we introduce the temporal consistency loss to explicitly supervise the warped features to guarantee the temporal consistency of video segmentation. Extensive experiments have shown that our method outperforms existing warping-based methods and achieves state-of-the-art on the Cityscapes and CamVid benchmark datasets.

Data Availability

The Cityscapes and CamVid data can be downloaded freely at https://www.cityscapes-dataset.com/file-handling/?packageID=3 and http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the China Central Universities of USTB (FRF-DF-19-002), Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB (BK20BE014).

References

[1] J. Long, E. Shelhamer, and T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*, CVPR, London, UK, 2015.
[2] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017.
[3] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, *Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation*, ECCV, London, UK, 2018.
[4] H. Zhang, K. Dana, J. Shi et al., *Context Encoding for Semantic Segmentation*, CVPR, London, UK, 2018.
[5] J. Fu, J. Liu, H. Tian et al., "Dual attention network for scene segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 22, pp. 3146–3154, 2019.
[6] R. Gadde, V. Jampani, and P. V. Gehler, *Semantic Video Cnns through Representation Warping*, ICCV, London, UK, 2017.
[7] X. Zhu, Y. Xiong, J. Dai, L. Yuan, and Y. Wei, *Deep Feature Flow for Video Recognition*, CVPR, London, UK, 2017.
[8] S. Jain, X. Wang, and J. E. Gonzalez, *Accel: A Corrective Fusion Network for Efficient Semantic Segmentation on Video*, CVPR, London, UK, 2019.
[9] J. Feng, S. Li, Y. Chen, F. Huang, J. Cui, and X. Li, *How to Train Your Dragon: Tamed Warping Network for Semantic Video Segmentation*, 2020.
[10] A. Dosovitskiy, P. Fischer, E. Ilg et al., *Flownet: Learning Optical Flow with Convolutional Networks*, ICCV, London, UK, 2015.
[11] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: evolution of optical flow estimation with deep networks," CVPR, London, UK, 2017.
[12] D. Sun, X. Yang, M. Y. Liu, and J. Kautz, *Pwc-net: Cnns for Optical Flow Using Pyramid, Warping, and Cost Volume*, CVPR, London, UK, 2018.
[13] M. Ding, Z. Wang, B. Zhou, J. Shi, Z. Lu, and P. Luo, *Every Frame Counts: Joint Learning of Video Segmentation and Optical Flow*, AAAI, London, UK, 2020.
[14] S. Chandra, C. Couprie, and I. Kokkinos, *Deep Spatio-Temporal Random Fields for Efficient Video Segmentation*, CVPR, London, UK, 2018.
[15] D. Nilsson and C. Sminchisescu, *Semantic Video Segmentation by Gated Recurrent Flow Propagation*, CVPR, London, UK, 2018.
[16] Y. S. Xu, T. J. Fu, H. K. Yang, and C. Y. Lee, *Dynamic Video Segmentation Network*, CVPR, London, UK, 2018.
[17] Y. Liu, C. Shen, C. Yu, and J. Wang, *Efficient Semantic Video Segmentation with Per-Frame Inference*, ECCV, London, UK, 2020.
[18] Y. Li, J. Shi, and D. Lin, *Low-Latency Video Semantic Segmentation*, CVPR, London, UK, 2018.
[19] P. Hu, F. Caba, O. Wang, Z. Lin, S. Sclaroff, and F. Perazzi, *Temporally Distributed Networks for Fast Video Semantic Segmentation*, CVPR, London, UK, 2020.
[20] M. Fayyaz, M. H. Saffar, M. Sabokrou, M. Fathy, R. Klette, and F. Huang, "STFCN: spatio-temporal FCN for semantic video segmentation," 2016.
[21] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, *Deep End2end Voxel2voxel Prediction*, CVPR, London, UK, 2016.
[22] F. S. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, and J. M. Alvarez, *Bringing Background into the Foreground: Making All Classes Equal in Weakly-Supervised Video Semantic Segmentation*, ICCV, London, UK, 2017.
[23] J. Zhuang, Z. Wang, and B. Wang, "Video semantic segmentation with distortion-aware feature correction," 2020.
[24] D. B. Bert, J. Xu, T. Tinne, and V. G. Luc, *Dynamic Filter Networks*, NIPS, London, UK, 2016.
[25] H. Su, V. Jampani, D. Sun, O. Gallo, E. Learned-Miller, and J. Kautz, *Pixel-Adaptive Convolutional Neural Networks*, CVPR, London, UK, 2019.
[26] J. Liu, J. He, S. R. Jimmy, Y. Qiao, and H. Li, *Learning to Predict Context-Adaptive Convolution for Semantic Segmentation*, ECCV, London, UK, 2020.
[27] J. Dai, H. Qi, Y. Xiong et al., *Deformable Convolutional Networks*, ICCV, London, UK, 2017.
[28] X. Zhu, H. Hu, S. Lin, and J. Dai, *Deformable ConvNets V2: More Deformable, Better Results*, CVPR, London, UK, 2019.

[29] M. Cordts, M. Omran, S. Ramos et al., *The Cityscapes Dataset for Semantic Urban Scene Understanding*, CVPR, London, UK, 2016.
[30] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, *Segmentation and Recognition Using Structure from Motion Point Clouds*, ICCV, London, UK, 2008.
[31] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, CVPR, London, UK, 2016.
[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, *MobileNetV2: Inverted Residuals and Linear Bottlenecks*, CVPR, London, UK, 2018.
[33] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, *Pyramid Scene Parsing Network*, CVPR, London, UK, 2017.
[34] E. Shelhamer, K. Rakelly, J. Hoffman, and T. Darrell, *Clockwork Convnets for Video Semantic Segmentation*, ECCV, London, UK, 2016.
[35] X. Jin, X. Li, H. Xiao et al., *Video Scene Parsing with Predictive Feature Learning*, ICCV, London, UK, 2017.

Hindawi
Complexity
Volume 2021, Article ID 6685306, 9 pages
https://doi.org/10.1155/2021/6685306

Research Article

On ISRC Rumor Spreading Model for Scale-Free Networks with Self-Purification Mechanism

Zijun Wang [1,2] **and An Chen** [1,2]

[1] *Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China*
[2] *University of Chinese Academy of Sciences, Beijing 100190, China*

Correspondence should be addressed to An Chen; change1970@163.com

Received 9 October 2020; Revised 26 December 2020; Accepted 16 January 2021; Published 3 February 2021

Academic Editor: Jianxiang Xi

At present, the feasibility of using self-purification mechanism to inhibit rumor spreading has been confirmed by studies from different perspectives. This paper improves the classical rumor spreading models with self-purification mechanism, analyzes the correlation between spreading threshold in the model and its self-purification level theoretically, and conducts numerical simulations to study the impact of the changes of model parameters on key indicators in the process of rumor spreading. The simulation results show that changes of model parameters, including self-purification level and forgetting rate, exert significant influences on rumor spreading exactly.

1. Introduction

Rumors generally refer to unconfirmed information. Although they contain uncertain elements, most of them are false and wrong [1, 2]. For example, after the nuclear leakage caused by the Japanese earthquake, the rumor that eating iodized salt can prevent radiation led to a large number of people buying salt in many places, which not only caused public panic but also seriously disturbed the normal social order.

The systematic study of rumors began in World War II. Knapp [3], whose research laid a foundation for the study, sorted out the rumors generated during the war and classified them. In 1947, Allport and Postman [4] believed that rumors were affected by personal values and other psychological factors and proposed the formula for the influence of rumors: R (influence of rumors) $= I$ (importance of event) $\times A$ (ambiguity of event). The research study on the dynamics of rumor spreading arose during the 1960s. Because the process of rumor spreading is similar to the process of disease infection, the nodes in the model of rumor spreading is frequently divided into three categories by many studies using the dynamic model of disease infection for reference. These are the ignorant, spreaders, and stiflers, corresponding to the susceptible, the infected, and the recovered, respectively, in the model of disease infection. The existing dynamical models of rumor spreading are based on two classical models: DK model [5] and MT model [6]. However, these models failed to accurately describe the rumor spreading process in large-scale social networks with increasing complexity of the objects studied. Therefore, scholars paid more attention to the phenomenon of rumor spreading in complex networks. Moreno et al. [7, 8] examined the rumor spreading thresholds in homogeneous and heterogeneous networks, respectively. Mo and Guo [9] proposed a new control protocol to force the multiagent systems to achieve robust consensus. Zanette [10, 11] applied the rumor spreading model to static and dynamic small-world networks and verified the existence of critical threshold of the rumor spreading. Pan et al. [12] proved that high clustering of the network could effectively resist the spread of rumors. Zhao et al. [13, 14] investigated the dynamic characteristics of the rumor spreading in the BA scale-free network and BBV network and held a viewpoint that the topology of the network has a great influence on the rumor spreading. Scholars also consider the influence of

forgetting mechanism [15–17], rejection mechanism [18–20], refutation mechanism [21–23], conformity and authoritative effect [24], ambiguity and attraction characteristics of rumors [25], noise in environment [26, 27], differences of spreading environment [28–31], prevention and control strategies [32–34], and other related factors on the process of rumor spreading. Based on the specific situations and problems, traditional models are modified to ensure their applicability.

Suppressing the spread of rumors and actively spreading the truth of information are two main methods to cope with rumor spreading in practice. However, some problems such as high coping costs, difficulty in confirming the truth in time, and insufficient effectiveness of external intervention keep emerging during this process [35]. Thus, scholars discuss how to apply the self-purification mechanism to solve relevant troubles [36], that is, to encourage users to release the complementation and correction of information, as well as refute false information. At present, the feasibility of using self-purification mechanism to inhibit rumor spreading has been confirmed by studies from different perspectives. Ge et al. [37] analyzed the inhibition effect of collective intelligence on the spread of false information in social media. Xia et al. [35] developed a new rumor spreading model for social media with self-purification mechanism and tested simulations, which unfolded the ability of social media to self-purify rumors. Tanaka et al. [38] proved that public questions and criticisms of rumors can affect individual judgments so that the spread of rumors can be suppressed.

Most of the current research studies on self-purification mechanism of networks are qualitative or case studies, and some other rumor spreading models are built with self-purification and skepticism mechanism. Zan et al. [19] who considered counterattack and self-resistance mechanism based on SIR model proposed the SICR rumor spreading model. Wang and Zhao [39] studied the SIQR rumor spreading model with skepticism mechanism and found that rumor truth disseminating rate plays an important role in rumor spreading process. Zhao et al. [40] added the group of doubters and constructed an SIHR rumor spreading model with self-protection awareness and skepticism mechanism, which suggests that many countermeasures can effectively prevent the dissemination of rumor, such as reducing the real contact rate between the ignorant and spreaders, improving the attention-degree of media to skeptics, and cutting down the depletion rate of mass media.

It is worth noting that the above models with self-purification mechanism are mostly based on homogeneous networks, while the theoretical analysis and related simulation studies of the complex networks, such as online social networks, are less involved. Moreover, once the status of doubters in most models is determined, even if they contact with other groups, the identity of doubters will not change. That is not completely consistent with the actual situation that rumor spreaders are mostly skeptical before spreading rumors [41]. In order to improve these shortcomings, the new rumor spreading model is introduced, traditional interaction rules are innovated, and correlation between spreading threshold in the model and its self-purification level is analyzed in this paper. In addition, simulations are designed and conducted to prove that the changes of model parameters, including self-purification level and forgetting rate, exert significant influences on rumor spreading exactly.

In this paper, the novelties are as follows. (1) adding the rule that "the criticizer may become a rumor spreader after hearing rumors for many times" as a complement to classical rules; (2) theoretical analysis of correlation between spreading threshold and self-purification level of the network is conducted; and (3) the influence of combinations of varying rumor spreading rate, forgetting rate, and self-purification capacity on the key indicators during rumor spreading process is explored. In the next section, the rumor spreading model for scale-free networks is built and relevant interaction rules, transformation relationships, and mean-field equations are introduced. Theoretical analysis about the correlation between spreading threshold and self-purification level is performed in Section 3. In Section 4, relevant numerical simulations are tested. Finally, conclusions and discussions are shown in Section 5.

2. Rumor Spreading Model Building

When building the rumor spreading model with self-purification mechanism, "whether an individual believes in rumor, whether an individual spreads rumor, and whether an individual criticizes rumor" are taken as a basis for grouping, and the situation of criticizing is a sufficient but non-necessary condition for the situation of disbelieving.

The rumor spreading model in this paper is called the ISRC model, where 'I' stands for the ignorant, 'S' means spreaders, 'R' denotes the recovered, and 'C' indicates criticizers. $I(t)$ refers to the density of the ignorant who do not hear the rumor at time t, and for the same reason, $S(t)$, $R(t)$, and $C(t)$ show the density of spreaders who believe in and spread the rumor at time t, the density of the recovered who can determine falsity of the rumor and do not spread or criticize it at time t, and the density of criticizers who criticize the rumor at time t, respectively. It is a remarkable fact that the criticizers consist of the individuals who can accurately identify the rumor and the others who criticize the rumor without understanding the truth of rumor information. The latter is likely to be assimilated by rumor spreaders due to the herd effect after getting access to the rumor for many times. That corresponds to the conclusion in the previous study that "the criticizer may become a rumor spreader after hearing rumors for many times [41]." Correspondingly, $I_k(t)$ represents the density of the ignorant with connectivity k at time t. The meanings of $S_k(t)$, $R_k(t)$, and $C_k(t)$ are similar to that of $I_k(t)$ and will not be repeated here. We have that $I(t) = \sum_k I_k(t)p(k)$ with $p(k)$ the degree distribution and so do $S_k(t)$, $R_k(t)$, and $C_k(t)$. In addition, $I(t) + S(t) + R(t) + C(t) = 1$ and $I_k(t) + S_k(t) + R_k(t) + C_k(t) = 1$.

Here, we assume that (1) the rumor spreading model targets a single rumor and does not consider the interaction among multiple rumors; (2) the method without nodes joining or leaving is used to keep the total number of nodes

in the network constant in the selected period; (3) the transformation of these four groups must be based on the information exchange between groups, and only spreaders and criticizers can spread the information, the others are receivers of the information; (4) considering the herd effect, some criticizers who do not understand the truth may be assimilated by rumor spreaders after repeated exposures to the rumor; similarly, rumor spreaders have chance to be the recovered after hearing criticizing information for many times; and (5) rumor exchange within the group of spreaders has no effect on the change of status of both sides.

For scale-free networks, the transformation relationships and interaction rules among I, S, R, and C are illustrated in Figure 1.

The following formulas hold $\alpha + a_1 + a_2 = 1$ and $b_1 + b_2 = 1$. According to the relevant dynamical method, the mean-field equations can be described as follows:

$$\begin{cases} \dfrac{\mathrm{d}I_k(t)}{\mathrm{d}t} = -kI_k(t)\left[\sum_{k'} C_{k'}(t)p\left(\dfrac{k'}{k}\right) + \sum_{k'} S_{k'}(t)p\left(\dfrac{k'}{k}\right)\right], \\ \dfrac{\mathrm{d}S_k(t)}{\mathrm{d}t} = \alpha kI_k(t)\sum_{k'} S_{k'}(t)p\left(\dfrac{k'}{k}\right) + a_3kC_k(t)\sum_{k'} S_{k'}(t)p\left(\dfrac{k'}{k}\right) - \beta S_k(t) - \theta kS_k(t)\sum_{k'} C_{k'}(t)p\left(\dfrac{k'}{k}\right), \\ \dfrac{\mathrm{d}R_k(t)}{\mathrm{d}t} = \beta S_k(t) + \beta C_k(t) + \theta kS_k(t)\sum_{k'} C_{k'}(t)p\left(\dfrac{k'}{k}\right) + a_1kI_k(t)\sum_{k'} S_{k'}(t)p\left(\dfrac{k'}{k}\right) + b_1kI_k(t)\sum_{k'} C_{k'}(t)p\left(\dfrac{k'}{k}\right), \\ \dfrac{\mathrm{d}C_k(t)}{\mathrm{d}t} = a_2kI_k(t)\sum_{k'} S_{k'}(t)p\left(\dfrac{k'}{k}\right) + b_2kI_k(t)\sum_{k'} C_{k'}(t)p\left(\dfrac{k'}{k}\right) - \beta C_k(t) - a_3kC_k(t)\sum_{k'} S_{k'}(t)p\left(\dfrac{k'}{k}\right), \end{cases} \tag{1}$$

where $p(k'/k)$ denotes the conditional probability that a node with k links is connected to a node with degree k'. It can be written as $p(k'/k) = k' \cdot p(k')/\langle k\rangle$, where $\langle k\rangle$ means the average degree and $p(k')$ denotes the degree distribution [15].

3. Analysis of Correlation

In this section, we analyze the correlation between spreading threshold in the ISRC model for scale-free networks and self-purification level. Considering the model shown in Figure 1, only nodes of type I and type R exist in the final network. Referring to the idea on the SIR-like model for complex networks in reference [42], spreaders S and criticizers C are grouped into one category in this section and recorded as Information Spreaders, denoted by X. This model is called IXR model, and its transformation relationships and interaction rules among I, X, and R are shown in Figure 2.

Mean-field equations are depicted by the following equations:

$$\frac{\mathrm{d}I_k(t)}{\mathrm{d}t} = -kI_k(t)\sum_{k'} X_{k'}(t)p\left(\frac{k'}{k}\right), \tag{2}$$

$$\frac{\mathrm{d}X_k(t)}{\mathrm{d}t} = p_1kI_k(t)\sum_{k'} X_{k'}(t)p\left(\frac{k'}{k}\right) - p_3kX_k(t)\sum_{k'} X_{k'}(t)p\left(\frac{k'}{k}\right) - p_4X_k(t), \tag{3}$$

$$\frac{\mathrm{d}R_k(t)}{\mathrm{d}t} = p_2kI_k(t)\sum_{k'} X_{k'}(t)p\left(\frac{k'}{k}\right) + p_3kX_k(t)\sum_{k'} X_{k'}(t)p\left(\frac{k'}{k}\right) + p_4X_k(t), \tag{4}$$

where $p_1 + p_2 = 1$ and $p(k'/k)$ stand for the conditional probability, represented as $p(k'/k) = k' \cdot p(k')/\langle k\rangle$, where $\langle k\rangle$ means average degree and $p(k')$ denotes degree distribution.

We assume a homogeneous initial distribution of the ignorant $I_k(0) = I(0)$, and set $I_k(0) \approx 1$ without loss of generality. In this case, equation (2) can be integrated directly yielding

$$I_k(t) = e^{-k\Phi(t)}, \tag{5}$$

where an auxiliary function is defined as follows:

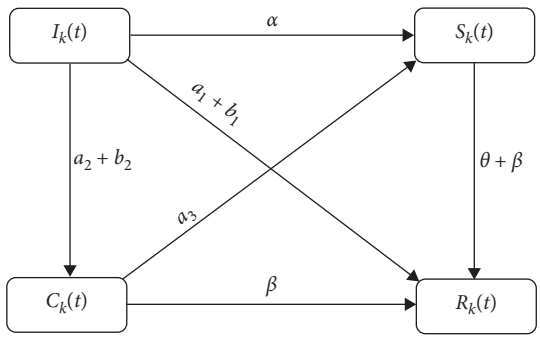


Figure 1: Transformation relationships and interaction rules among I, S, R, and C for scale-free networks where α represents probability of transforming from I to S after contacting (S) called rumor spreading rate; a_1 represents probability of transforming from I to R after contacting (S); a_2 represents probability of transforming from I to C after contacting (S); a_3 represents probability of transforming from C to S after contacting (S); b_1 represents probability of transforming from I to R after contacting (C); b_2 represents probability of transforming from I to C after contacting (C); θ represents probability of transforming from S to R after contacting (C) called self-purification level; and β represents probability of transforming from S or C to (R) called forgetting rate.

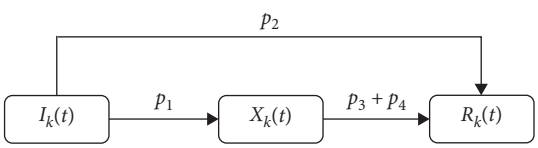


Figure 2: Transformation relationships and interaction rules among I, X, and R where p_1 represent probability of transforming from I to X after contacting X called information spreading rate; p_2 represents probability of transforming from I to R after contacting X; p_3 represents probability of transforming from X to R after contacting X called self-purification level; and p_4 represents probability of transforming from X to R spontaneously, called forgetting rate.

$$\Phi(t) = \int_0^t \sum_k \langle\langle X_k(t')\rangle\rangle \mathrm{d}t', \tag{6}$$

with the shortened form $\langle\langle M_k(t')\rangle\rangle = \sum_k M_k(t')q(k)$, and $q(k')$ is the simple mark for $p(k'/k)$ for convenience.

Multiplying equation (3) with $q(k)$, summing over, and integrating terms in the equation, we obtain

$$\begin{aligned}\frac{\mathrm{d}\Phi(t)}{\mathrm{d}t} &= p_1 \int_0^t \sum_k kI_k(t')q(k)\langle\langle X_{k'}(t')\rangle\rangle \mathrm{d}t' \\ &\quad - p_3 \int_0^t \sum_k kX_k(t')q(k)\langle\langle X_{k'}(t')\rangle\rangle \mathrm{d}t' - p_4\Phi(t) \\ &= p_1\left[1 - \langle\langle e^{-k\Phi(t)}\rangle\rangle\right] - p_3 \int_0^t \langle\langle kX_k(t')\rangle\rangle\langle\langle X_k(t')\rangle\rangle \mathrm{d}t' - p_4\Phi(t).\end{aligned} \tag{7}$$

Because $\Phi(t)$ and $\Phi(\infty) = \lim_{t\longrightarrow\infty}\Phi(t)$ are very small when close to the critical threshold, we assert $\Phi(t) = f(t)\Phi(\infty)$, where $f(t)$ denotes a finite function. We solve equation (3) using the method to ODE, solve equation (7) using Taylor series expansion and retaining leading terms and eventually derive the expression of $X_k(t)$ as follows:

$$\begin{aligned}X_k(t) &= p_1 \int_0^t kI_k(t') \sum_{k'} X_{k'}(t')q(k')\,\mathrm{d}t' \\ &\quad - p_3 \int_0^t kX_k(t') \sum_{k'} (X_{k'}(t'))q(k')\,\mathrm{d}t' \\ &\quad - p_4 \int_0^t X_k(t')\mathrm{d}t' \\ &= p_1 k\Phi(\infty)\left[f(t) - p_4 \int_0^t f(t')e^{p_4(t'-t)}\mathrm{d}t'\right] \\ &\quad + O\left(\Phi^2(\infty)\right) + O(p_3).\end{aligned} \tag{8}$$

When $t \longrightarrow \infty$, $(\mathrm{d}\Phi(t)/\mathrm{d}t) \longrightarrow 0$. Therefore, according to equation (7), we have

$$\begin{aligned}&p_1\left[1 - \langle\langle e^{-k\Phi(\infty)}\rangle\rangle\right] - p_3 \int_0^\infty \langle\langle kX_k(t')\rangle\rangle\langle\langle X_{k'}(t')\rangle\rangle \mathrm{d}t' \\ &\quad - p_4\Phi(\infty) = 0.\end{aligned} \tag{9}$$

Insert equations (8) into (9) and expand the exponential to the relevant order in $\Phi(\infty)$ yielding

$$\begin{aligned}&p_1\left[1 - \langle\langle 1 - k\Phi(\infty) + \frac{k^2\Phi^2(\infty)}{2}\rangle\rangle\right] - p_3\langle\langle k^2\rangle\rangle\langle\langle k\rangle\rangle\Phi^2(\infty)L \\ &\quad - p_4\Phi(\infty) + O\left(\Phi^3(\infty)\right) \\ &\quad + O\left((p_3)^2\right) = 0,\end{aligned} \tag{10}$$

where $L = \int_0^\infty [p_1(f(t) - p_4\int_0^t f(t')e^{p_4(t'-t)}\mathrm{d}t')]^2\mathrm{d}t$ is a positive-defined integral.

Consequently,

$$\Phi(\infty)\left[p_1\langle\langle k\rangle\rangle - \frac{p_1}{2}\langle\langle k^2\rangle\rangle\Phi(\infty) - p_3\langle\langle k^2\rangle\rangle\langle\langle k\rangle\rangle L\Phi(\infty) - p_4\right] + O\left(\Phi^3(\infty)\right) + O\left((p_3)^2\right) = 0, \tag{11}$$

where $\Phi(\infty) = 0$ is always a solution. We can find the nonzero solution as follows:

$$\Phi(\infty) = \frac{2\left(p_1\langle\langle k\rangle\rangle - p_4\right)}{\langle\langle k^2\rangle\rangle\left[p_1 + 2p_3\langle\langle k\rangle\rangle L\right]}. \tag{12}$$

From expression of $\Phi(\infty)$ in equation (12), we learn $\langle\langle k^2\rangle\rangle[p_1 + 2p_3\langle\langle k\rangle\rangle L]$ is always positive. $2(p_1\langle\langle k\rangle\rangle - p_4) > 0$ needs to be true to ensure $\Phi(\infty)$ is positive, i.e., $p_1 > p_4/\langle\langle k\rangle\rangle = p_4\langle k\rangle/\langle k^2\rangle$. Therefore, we obtain the spreading threshold in the IXR model:

$$\lambda = \frac{p_4\langle k\rangle}{\langle k^2\rangle}. \tag{13}$$

This result shows that the spreading threshold in the IXR model is not only related to the degree of nodes in the network k but also depends on the forgetting rate p_4. That actually indicates that spreading threshold in the IXR model for scale-free networks is uncorrelated with the self-purification level p_3. It can be deduced that there is no correlation between spreading threshold in the ISRC model for scale-free networks and self-purification level of the network.

4. Numerical Simulations

In this section, the Runge–Kutta method is used to solve the system of differential equation (1), and the numerical simulations are conducted by using NetLogo to analyze the influence of changes of rumor spreading rate, forgetting rate, and self-purification level on process and results of rumor spreading in scale-free networks. Maximum value of the sum of the density of group S in the process of rumor spreading is regarded as peak value of rumor influence (abbreviated as PVI), and the moment when this situation is reached is called arrival time of peak value of rumor influence (abbreviated as TPVI). Since the types of remaining nodes in the final network will only be part or all of I and R, the situation in which the density of I and the density of R is no longer changing is set a sign of the end of the rumor spreading process. The duration of rumor spreading is abbreviated as DRS. These three indicators (PVI, TPVI, and DRS) reflect the pros and cons of the effect to inhibit rumor spreading. For example, if with smaller PVI, earlier TPVI, and shorter DRS, there will be better effect to inhibit rumor spreading.

In this section, the rumor is set to spread in a scale-free network with $N = 10^3$ nodes, power exponent $\gamma = 2.8$, and average degree $\langle k\rangle = 5$. In simulations, there are 10 spreaders in the initial network, i.e., $S(0) = 10/10^3$, $I(0) = (10^3 - 10/10^3)$, $R(0) = 0$, and $C(0) = 0$. 30 simulations under each condition is performed, and average value of all results under each condition as final result is taken (the values of PVI are account to two decimal places, and the values of TPVI and DRS are accurate to one decimal place).

Figure 3 shows the influence of changes of rumor spreading rate α and self-purification level θ on three indicators (PVI, TPVI, and DRS) during the process of rumor spreading. Figure 3(a) displays how PVI changes for four values of α. It can be seen from Figure 3(a) that PVI decreases as θ increases, and for the same level of self-purification, the higher the rumor spreading rate α is, the greater the PVI is. Figure 3(a) reflects that the improving self-purification level and reducing rumor spreading rate are conducive to reducing the peak value of rumor influence in process of rumor spreading. Correspondingly, Figures 3(b) and 3(c) reveal how TPVI and DRS change. From Figure 3(b), we can obtain that with small α ($\alpha = 0.2, 0.4, 0.6$), TPVI is negatively correlated with θ, but TPVI has little changes in the situation of $\alpha = 0.8$. When $\theta < 0.3$, TPVI decreases as α increases, while $\theta > 0.3$, TPVI decreases in α. Therefore, Figure 3(b) implies that if rumor spreading rate remains at a high level, the improvement of self-purification level of networks may not affect the arrival time of peak value of rumor influence, and Figure 3(b) also indicates that $\theta = 0.3$ is a critical value; when $\theta < 0.3$, the higher the rumor spreading rate is, the earlier the arrival time of peak value of rumor influence is; when $\theta > 0.3$, result is the opposite. Figure 3(c) manifests that α and θ have minor effects on DRS, which also means duration of rumor spreading is not clearly related to rumor spreading rate or self-purification level. Therefore, as a whole, at a low rumor spreading rate and high self-purification level, there is a better effect to inhibit rumor spreading in the network, while in the case of high rumor spreading rate, as the level of self-purification increases, the effect to inhibit rumor spreading may become worse.

Figure 4 displays the influence of changes of forgetting rate β and self-purification level θ on three indicators (PVI, TPVI, and DRS) during the process of rumor spreading. Figure 4(a) shows how PVI changes for four values of β. We can learn from Figure 4(a) that PVI is negatively correlated with θ, and for the same level of self-purification, the higher the forgetting rate β is, the smaller the PVI is. Figure 4(a) reveals that the improvement of self-purification level and forgetting rate is conducive to reducing the peak value of rumor influence in process of rumor spreading. Accordingly, Figure 4(b) demonstrates how TPVI changes. For same β, as θ increases, the values of TPVI tend to decrease. However, as forgetting rate β increases, the changes of TPVI become more and more irregular and erratic. It is conveyed by Figure 4(c) that DRS is negatively correlated with β, and for same forgetting rate, the change of θ has a minor effect on DRS. Figure 4(c) also means that duration of rumor spreading is considerably related to forgetting rate. Thus, whether the network is at a high or low forgetting rate, on the whole, the increase in self-purification level will lead to a tendency to suppress rumor spreading. However, compared with the case of high forgetting rate, when the forgetting rate

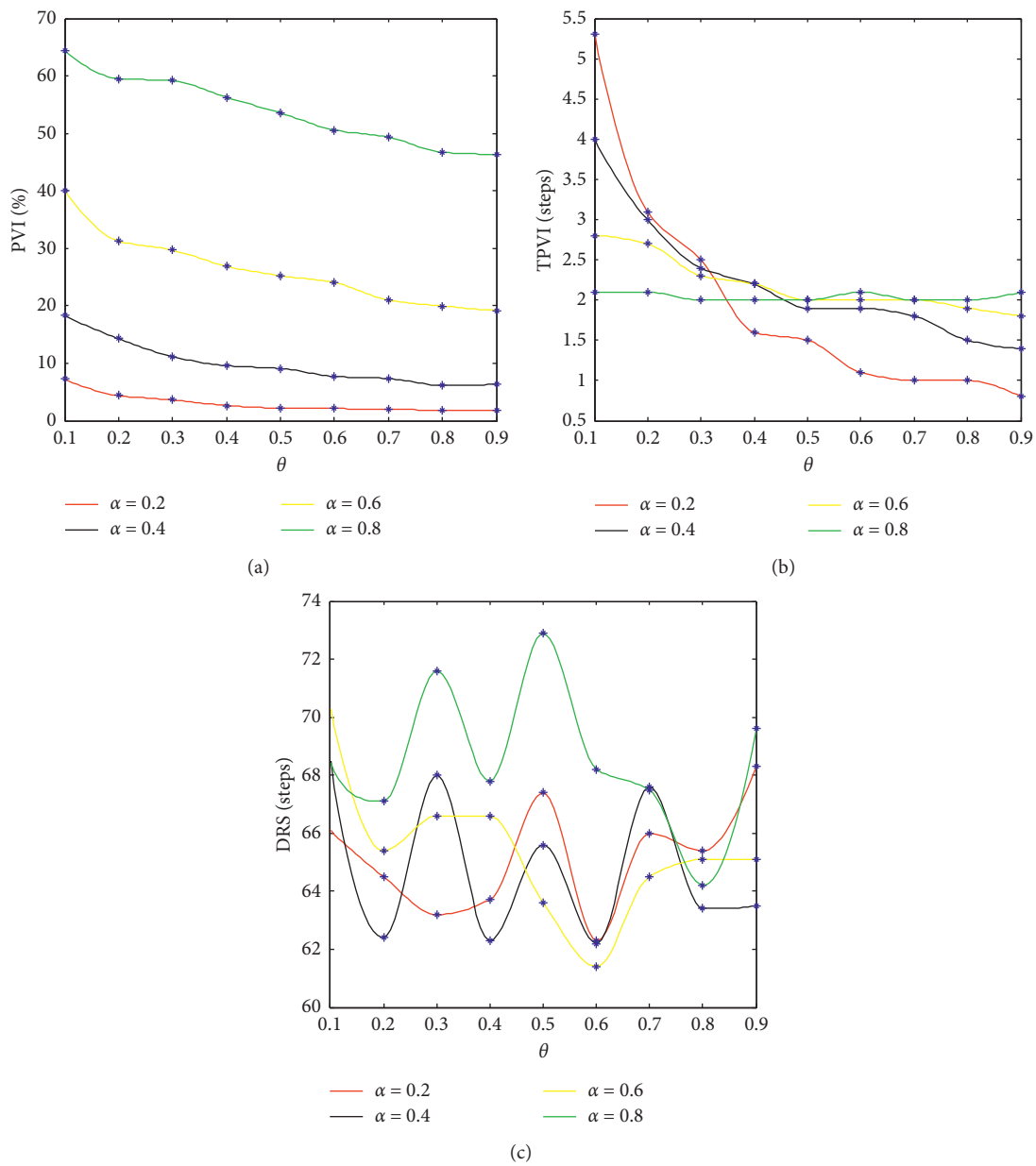


Figure 3: Influence of changes of rumor spreading rate α and self-purification level θ on three indicators PVI, TPVI, and DRS. Simulations are conducted under the condition $a_3 = 0.33$, $b_1 = b_2 = 0.5$, $\beta = 0.1$, and $a_1 = a_2$. (a) PVI change for four values of α ($\alpha = 0.2, 0.4, 0.6, 0.8$), (b) TPVI change, and (c) DRS change.

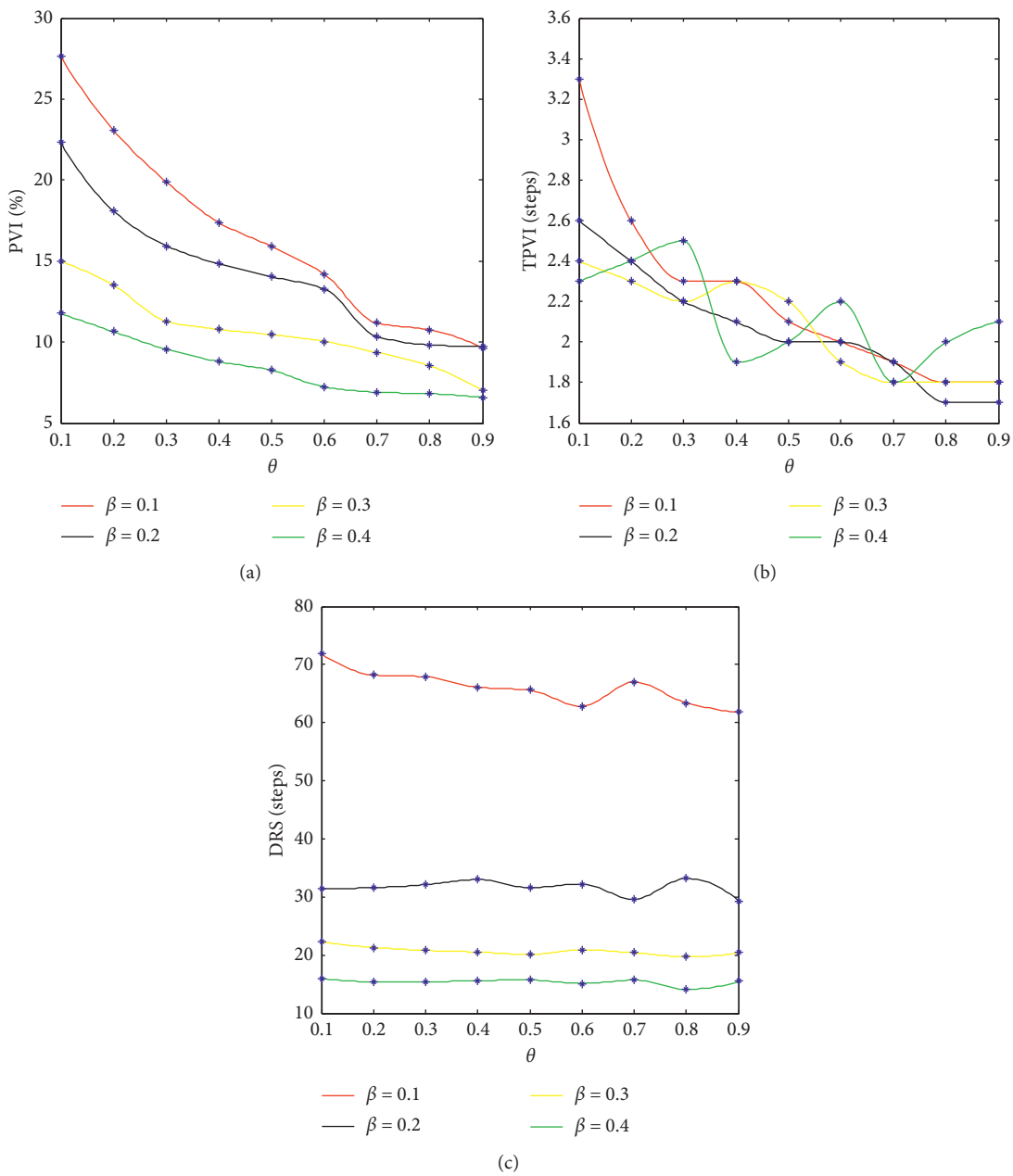


Figure 4: Influence of changes of forgetting rate β and self-purification level θ on above three indicators PVI, TPVI, and DRS. Simulations are conducted under the condition $\alpha = 0.5$, $a_1 = a_2 = 0.25$, $a_3 = 0.33$, and $b_1 = b_2 = 0.5$. (a) PVI change for four values of β ($\beta = 0.1, 0.2, 0.3, 0.4$), (b) TPVI change, and (c) DRS change.

is low, the improvement of self-purification level has a more significant effect to inhibit rumor spreading.

5. Conclusions and Discussion

This paper improves classical rumor spreading models with self-purification mechanism, analyzes the correlation between spreading threshold in the model and its self-purification level theoretically, and conducts numerical simulations to prove that the changes of model parameters, including self-purification level and forgetting rate, exert significant influences on rumor spreading. Novel features and significant results are summarized into three respects:

(1) When building the model, take "the criticizer may become a rumor spreader after hearing rumors for many times" into account and add variable a_3 to convey the probability of a criticizer transforming to a rumor spreader under the influence of rumor spreaders.

(2) Through theoretical analysis about the correlation between spreading threshold in the model and its self-purification level, we find the spreading

threshold in the ISRC rumor spreading model for scale-free networks has no correlation with self-purification mechanism. More precisely, the spreading threshold has no correlation with the probability of a criticizer transforming to a rumor spreader under the influence of rumor spreaders.

(3) Numerical simulations are conducted to study the impact of changes of model parameters on key indicators in process of rumor spreading. The results manifest that changes of model parameters exert significant influences on rumor spreading exactly. On the one hand, at a low rumor spreading rate and high self-purification level, there is a better effect to inhibit rumor spreading in the network, while in the case of high rumor spreading rate, as self-purification level increases, the effect may become worse; on the other hand, whether the network is at a high or low forgetting rate, on the whole, the increase in self-purification level will lead to a tendency to suppress rumor spreading. However, compared with the case of high forgetting rate, when the forgetting rate is low, the improvement of self-purification level has a more significant effect to inhibit rumor spreading.

It should be noted that the authoritative effect of individuals is not considered in this paper, and the forgetting rate is regarded as a fixed value. In real social networks, there are "opinion leaders," which mirror that the more fans a person has, the greater his or her assimilation influence on other individuals is. Considering the forgetting rate tends to change over time in real networks, the authoritative effect of individuals and the dynamic forgetting rate can be combined in the follow-up research to further explore the influence of changes of related parameters on the process of rumor spreading.

Data Availability

The data in this paper are obtained through simulations by NetLogo 6.1.0. The specific code and data are available from the author (fightingzj@163.com) on reasonable request.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Supplementary Materials

The code within Code.txt is run using Netlogo 6.1.0. and adjusted the parameters to get simulation results (results of authors are shown in Data.txt). Figures are plotted according to the data within Data.txt and eventually Figures 3(a)–3(c) and Figures 4(a)–4(c) are obtained in this paper. (*Supplementary Materials*)

References

[1] S. A. Thomas, "Lies, damn lies, and rumors: an analysis of collective efficacy, rumors, and fear in the wake of Katrina," *Sociological Spectrum*, vol. 27, no. 6, pp. 679–703, 2007.

[2] J. Kostka, Y. A. Oswald, and R. Wattenhofer, "Word of mouth: rumor dissemination in social networks," *Lecture Notes in Computer Science*, vol. 5058, pp. 185–196, 2008.

[3] R. H. Knapp, "A psychology of rumor," *Public Opinion Quarterly*, vol. 8, no. 1, pp. 22–37, 1944.

[4] G. W. Allport and L. Postman, *The Psychology of Rumor*, Henry Holt and Company, New York, NY, USA, 1947.

[5] D. J. Daley and D. G. Kendall, "Stochastic rumours," *IMA Journal of Applied Mathematics*, vol. 1, no. 1, pp. 42–55, 1965.

[6] D. P. Maki and M. Thomson, *Mathematical Models and Applications: With Emphasis on the Social, Life and Management Sciences*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1973.

[7] Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of rumor spreading in complex networks," *Physical Review E*, vol. 69, Article ID 066130, 7 pages, 2004.

[8] Y. Moreno, M. Nekovee, and A. Vespignani, "Efficiency and reliability of epidemic data dissemination in complex networks," *Physical Review E*, vol. 69, Article ID 055101, 4 pages, 2004.

[9] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[10] D. H. Zanette, "Critical behavior of propagation on small-world networks," *Physical Review E*, vol. 64, p. 4, Article ID 050901, 2001.

[11] D. H. Zanette, "Dynamics of rumor propagation on small-world networks," *Physical Review E*, vol. 65, p. 9, Article ID 041908, 2002.

[12] Z. Pan, X. Wang, and X. Li, "Simulation investigation on rumor spreading on scale-free network with tunable clustering," *Journal of System Simulation*, vol. 18, no. 8, pp. 2346–2348, 2006.

[13] L. Zhao, X. Qiu, X. Wang, and J. Wang, "Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 4, pp. 987–994, 2013.

[14] L. Zhao, X. Wang, X. Qiu, and J. Wang, "A model for the spread of rumors in Barrat-Barthelemy-Vespignani (BBV) networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 21, pp. 5542–5551, 2013.

[15] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, "Mechanism of rumour spreading in complex social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 374, no. 1, pp. 457–470, 2006.

[16] L. Zhao, Q. Wang, J. Cheng, Y. Chen, J. Wang, and W. Huang, "Rumor spreading model with consideration of forgetting mechanism: a case of online blogging LiveJournal," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 13, pp. 2619–2625, 2011.

[17] X. Wang, L. Zhao, and W. Xie, "Rumor spreading model with variable forgetting rate in scale-free network," *System Engineering Mechanism and Practice*, vol. 35, no. 2, pp. 458–465, 2015.

[18] N. Ren and J. Li, "Rumor-spreading model with a rejection mechanism," *Journal of Yunnan Minzu University (Natural Sciences Edition)*, vol. 28, no. 1, pp. 67–71, 2019.

[19] Y. Zan, J. Wu, P. Li, and Q. Yu, "SICR rumor spreading model in complex networks: counterattack and self-resistance," *Physica A: Statistical Mechanics and Its Applications*, vol. 405, pp. 159–170, 2014.

[20] L. Zhao, X. Wang, J. Wang, X. Qiu, and W. Xie, "Rumor-propagation model with consideration of refutation

mechanism in homogeneous social networks," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 659273, 11 pages, 2014.

[21] X. Wang, L. Zhao, and Z. Wu, "Rumor propagation model with considering refutation mechanism in inhomogeneous networks," *Systems Engineering*, vol. 2015, no. 12, pp. 139–145, 2015.

[22] N. Zhang, H. Huang, B. Su, J. Zhao, and B. Zhang, "Dynamic 8-state ICSAR rumor propagation model considering official rumor refutation," *Physica A: Statistical Mechanics and Its Applications*, vol. 415, pp. 333–346, 2014.

[23] Q. Guo, X. Liu, and Z. Hu, "Effect of factual information release on rumor spreading," *Application Research of Computers*, vol. 31, no. 4, pp. 1031–1034, 2014.

[24] Y. Ma, Y. Zhao, and Y. Qiang, "Conformity effect and authoritative effect of rumor spreading in social network," *Journal of Computer Applications*, vol. 39, no. 1, pp. 232–238, 2019.

[25] L.-L. Xia, G.-P. Jiang, B. Song, and Y.-R. Song, "Rumor spreading model considering hesitating mechanism in complex social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 437, pp. 295–303, 2015.

[26] H. Chen, "Analysis on the dynamics behavior of a rumor transmission model with stochastic perturbation," *Bulletin of Science and Technology*, vol. 32, no. 6, pp. 139–144, 2016.

[27] N. Cai, M. He, Q. Wu, and M. J. Khan, "On almost controllability of dynamical complex networks with noises," *Journal of Systems Science and Complexity*, vol. 32, no. 4, pp. 1125–1139, 2019.

[28] Y. Gu and F. Meng, "Rumor spreading in the online social network: A case of a Renren account," in *Proceedings of the Third International Conference on Digital Manufacturing and Automation*, pp. 751–754, Guilin, China, August 2012.

[29] D. Lu, J. Guo, and Y. He, "Based on the double S model research in WeChat rumors spreading," *Mathematics in Practice and Mechanism*, vol. 47, no. 16, pp. 157–163, 2017.

[30] C. Fan, H. Song, and G. Ding, "Research on an improved SEIR network rumor propagation model," *Journal of Intelligence*, vol. 36, no. 3, pp. 86–91, 2017.

[31] L. Zhu and H. Zhao, "Dynamical behaviours and control measures of rumour-spreading model with consideration of network topology," *International Journal of Systems Science*, vol. 48, no. 10, pp. 2064–2078, 2017.

[32] Z. Song, J. Wang, and R. Shi, "Emergency rumor spreading on Weibo: a research based on scale-free networks," *Journal of Intelligence*, vol. 34, no. 12, pp. 111–115, 2015.

[33] Y. Wan, D. Zhang, and Q. Ren, "Propagation and inhibition of online rumor with considering rumor elimination process," *Acta Physica Sinica (Chinese Edition)*, vol. 64, no. 24, pp. 69–79, 2015.

[34] Y. Gu and L. Xia, "The propagation and inhibition of rumors in online social network," *Acta Physica Sinica (Chinese Edition)*, vol. 61, no. 23, 7 pages, Article ID 238701, 2012.

[35] Z. Xia, Z. Wu, X. Wang, and Y. Xie, "The quantitative simulation of social media rumors self-purification mechanism," *Journal of Modern Information*, vol. 39, no. 3, pp. 103–110, 2019.

[36] P. Ozturk, H. Li, and Y. Sakamoto, "Combating rumor spread on social media: the effectiveness of refutation and warning," in *Proceedings of the IEEE 2015 48th Hawaii International Conference on System Sciences*, pp. 2406–2414, Washington, DC, USA, January 2015.

[37] T. Ge, Z. Xia, and Y. Zhai, "Analysis of collective intelligence effect on social media false information transmission," *Journal of Intelligence*, vol. 34, no. 7, pp. 148–152, 2015.

[38] Y. Tanaka, Y. Sakamoto, and T. Matsuka, "Toward a social-technological system that inactivates false rumors through the critical thinking of crowds," in *Proceedings of the IEEE 2013 46th Hawaii International Conference on System Sciences*, pp. 649–658, Washington, DC, USA, January 2013.

[39] X. Wang and L. Zhao, "Rumor spreading model with skepticism mechanism in social networks," *Journal of University of Shanghai for Science and Technology*, vol. 34, no. 5, pp. 424–428, 2012.

[40] L. Zhao, J. Wang, Y. Chen, Q. Wang, J. Cheng, and H. Cui, "SIHR rumor spreading model in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 7, pp. 2444–2453, 2012.

[41] H. Yuan and Y. Xie, "Research on rumormonger in public events based on content analysis of 118 Internet rumors in public events," *Shanghai Journalism Review*, vol. 2015, no. 5, pp. 58–65, 2015.

[42] X. Qiu, L. Zhao, J. Wang, X. Wang, and Q. Wang, "Effects of time-dependent diffusion behaviors on the rumor spreading in social networks," *Physics Letters A*, vol. 380, no. 24, pp. 2054–2063, 2016.

Hindawi
Complexity
Volume 2021, Article ID 8897105, 12 pages
https://doi.org/10.1155/2021/8897105

Research Article

Construction and Analysis of Emotion Computing Model Based on LSTM

Huiping Jiang, Rui Jiao, Zequn Wang, Ting Zhang, and Licheng Wu

Brain Cognitive Computing Lab, School of Information and Engineering, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Licheng Wu; wulicheng@tsinghua.edu.cn

Received 17 September 2020; Revised 25 November 2020; Accepted 21 January 2021; Published 3 February 2021

Academic Editor: Ning Cai

The electroencephalogram (EEG) is the most common method used to study emotions and capture electrical brain activity changes. Long short-term memory (LSTM) processes the temporal characteristics of data and is mostly used for emotional text and speech recognition. Since an EEG involves a time series signal, this article mainly studied the introduction of LSTM for emotional EEG recognition. First, an ALL-LSTM model with a four-layered LSTM network was established in which the average accuracy rate for emotional classification reached 86.48%. Second, four EEG characteristics were extracted via the wavelet transform (WT) using the LSTM-based sentiment classification network. The experimental results showed that the best average classification accuracy of these four features was 73.48%. This was 13% lower than in the ALL-LSTM model, indicating that inappropriate feature extraction methods could destroy the timing of EEG signals. LSTM can be used to thoroughly examine EEG signal timing and preprocessed EEG data. The accuracy and stability of the ALL-LSTM model are significantly superior to those of the WT-LSTM model. The result showed that the process of emotion generation based on EEG is sequential. Compared with EEG emotion extraction using WT, the raw EEG signal's timing is more suitable for the LSTM network.

1. Introduction

Suppose a high level of human-computer interaction is to be achieved. In that case, it is essential for computers to effectively recognize human emotions, which is significantly useful for realizing a brain-computer interface and intelligent machines.

People expect computers that are easier to control and anticipate a gradual change from human-operated computers to computer-aided people, signaling a transition from passive cognitive to active. The concept of affective computing was proposed by Professor Picard of the MIT Media Lab in 1997. She indicated [1] that sentiment calculation involves certain techniques to classify and interpret emotions according to specific data. Zhou [2] believes that the purpose of emotional computing is to establish a harmonious human-machine environment by providing computers with the ability to recognize, understand, express, and adapt to human emotions, equipping computers with higher and comprehensive intelligence.

Scientific evidence shows that the appearance and development of emotions occur parallel to the brain's evolution, while brain development corresponds with the differentiation and development of facial expressions [3]. Together, the nervous and endocrine systems determine the physiological changes in the human body, the signals of which are challenging to control artificially. Many methods are employed for emotion recognition, such as those based on facial expressions and physiological signals.

Using physiological signals for the recognition of emotions usually yields accurate and objective results. Furthermore, this technique aids in the safety improvement of equipment used by reducing the security risks associated with emotional factors. As a specific physiological modality, EEG signals are exceptionally valuable in emotional classification, and this method has been extensively studied. The EEG detection instrument is inexpensive and exhibits a high time resolution and a bearable space resolution. Besides these advantages, an EEG obtains more detailed, complex information in a noninvasive manner for emotion

recognition. The data cannot be deliberately modified or concealed, making EEG-based emotion recognition more effective and reliable [4].

Human-computer emotional interaction can render many computer applications more convenient and feasible. The computer can use human physiological signals to make judgments without the need for cumbersome behavioral responses. This is of considerable significance to people with disabilities involving the facial muscular system or limbs.

Many machine-learning and pattern-recognition algorithms are applied to EEG-based emotion recognition, but the generation of emotions remains a complex cognitive activity. The mechanism and process of emotion generation are still being investigated, and applying EEG signals for emotion calculation shows significant potential.

Since EEG signals are nonstationary and highly random, it is challenging to extract EEG features related to a particular cognitive task. An essential feature of the extraction method is to minimize the loss of the raw signal and simplify the raw dataset. Therefore, feature extraction aims to reduce the complexity of the application to render information processing more cost-efficient. Since Dietch first used a Fourier transform for EEG analysis in 1932, classical methods, such as frequency domain analysis, time-domain analysis, and WT, were introduced [5, 6]. Because WT is more suitable for analyzing nonstationary signals, as well as the signal in the time and frequency domain, it can be used to resolve the contradiction between the time and frequency resolutions [7].

The EEG classification challenge is essentially a pattern-recognition problem. The current methods used for classifying EEG signals include linear discriminant analysis, support vector machine (SVM), and deep learning models [8, 9]. Deep learning is a general term for this type of neural network learning algorithm depth and has attracted significant attention in recent years [10]. Deep learning models, such as the autoencoder (AE), deep belief networks (DBN), convolutional neural networks (CNN), and recurrent neural networks (RNN), are widely used [11–13]. An unsupervised DBN was applied for the depth level feature extraction from fused observation signals. Experiments involving a public multimodal physiological signal dataset show that these models significantly increase the emotion recognition rate accuracy [14]. A novel computer model [15] is presented for the EEG-based screening of depression using a CNN. The algorithm attained 93.5% and 96.0% accuracy using EEG signals from the left and right hemispheres, respectively. Results reveal that the EEG signals from the right hemisphere are more distinctive than those from the left hemisphere in depression. A compact CNN is introduced for EEG-based brain-computer interface (BCI) [16], allowing for EEG feature extraction. EEGNet can better generalize across paradigms than the reference algorithms when only limited data is available across all tested programs while achieving comparable high performance. These techniques and others have been applied to explore EEG in machine learning and deep learning, achieving positive results. However, EEG signals are composed of multilead signals and contain important time-frequency information. Without sufficient time and frequency domain information, it is difficult to obtain a good classification result. The recurrent structure of the RNN can be used to obtain contextual information about the time series [17]. However, when the training sequence is too long, the traditional RNN faces the problem of gradient disappearance or explosion due to its structural design. Therefore, this paper proposes a new emotion recognition model based on the LSTM network to resolve this issue. Compared with standard RNN, LSTM performs better in longer sequences and can be applied widely in the technology field. LSTM-based systems can perform image analysis, speech recognition, and disease prediction [18, 19]. However, establishing an LSTM emotion model based on EEG and its application in emotion recognition requires further study.

The emotion recognition method based on EEG generally proceeds as follows. First, the EEG signals induced by specific images or videos corresponding to emotions are read. Second, models are established by learning the EEG samplings. Finally, the models are applied to the real systems.

In conclusion, one of the most critical factors of emotion recognition based on EEG is acquiring EEG ground truth data for training the models, substantially affecting the accuracy rate and stationary ability regardless of which model is adopted. Furthermore, according to the temporal characteristics of emotion induction, it can improve emotion classification accuracy by constructing an emotion recognition system with a sequential effect.

This paper aims to establish a corresponding neural network based on the experimental EEG data, fully utilizing the sequential information implicit in the EEG signals, and building a suitable deep learning neural network. The method proposed in this paper mainly solves two problems: one is the effect of the feature extraction of the preprocessed EEG signal on emotion recognition before establishing a related model and the other is the influence of the model on emotion recognition compared with existing research.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 discusses the overall framework design, experimental dataset, experimental environment, and classification. Sections 4 and 5 discuss the results for different models and compare them.

2. Related Work

2.1. Emotion Features Based on the EEG Signal. The EEG signals exhibit various frequencies. Neuroscientists have divided them into frequency bands, each of which is responsible for specific brain activity. The different brainwaves and the activity responsible for them are as follows:

Delta (0.5–4 Hz): its amplitude is about 0–200 μV, which only occurs during sleep, deep anesthesia, hypoxia, or brain lesions.

Theta (4–8 Hz): its amplitude is about 100–150 μV, which appears during drowsiness and corresponds to daydreaming, drowsiness, or sleep.

Alpha (8–12 Hz): its amplitude is about 5–20 μV, corresponding to the resting state of the brain.

Beta (12–30 Hz): it is associated with active, task-oriented, busy or anxious thinking, and active concentration.

Gamma (>30 Hz): it occurs when different populations of neurons work together to perform demanding cognitive or motor functions.

The methods for EEG signal feature extraction are relatively mature. The common emotional features include the following three categories: the time-domain features, such as the mean, standard deviation, skewness, peak amplitude, variance, skewness, and kurtosis. The second is the frequency domain features, including the features extracted via the Fourier transform and those extracted via the parameter model (such as AR, Ma, ARMA, and the harmonic signal model). Finally, there are the time-frequency features, such as the short-time Fourier transform, WT, and nonlinear dynamic features.

The WT decomposes the input signals into various constituting small range frequency bands. This is done by obtaining the approximation and detail coefficients via multiple-level decomposition.

The key to the efficient extraction of EEG features is to choose the appropriate wavelet base. Standard wavelet bases include Daubechies (dbN) wavelet, Symlets (symN) wavelet, and Coiflet (coifN) wavelet. There is no unified standard for the selected wavelet base and it primarily relies on the classification accuracy. During early research, EEG emotion classification involving different wavelet bases was shaped according to CNN [20]. The results show that the Sym8 wavelet could better classify emotion based on the raw EEG signal. Therefore, this study used Sym8 wavelet for further experimentation.

After the WT of the EEG signal, the wavelet coefficients of each layer of the frequency band were obtained. Still, the wavelet coefficients cannot be sent directly to the classifier as features and require further processing to extract the EEG features. The features selected in this paper include the band energy (E), the band energy ratio (REE), the logarithm of the band energy ratio (LREE), and the differential entropy (DE) and are described below.

The E refers to the energy of each frequency band after WT and is obtained by square-summing the coefficients of each frequency band. The solution formula is shown as follows:

$$E_i = \sum_{j=1}^{n_i} d_{ij}^2, \tag{1}$$

where E_i is the energy of the i-th band, n_i is the number of coefficients decomposed by the i-th layer, and d_{ij} is the j-th wavelet coefficient of the i-th layer.

The REE refers to the ratio of each layer of energy to the total energy and is expressed as follows:

$$\mathrm{REE}_i = \frac{E_i}{\sum_{j=1}^{n} E_j}, \tag{2}$$

where REE_i is the band energy ratio of the i-th band and n is the number of bands.

The LREE for each band is based on 10 and is expressed as follows:

$$\mathrm{LREE}_i = \log_{10}\mathrm{REE}_i, \tag{3}$$

where LREE_i represents the logarithm of the energy ratio of the i-th band.

If the signal obeys a different distribution, the DE is solved differently. It is assumed that the acquired EEG signal, X, is affected by the Gaussian distribution, as shown in equations (4)–(6):

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left((x-m)^2/2\sigma^2\right)}, \quad -\infty < x < \infty, \tag{4}$$

$$m = E(x) = \int_{-\infty}^{\infty} x p(x) \mathrm{d}x, \tag{5}$$

$$\sigma^2 = E\left[(x-m)^2\right] = \int_{-\infty}^{\infty} (x-m)^2 \, p(x) \mathrm{d}x, \tag{6}$$

The solution process of the DE is shown as follows:

$$\begin{aligned} h(x) &= -\int_{-\infty}^{\infty} p(x) \log p(x) \mathrm{d}x = -\int_{-\infty}^{\infty} p(x) \log p(x) \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left((x-m)^2/2\sigma^2\right)}\right) \mathrm{d}x \\ &= \log\sqrt{2\pi\sigma^2} \int_{-\infty}^{\infty} p(x) \mathrm{d}x + \log e \int_{-\infty}^{\infty} p(x) \frac{(x-m)^2}{2\sigma^2} = \log\sqrt{2\pi\sigma^2} + \frac{1}{2}\log e = \frac{1}{2}\log\left(2\pi e \sigma^2\right). \end{aligned} \tag{7}$$

This formula indicates that the key to solving DE is acquiring the EEG signal variance, which is approximately the same as the average of the energy of the EEG signal in each band. In practical applications, the E value is commonly used as a logarithm instead of DE. The simplified formula for DE is expressed by equation (8):

$$\mathrm{DE}_i = \log_{10} E_i, \tag{8}$$

where DE_i represents differential entropy.

2.2. Deep Learning for EEG Analysis. Over the past few years, traditional machine learning technology (i.e., nondeep

learning algorithm) has been the only feasible EEG analysis option. It continues to be widely used in combination with various feature extraction and feature selection algorithms [21–24].

As a relatively new trend, the deep learning algorithm has been applied in medical image and signal processing due to the improvement and availability of computing power and big data. In most cases, its performance exceeds the rates that have been previously achieved with traditional machine learning techniques [25].

Many methods have been proposed for studying appropriate computational models for emotion recognition using EEG signals. Various deep learning structures have been used to classify EEG signals to solve different recognition tasks. Generally, most existing EEG research based on deep learning can be summarized into two categories. The first is based on EEG signals input to the network. The second type is based on features extracted from EEG signals as input to the network.

An EEG analysis task requires the developed model to capture private information from EEG signals. Traditional machine learning methods need to design and extract the features of EEG signals manually. The redundancy of the features is exceedingly high and does not consider the temporal dynamics of the EEG signals crucial for emotion recognition.

Tripathi et al. [26] proposed an emotion recognition method based on CNN from EEG signals in the Database for Emotion Analysis using Physiological Signals (DEAP) dataset. They explored two different neural models: a simple deep neural network and a CNN. The performance of the latter is 4.96% higher than that in state-of-the-art techniques.

Shawky et al. [27] presented a three-dimensional CNN approach for recognizing emotions from multichannel EEG signals. They developed a data enhancement phase to improve the performance of their 3D CNN model. They achieved 87.44% accuracy for valence and 88.49% for arousal.

Moon et al. [28] applied CNN to recognize emotion based on EEG. They employed brain connectivity features to explain the synchronous activation of different brain regions, an approach that has not been used in previous studies. Therefore, their method effectively captures the asymmetric brain activity patterns, playing a vital role in emotion recognition.

LSTM [29–31] is one form of RNN that overcomes the problem of exploding and vanishing gradients. The building blocks of LSTM include a cell, an input gate, an output gate, and a forget gate. The cell is responsible for handling long-term dependency while the three gates regulate the flow of values between the different layers of the LSTM network.

The innovation of LSTM networks compared to traditional RNNs is the inclusion of "gates" to solve the vanishing gradient problem and allow the algorithm to control more precisely what information needs to be retained in its memory and what must be removed [32, 33]. By controlling the learning rate with the three gates (i.e., input gate, forget gate, and output gate), the LSTM network can better adjust to large data series sequences than RNNs and other deep learning techniques. Considering that EEG signals are essentially highly dynamic, nonlinear time-series data, LSTM networks are better than CNN in isolating the temporal characteristics of brain activity during different states as reported in various applications, such as emotion recognition, confusion estimation, and estimation prediction [26, 34–36]. Despite their inherent advantages in EEG analysis, LSTM models have not been examined combined with emotion feature extraction.

This paper analyzes an emotion recognition framework based on the LSTM. First, the multichannel EEG signal is divided into multiple segments, and the time domain, frequency domain, and nonlinear dynamic features are extracted from each segment of the signal to form a feature sequence along with time, respectively. Each feature sequence consists of characteristics representing specific feature information of the signal. Second, an LSTM neural network is used to obtain the time dynamic information from various feature sequences and make the final emotion prediction.

3. Methods

3.1. Framework Design. Figure 1 shows that the framework includes three parts: source signal processing, feature extraction, and sentiment classification.

3.2. Experimental Dataset. Here, 12 videos were selected as emotion-evoking stimuli to cover the entire emotional spectrum. Six of 12 videos were excerpts from movies and were chosen based on the preliminary study. During the initial investigation, the participants self-assessed their emotions by reporting their arousal feeling (ranging from calm to excited/activated) and valence (ranging from unpleasant to pleasant) on a nine-point scale. Sam Manikins were shown to facilitate the self-assessments of valence and arousal [37]. Ultimately, six videos between 110 s and 120 s long were selected to be shown. Psychologists recommend videos from 1 min to 10 min long to elicit a single emotion [38]. Here, the video clips were kept as short as possible to avoid multiple emotions or habituation to the stimuli [39] while keeping them long enough to observe the effect—data collection. The stimulus file is shown in Table 1.

Twenty younger adults (11 women), mainly students at the Minzu University of China (mean age: 21.4 years, range: 20–23), participated in the experiment. They were paid 150 RMB per hour for their participation. Participants were right-handed as assessed by a German version of the Edinburgh handedness inventory [40] and had a normal or corrected-to-normal vision. None of them reported neurological or psychological disorders. All participants were fully aware of the purpose of the study. The study was approved by the Local Ethics Committee (Minzu University of China, Beijing, ECMUC2019008CO).

3.3. Experiment Environment. In the cognitive brain laboratory, the electrode Synamps2 amplifier and Scan4.5 software developed by the Neuroscan Company, a cap with

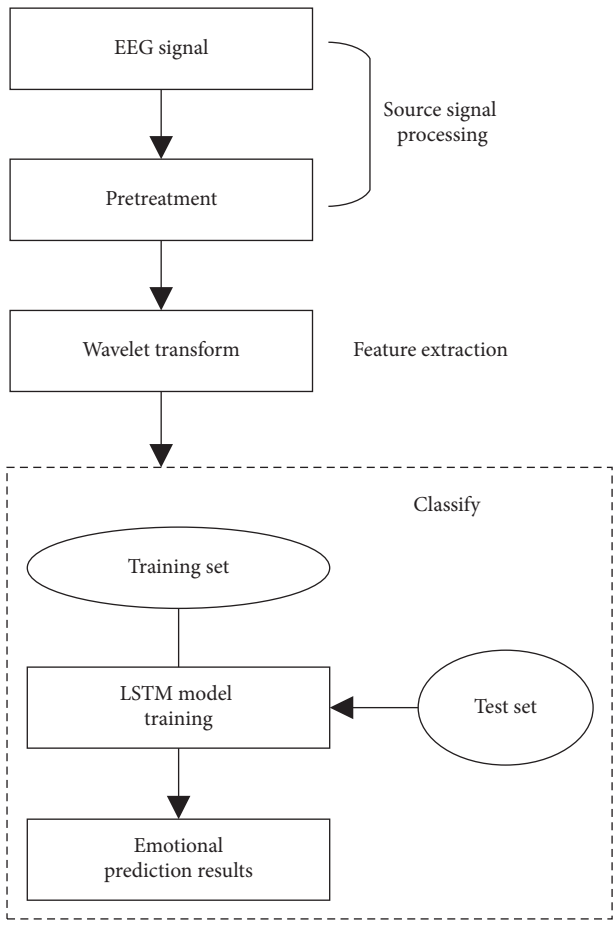


Figure 1: Framework design.

Table 1: The movie list.

Movie slices sources	Label
Martial arts biography	Positive
Love apartments	Positive
Pet funny video	Positive
Death is coming	Negative
The chainsaw cry	Negative
Teacher's grace	Negative

64 electrodes, and a computer (labeled as computer 1) were used to collect the EEG signals, while a webcam camera and another computer (labeled as computer 2) were used for documenting the facial expressions. A dedicated server was employed to generate the induction files developed with the E-prime application software while coordinating computer 1 and computer 2 to obtain the EEG records and facial expressions simultaneously.

The Neuralscan-64 system was selected for EEG acquisition, while the electrode cap collected 64-channel EEG signals. The EEG sampling frequency was set to 1000 HZ, which fulfilled the requirements of rapidly changing EEG signals. The electrode distribution of the electrode cap was based on the currently used 10/20 system electrode placement method. Figure 2 shows the specific experimental process.

Furthermore, to obtain a better mood in the wake of the formal experiment, the video, as well as positive and negative video capture emotions samples, was randomly presented. Each test lasted about 25 min in total, and the screen displayed a 3000 ms gaze point "+" to prompt the participant to focus, immediately playing a stimulus video. The video was displayed for about 3 min. After the video had completed playing, the subject had to provide feedback regarding the subjective feelings after watching the material by pressing a button. The participants had a choice between three alternatives: "positive," "neutral," and "negative." After providing feedback, a black screen appeared for 7000 ms to clear the participant's thoughts and reduce mutual interference between videos. After the experiment was completed, the EEG data samples, including the positive emotions and negative emotions, were mixed, and part of the EEG data was proportionally selected as the training set for the model, while the remaining part of the EEG data denoted the test set.

The EEG signal is fragile and extremely susceptible to the internal or external environment during the measurement process. This rendered the collected signal unreliable, while it was subject to interference by many electrical activities not originating from the brain. These interferences are known as artifacts. Common artifacts originate from electrooculograms, electrocardiograms, electromyography, and electrode

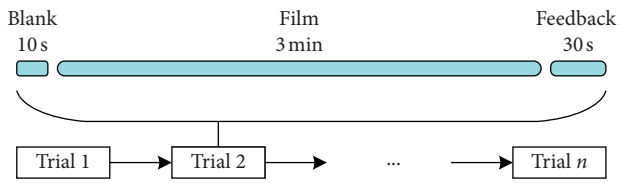


Figure 2: Emotional induction experimental process.

movements. The experimental environment used in the acquisition of EEG signals can be controlled artificially. However, it is challenging to artificially intervene in the human body's unconscious activities. Therefore, the unsatisfactory EEG area was deleted, while the electrooculogram artifacts and other interferences were removed. Digital filtering was performed to preprocess the EEG signals in preparation for the next step of feature extraction and EEG signal classification.

3.4. Classifier. The manuscript uses the LSTM network to classify the emotions obtained via the EEG. Two classification models were compared based on LSTM. One was to perform feature extraction and provide input to the LSTM network for classification, while the other used LSTM directly for classification.

4. Result

4.1. The Construction of an LSTM Model Based on EEG. The LSTM-based emotion classification model established in this paper consisted of four layers: the input layer, the LSTM layer, the fully connected layer, and the output layer. In establishing the LSTM network, different parameters determined different network structures. It was necessary to select the appropriate number of layers and determine the number of hidden nodes in each layer. These parameters directly determined the training speed of the network, the level of classification accuracy, and the stability of the network.

The advantage of deep learning is its colossal network scale. However, as the network scale expands, more computing resources are required. Too many nodes in the hidden layer may cause the training speed to decline or even overfit. The experiments indicated that when the number of hidden layer units was below a specific value, it became challenging to fit the model. It required the design of a more streamlined model structure on the premise of meeting the accuracy requirements. Therefore, as few nodes in the hidden layer and the number of LSTM layers as possible should be selected while ensuring accuracy. Due to the LSTM structure, the neuron's state continuously changed as the input increased, while the historical information of the data was saved, and the hidden layer could utilize the output of each step as the next input. Experiments were conducted on the single-layered and multilayered LSTM structures. After adjustment, it was found that the classification of the multilayered LSTM was superior. The multilayered LSTM sent the output values of the front-end LSTM as input to the back-end LSTM, while the LSTM could be stacked infinitely in a similar way. Finally, an LSTM-based emotion classification model was established to classify the features of wavelet extraction. The model consisted of four layers, each of which exhibited 32 hidden nodes.

Because of the distinct individual differences in EEG, the EEG classification tasks in this paper are based on the EEG signals collected by individual people. The single-person EEG data is divided into a training set and a test set, where the model was trained using the former and tested using the latter.

Via continuous debugging, the Adam algorithm is used for parameter optimization and denotes an adaptive moment estimation method to calculate the adaptive learning rate for each parameter. In practical applications, the convergence speed of the network can be accelerated, which achieves excellent results. Compared with other adaptive learning rate algorithms, the convergence speed is faster, obtaining a more effective learning effect. The learning rate is set to 0.005. During the process of neural network training, regularization or Dropout is usually used to avoid overfitting. The model uses the Dropout method, with the parameter value set to 0.5. During the training process, this section uses the Batch technology. During the experiment, the batch was set to 16, 32, and 64, respectively. The analysis found that an appropriate increase in batch size could improve memory utilization and enhance the running speed, each finalizing a batch size of 64 training examples. Google's Tensor Flow framework implements an LSTM network model. The specific parameters are shown in Table 2.

4.2. ALL-LSTM. The LSTM model's structural properties allow it to learn the timing characteristics of the data, facilitating long-term memory. Therefore, the ALL-LSTM model does not perform artificial feature extraction from the EEG data but selects the preprocessed full-scale EEG information and sends it directly to the LSTM-based emotion classification model shown in Figure 3.

The ALL-LSTM emotion classification model consists of four layers. The first layer takes the preprocessed full EEG sequence as input. The second is the LSTM layer, which extracts contextual related features from the input EEG sequence, such as the time-domain information. The third is a fully connected layer, which is used to integrate the features extracted by the LSTM layer. It is a linear combination of the output of all LSTM units during the last time step. The function of this layer is to combine different feature-dynamic information learned from each LSTM unit. The output of this layer represents the input to the SoftMax layer to predict the emotional state. The fourth is the output layer, producing the recognized emotion category.

The dataset used in this paper is represented by the raw EEG data collected from subjects watching the stimulation material. Each subject's EEG data was divided into 10 ms (10 sampling points), obtaining 100440 EEG data from each. Each EEG data dimension collected via the 64-electrode cap presented a matrix of 64×10. According to the LSTM principle, each column of the matrix (the voltage value collected by the 64-lead electrode) was selected as the data read in one step. Each row of the matrix

Table 2: LSTM emotion classification model specific parameter settings.

Name	Parameter
Learning rate	0.005
Input dimension	64∗5, 64∗10
Output category	2
Batch	64
Dropout	0.5
Hidden node number	32
LSTM layer number	4

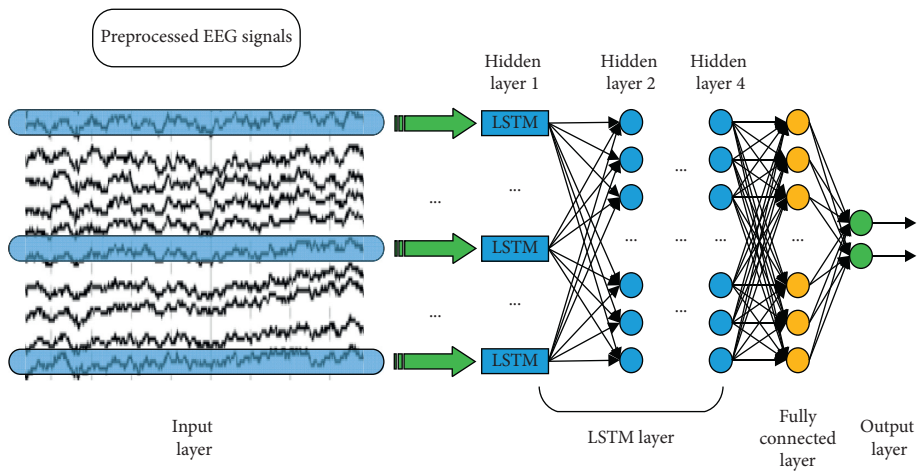


Figure 3: ALL-LSTM model.

(10 sampling points with a duration of 10 ms) was considered the time step number. One of the advantages of intercepting EEG data in this way is that the amount of data is sufficient since deep learning requires an abundant amount of data as a basis. Seventy-five thousand were selected as the training set, accounting for about 75%, and 25440 were selected as the test set, accounting for about 25%. The ratio of the training set to the test set was about 3 : 1.

All the EEG training data obtained from a single person was sent to the LSTM emotion classification model in batch polls for training, a process known as completing an epoch. After each epoch, the loss and accuracy rate of the training set based on the existing training model was provided. After 200 epochs, the model tended to converge, and the training was completed. The training process is shown in Figure 4, in which the abscissa represents the training number, and the ordinate denotes the accuracy during the training.

The test set was sent to the trained model. The final classification results of the eight subjects are shown in Table 3.

These analytical results indicate that the ALL-LSTM model classification's average accuracy is 86.48%, and the variance is 0.0039.

4.3. WT-LSTM. The WT-LSTM was established using EEG data after performing the WT emotion recognition process based on the LSTM emotion classification model. This paper examines four standard, efficient wavelet features, namely, E, REE, LREE, and DE. The same source data is used throughout the experiment and sent to the LSTM emotion classification model using the same parameters. The classification results corresponding to these four factors were obtained, and the EEG features that were most compatible with the LSTM network were identified. Therefore, the classification results obtained by changing only the feature parameters can clearly reflect the efficacy of the features.

Figure 5 shows that the sentiment classification model based on the WT-LSTM established in this section consists of four layers. First, the input layer considers the extracted wavelet features as input. The second layer represents the LSTM layer, extracting context-related features from the input information. The third layer is a fully connected layer, which is used to integrate the features extracted by the LSTM layer and convert the information from the LSTM into the desired output. The fourth layer is the output layer, which aims to output the recognized emotion categories.

The EEG data represent the dataset used in this article that collected the subjects watching the stimulation material simultaneously. Because the EEG signals recorded too many potential values, which were doped with some unwanted potential values, it was necessary to segment each person's EEG signal samples. However, since there is no clear conclusion regarding the time range of human emotional change, the length of each sample selected in this paper is 3000 ms (3000 sampling points). Each 3,000 ms was divided into one EEG data unit. Since the electrode cap used in the experiment was a 64-lead, the dimension of

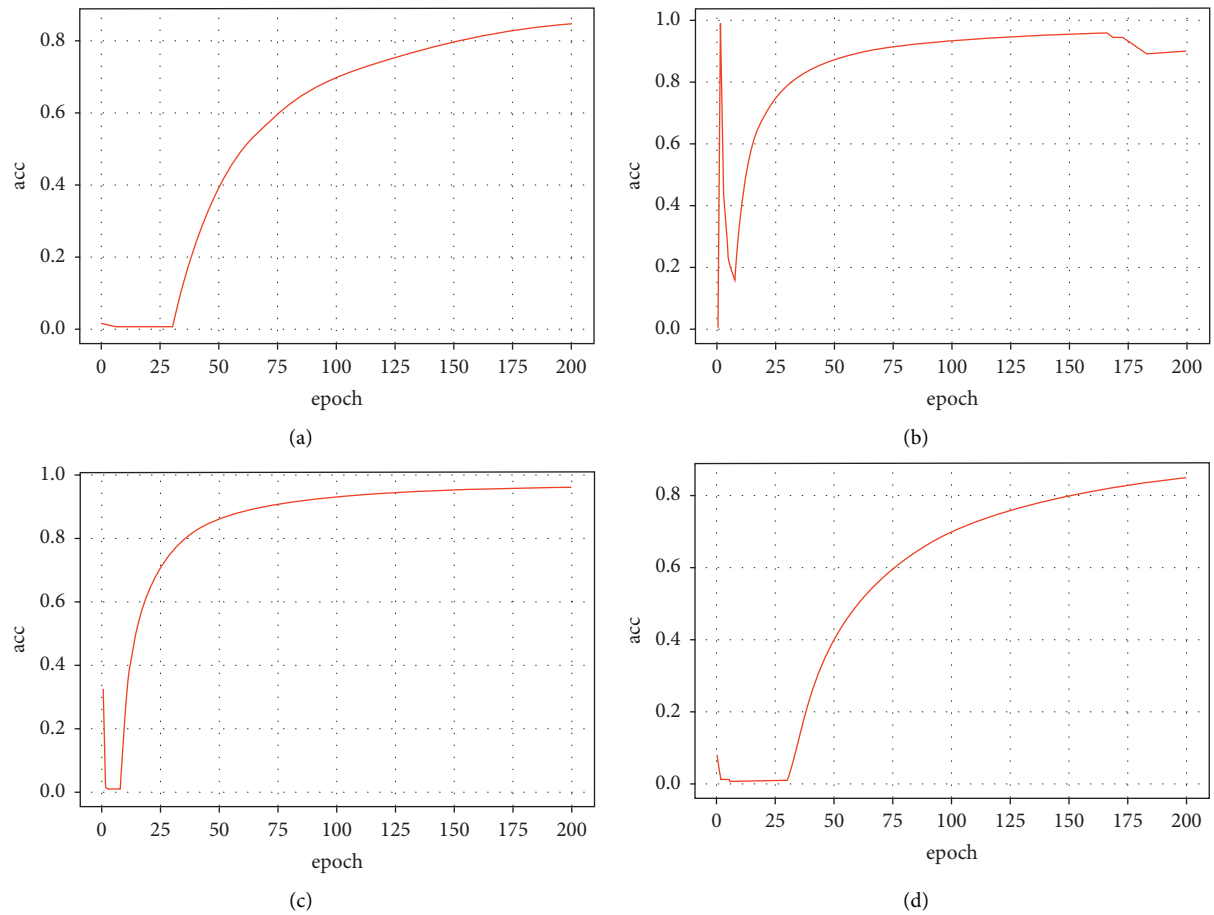


Figure 4: The random training process of four participants.

Table 3: ALL-LSTM model classification results.

Experimenter	Accuracy %
Subject 1	84.72
Subject 2	84.57
Subject 3	96.04
Subject 4	84.57
Subject 5	90.36
Subject 6	93.02
Subject 7	81.91
Subject 8	76.64
AVG	86.48
VAR	0.0039

each EEG data unit was 64×3000. Of the total 348 EEG data units obtained after each participant's segmentation, 260 were randomly selected as the training set, accounting for about 75%, and 88 were chosen as the test set, accounting for about 25%, while the ratio of the data in the training set to the test set was approximately 3 : 1. After the five-layered WT, each raw EEG data dimension was reduced to a matrix of 64×5. According to LSTM principles, this paper selected each column of the matrix as the data read in one step and considered each row of the matrix as the number of time steps.

The frequency E, frequency REE, frequency LREE, and DE were extracted as EEG features for training. Figure 6 shows a training process using the frequency LREE as a feature. The abscissa represents the training number in the figure, while the ordinate denotes the accuracy rate during training. Each image randomly shows the training process of four participants.

After 40 training epochs, the model tended to converge. The classification results of the eight subjects and the average classification accuracy and variance corresponding to the four characteristics are shown in Table 4.

It can be concluded from the above table that the classification accuracy of the four features of Subject 1 is low, and the classification accuracy of the four features of Subject 6 exceeds 80%. The classification accuracy is significantly affected by individual differences. Analysis of the four characteristics indicated an accuracy rate of 50% or less exhibiting one E, two REE, two DE, and zero LREE. An accuracy rate of 80% or more displays one E, one REE, four DE, and three LREE. An accuracy rate of 90% or more shows zero E, zero REE, zero DE, and one LREE. Furthermore, when the frequency LREE is used as the feature, the average classification accuracy is the highest, and the variance value is the lowest. Therefore, the use of LREE as a feature leads to

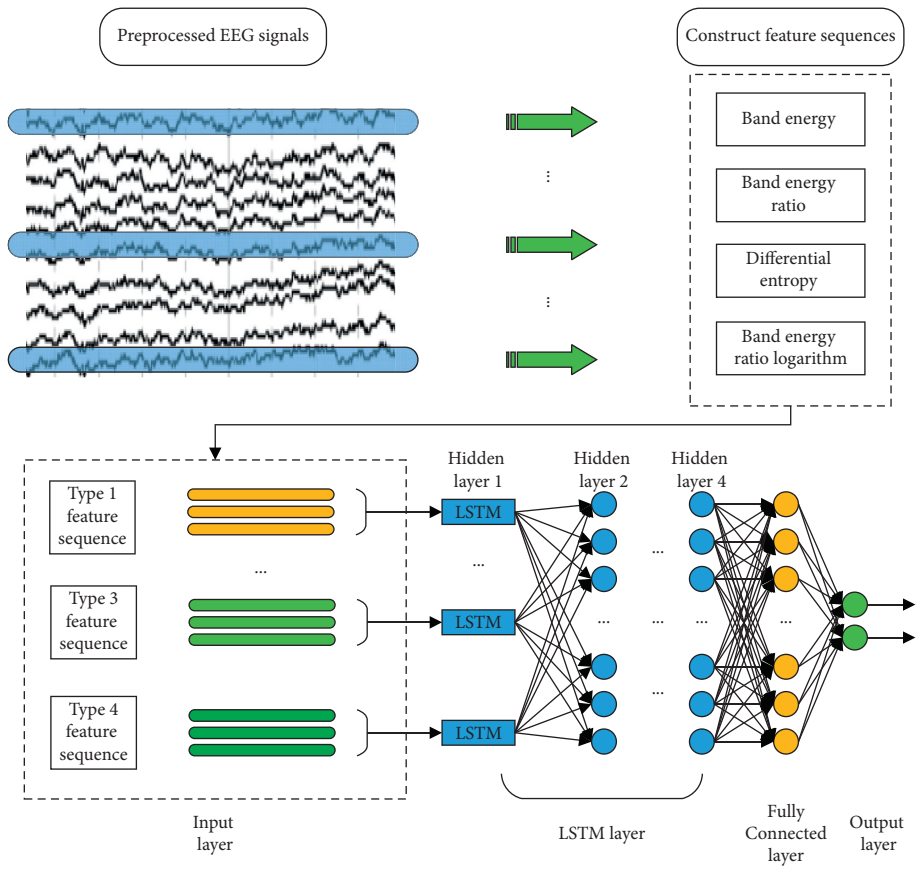


Figure 5: WT-LSTM model.

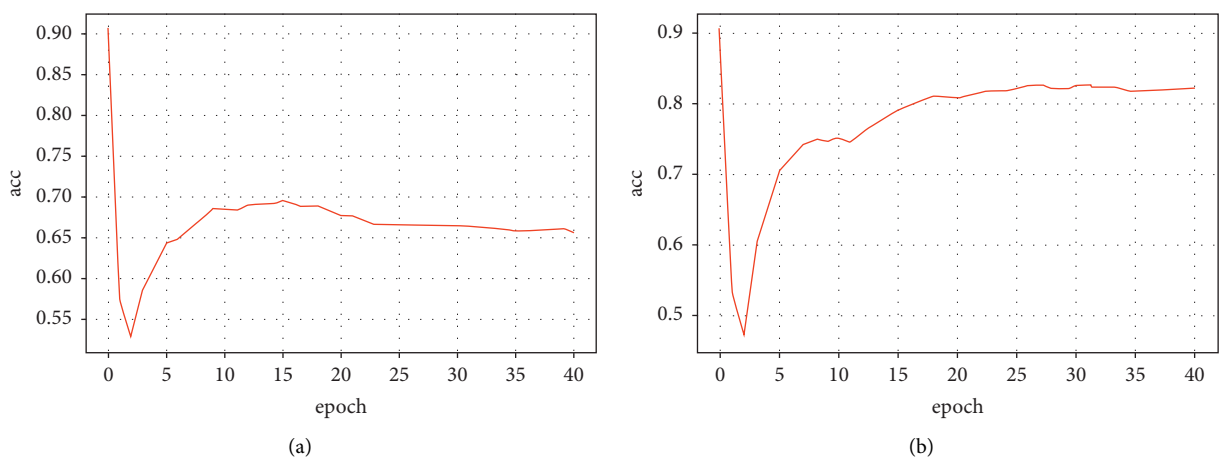
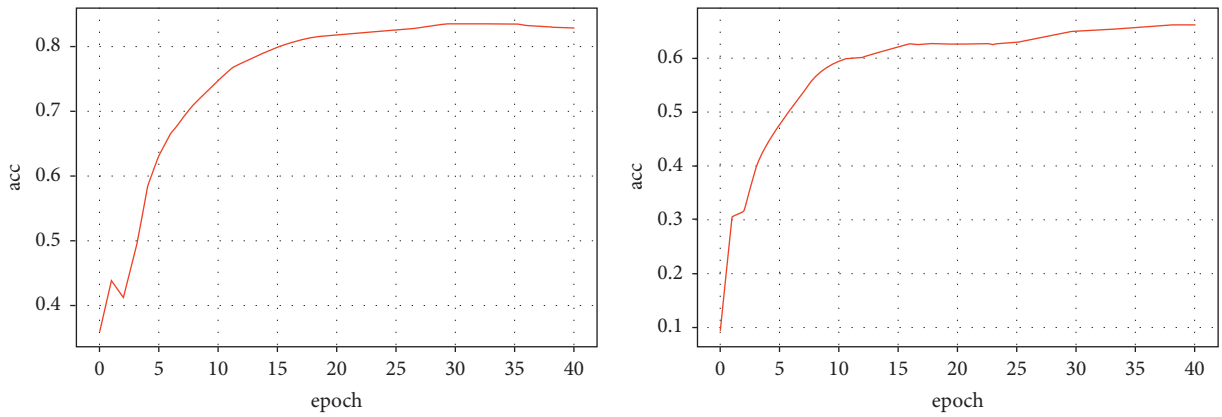


Figure 6: Continued.



(c) (d)

Figure 6: The training process of the LREE as an EEG feature.

Table 4: Test set classification results.

Sample	E (%)	REE (%)	DE (%)	LREE (%)
Subject 1	57.05	68.52	85.09	83.48
Subject 2	57.32	43.62	45.85	66.73
Subject 3	58.17	58.97	70.91	66.38
Subject 4	71.29	74.15	72.73	65.86
Subject 5	75.20	60.23	85.47	91.17
Subject 6	89.95	86.15	86.78	82.14
Subject 7	45.93	47.11	47.83	59.47
Subject 8	72.95	74.97	84.33	72.65
AVG	65.98	64.22	72.37	73.48
VAR	0.0193	0.0210	0.0284	0.0119

Table 5: Comparison of results of the two classification models.

Model	Average accuracy (%)	The variance of accuracy
ALL-LSTM	86.48	0.0039
WT-LSTM	73.48	0.0119

higher accuracy and higher stability when used for emotion classification.

These results indicate that when the LREE is selected as the feature for the WT-LSTM model, the classification effect is the best, and the average accuracy of the classification is 73.48%. The LREE is also the most suitable EEG feature of the four wavelet features in the emotion classification problem and LSTM network.

5. Discussion and Conclusion

In theory, the advantage of feature extraction for the WT-LSTM model is that it provides a straightforward solution for obtaining hidden information about the signal's frequency contents or brain area connectivity from the available channels, compared to the information gained by using the EEG signals as a time series. However, Table 5 shows that the best classification accuracy of the WT-LSTM model with features extracted via WT is still significantly lower than that of the ALL-LSTM model.

The best average accuracy of the WT-LSTM model is about 13% lower than the average accuracy of the ALL-LSTM model, while the variance also increases. The results showed that the ALL-LSTM model's stability was slightly better than that of the WT-LSTM model in the emotion classification of these eight subjects.

This could be attributed to LSTM displaying a strong ability to use context. Feature extraction based on WT is currently based on more complex and mature feature extraction methods used during EEG emotion recognition. However, feature extraction via WT may destroy the timing of the EEG signal itself. Timing information is vital for emotional classification, and the LSTM model can make full use of the timing information implicit in the EEG data. The significant improvement in the classification ability of the ALL-LSTM model also shows that the method is feasible for EEG-based emotion recognition.

Furthermore, by adding more layers and memory units, better EEG signal representation could be learned when the LSTM network's size was substantially increased, compensating for the more extensive input size by directly providing the EEG signals. However, the computational cost of training larger LSTM networks increases rapidly, requiring extended training time or GPU arrays. Regardless of the computational cost, this method could need even more EEG data to effectively train the millions of network parameters.

Therefore, the goal is to train the LSTM network by learning the suitable emotion features, which can be realized by essentially simulating a more profound and more complex LSTM model without increasing the time and data training limitations. Therefore, the emotion recognition system can run under more suitable conditions.

Data Availability

The EEG data used to support the findings of this study were supplied by National Nature Science Foundation of China

under license and so cannot be made freely available. Requests for access to these data should be made to Huiping Jiang, jianghp@muc.edu.cn.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Huiping Jiang was supported by the National Nature Science Foundation of China (No. 61503423). This work was supported in part by the Leading Talent Program of State Ethnic Affairs Commission and Double First-Class Special Funding of MUC. The authors thank all the participants in this research and the technical support from FISTAR Technology Inc.

References

[1] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55–64, 2003.

[2] Q. Zhou, "Multi-layer affective computing model based on emotional psychology," *Electronic Commerce Research*, vol. 18, no. 1, pp. 109–124, 2018.

[3] K. Bernhard, L. Suzana, K. Mathias et al., "Decision support with text-based emotion recognition: deep learning for affective computing," *Decision Support Systems*, vol. 115, pp. 24–35, 2018.

[4] F. Wang, S.-H. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for EEG-based emotion recognition with deep convolutional neural networks," in *MultiMedia Modeling*, pp. 82–93, Springer, Cham, Switzerland, 2018.

[5] T. Tanaka, T. Uehara, and Y. Tanaka, "Dimensionality reduction of sample covariance matrices by graph fourier transform for motor imagery brain-machine interface," in *Proceedings of the 2016 IEEE Statistical Signal Processing Workshop (SSP) IEEE*, Palma de Mallorca, Spain, June 2016.

[6] B. K. Kim, E. C. Lee, B. Suhng, D. Ryu, and W. Lee, "Feature extraction using FFT for banknotes recognition in a variety of lighting conditions," in *Proceedings of the 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013)*, pp. 698–700, Gwangju, South Korea, October 2013.

[7] D. Wang, D. Q. Miao, and R. Z. Wang, "A new method of EEG classification with feature extraction based on wavelet packet decomposition," *Acta Electronica Sinica*, vol. 41, no. 1, pp. 193–198, 2013.

[8] G. Akshansh, R. K. Agrawal, and B. Kaur, "Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods," *Soft Computing*, vol. 19, no. 10, pp. 2799–2812, 2015.

[9] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010.

[10] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141-142, 2012.

[11] M. Yanagimoto and C. Sugimoto, "Recognition of persisting emotional valence from EEG using convolutional neural networks," in *Proceedings of the 2016 IEEE 9th International Workshop on Computational Intelligence and Applications (IWCIA)*, Hiroshima, Japan, November 2016.

[12] T. Muhammed, U. B. Baloglu, Ö. Yildirim, and U. Rajendra Achary, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cognitive Systems Research*, vol. 54, pp. 176–188, 2019.

[13] N. Seijdel, K. Ramakrishnan, M. Losch, and S. Scholte, "Overlap in performance of CNN's, human behavior and EEG classification," *Journal of Vision*, vol. 16, no. 12, p. 501, 2016.

[14] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10–18, 2018.

[15] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 103–113, 2018.

[16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[17] P. S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*, San Francisco, CA, USA, October 2013.

[18] Z. Ni, A. C. Yuksel, X. Ni, M. I. Mandel, and L. Xie, "Confused or not confused? disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks," in *Proceedings of the 8th ACM International Conference ACM*, Boston, MA, USA, August 2017.

[19] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17–28, 2016.

[20] B. Y. Zhang, H. P. Jiang, and L. Dong, "Classification of EEG signal by WT-CNN model in emotion recognition system," in *Proceedings of the 2017 IEEE 16TH International Conference on Cognitive Informatics & Cognitive Computing (ICCI CC)*, pp. 109–114, Oxford, UK, July 2015.

[21] G. Ruffini, D. Ibañez, M. Castellano, S. Dunne, and A. Soria-Frisch, "EEG-driven RNN classification for prognosis of neurodegeneration in at-risk patients," in *25th International Conference on Artificial Neural Networks (ICANN 2016)*, pp. 306–313, Barcelona, Spain, October 2016.

[22] U. Elif Derya, "Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks," *Digital Signal Processing*, vol. 19, no. 1, pp. 134–143, 2009.

[23] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, 2017.

[24] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[25] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 851–869, 2017.

[26] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proceedings of the Twenty-Ninth AAAI Conference on*

Innovative Applications (IAAI-2017), San Francisco, CA, USA, February 2017.

[27] E. Shawky, R. El-Khoribi, M. Shoman, and M. Wahby Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, pp. 329–337, 2018.

[28] S.-E. Moon, S. Jang, and J. Lee, "Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2556–2560, Calgary, Canada, April 2018.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing*, vol. 30, no. 1–4, pp. 201–218, 2000.

[31] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, Berlin Germany, 2012.

[32] F. A. Gers, I. Galleria, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *Journal of Machine Learning Research*, vol. 3, no. 1, pp. 115–143, 2003.

[33] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.

[34] R. DiPietro and G. D. Hager, "Deep learning: RNNs and LSTM," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, pp. 503–519, Academic Press, Cambridge, MA, USA, 2020.

[35] M. M. Hasib, T. Nayak, and Y. Huang, "A hierarchical LSTM model with attention for modeling EEG non-stationarity for human decision prediction," in *Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, USA, March 2018.

[36] S. Alex, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, 2020.

[37] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[38] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers," *Cognition & Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.

[39] R. J. Barry and E. N. Sokolov, "Habituation of phasic and tonic components of the orienting reflex," *International Journal of Psychophysiology*, vol. 15, no. 1, pp. 39–42, 1993.

[40] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.

Hindawi
Complexity
Volume 2021, Article ID 6691117, 10 pages
https://doi.org/10.1155/2021/6691117

Research Article

Image-Based Iron Slag Segmentation via Graph Convolutional Networks

Wang Long,[1] **Zheng Junfeng**,[2] **Yu Hong**,[2] **Ding Meng**,[3] **and Li Jiangyun**[2,4]

[1]*State Key Laboratory of Advanced Special Steel & Shanghai Key Laboratory of Advanced Ferrometallurgy & School of Materials Science and Engineering, Shanghai University, Shanghai, China*
[2]*School of Automation & Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[3]*Scoop Medical, Inc., Houston 77007, TX, USA*
[4]*Shunde Graduate School of University of Science and Technology Beijing, Foshan 528000, China*

Correspondence should be addressed to Li Jiangyun; leejy@ustb.edu.cn

Received 10 October 2020; Revised 6 December 2020; Accepted 6 January 2021; Published 2 February 2021

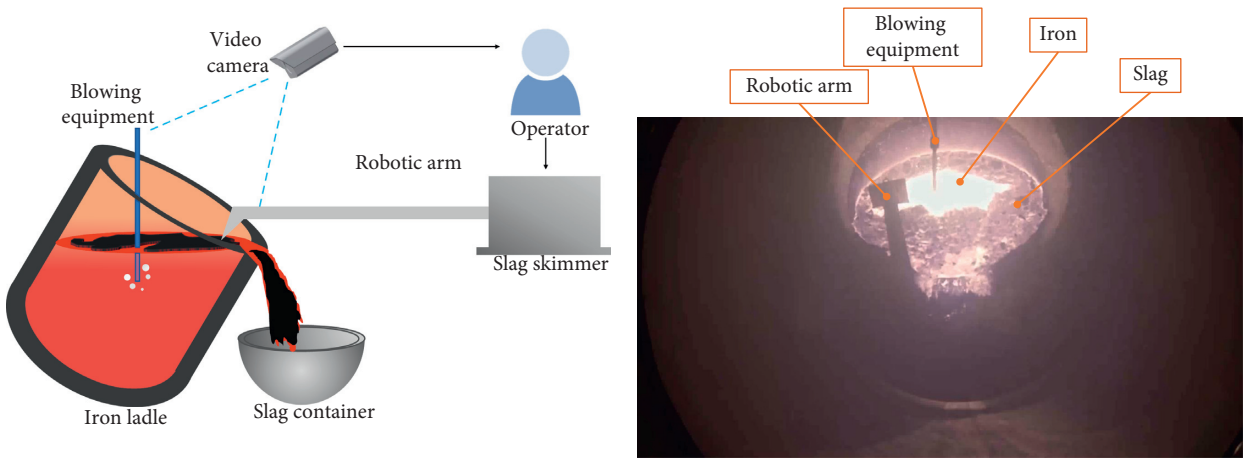
Academic Editor: Ning Cai

Slagging-off (i.e., slag removal) is an important preprocessing operation of steel-making to improve the purity of iron. Current manual-operated slag removal schemes are inefficient and labor-intensive. Automatic slagging-off is desirable but challenging as the reliable recognition of iron and slag is difficult. This work focuses on realizing an efficient and accurate recognition algorithm of iron and slag, which is conducive to realize automatic slagging-off operation. Motivated by the recent success of deep learning techniques in smart manufacturing, we introduce deep learning methods to this field for the first time. The monotonous gray value of industry images, poor image quality, and nonrigid feature of iron and slag challenge the existing fully convolutional networks (FCNs). To this end, we propose a novel spatial and feature graph convolutional network (SFGCN) module. SFGCN module can be easily inserted in FCNs to improve the reasoning ability of global contextual information, which is helpful to enhance the segmentation accuracy of small objects and isolated areas. To verify the validity of the SFGCN module, we create an industrial dataset and conduct extensive experiments. Finally, the results show that our SFGCN module brings a consistent performance boost for a wide range of FCNs. Moreover, by adopting a lightweight network as backbone, our method achieves real-time iron and slag segmentation. In the future work, we will dedicate our efforts to the weakly supervised learning for quick annotation of big data stream to improve the generalization ability of current models.

1. Introduction

Slagging-off is an essential operation in steel-making. It is used to remove high sulfur slag from molten iron to improve the purity of iron. The process of slagging-off is shown in Figure 1(a) and the actual image obtained by video capture is shown in Figure 1(b). In this process, molten iron is inevitably brought out, and the loss of molten iron is directly proportional to the clean rate of slagging-off. Meanwhile, slagging-off operation will be accompanied by the decrease of molten iron temperature. Therefore, accuracy and efficiency are two key factors of slagging-off operation, which are directly related to production energy consumption. At present, manual operation of machinery for slag removal is a commonly employed scheme in industrial applications. However, affected by the long-term strong light and dense smoke condition, it can easily lead to misidentification and misoperation. Besides, manual operation is inefficient. With the introduction of Industry 4.0 paradigm, the trend is moving towards to intelligent production line, where automatic slagging-off will benefit the modern smart manufacturing greatly.

Recognition of iron and slag is the premise of automatic slagging-off operation. We formulate this problem as a semantic segmentation task, which is a fundamental problem of computer vision and aims to assign categories for



(a) (b)

Figure 1: (a) The process of slagging-off. The operator obtains real-time images of the iron ladle through the video camera and operates the slag skimmer to move the slag to the slag container. When the slag is less or dispersed, the blowing equipment will blow nitrogen to polymerize the slag. During this process, it is necessary to avoid the collision between the blowing equipment, the inner wall of the iron ladle, and the robotic arm. (b) The image of actual slagging-off operation working condition from video capture.

each pixel in an image. Many classical machine vision methods have been proposed for image segmentation. However, the monotonous gray value of industry images and poor image quality caused by strong light and dense smoke condition challenge the performance of traditional computer vision algorithms. As far as the task of iron and slag segmentation is concerned, results of some traditional algorithms including K-means, Markov random field, and mean shift are shown in Figure 2, which are obviously cannot meet the requirement of industrial application. Currently, the state-of-the-art methods for segmentation mainly based on fully convolutional networks (FCNs) are used [1]. However, only modeling local correlation with convolutional operations, FCNs are not effective to reason relation between distant regions with arbitrary shape without stacking multiple convolution layers. To tackle this problem, many algorithms have been proposed to expand the receptive field of FCNs to capture long-range contextual information in the scene. Dilated convolution has been implemented to capture large objects, thus introducing another problem that small objects may be ignored. Another research direction is fusing multiscale features [2], which is inefficient. Recently, self-attention mechanism-based methods [3] make use of affinity matrix to model the relation between each spatial position and its neighborhoods. However, the memory and computational requirements of large affinity matrix prevent the application of these methods for high-resolution image segmentation application, such as the iron and slag segmentation with a resolution of 1920×1080.

The monotonous gray value of industry images, poor image quality, and irregular and scattered shape of slag also challenge the existing FCNs. Graph convolution is an efficient and effective operation to model global contextual information over regions in a single layer, which has been widely employed in recent scene understanding works [4, 5]. Motivated by these works, we propose an effective and efficient spatial and feature graph convolutional network (SFGCN) module based on graph convolution. Different from previous works, our SFGCN module makes use of latent interaction space to efficiently perform global reasoning function. Our SFGCN module consists of two parallel branches to project feature maps to latent spatial space and feature space, respectively. Then, graph convolutions are employed to perform relation reasoning. After graph reasoning, the updated information is reprojected back into the original coordinate space for further information extraction. Extensive experiments prove that our SFGCN module can consistently improve the performance of current mainstream convolutional neural network backbones for iron and slag segmentation.

Our contributions can be summarized as follows:

(1) We formulate the slagging-off problem as an image-based semantic segmentation task and explore deep learning methods to tackle the automatic iron and slag recognition task for the first time.

(2) Considering the limitation of convolution operations for modeling local correlation, we propose a SFGCN module to effectively reason global information interaction via weighted spatial graph convolution and feature graph convolution branches. The proposed network is termed as SFGCNet.

(3) We establish an industrial slagging-off dataset and conduct extensive experiments, and the results show that our SFGCN module brings consistent performance improvement for a wide range of network backbones for iron and slag segmentation. Moreover, taking a lightweight network as backbone, our method is able to achieve real-time segmentation of iron and slag.

Figure 2: The segmentation results of traditional methods including K-means, Markov random field, and mean shift. The last column is the segmentation we expect. In the expected result, white, green, blue, and pink represent background, robotic arm, iron, and slag, respectively.

2. Related Work

Fully convolutional networks (FCNs) have made great progress in semantic segmentation [1,6]. There are many variants to improve the performance of segmentation; we briefly review several main research directions in scene understanding domain, including network architecture implementation, global context reasoning, and graph-based reasoning.

2.1. Network Architecture Implementation. Atrous Spatial Pyramid Pooling (ASPP) has been proposed and employed in Deeplabv2, v3 [7] to integrate multiscale contextual information, which contains multiple parallel dilated convolutions with different dilated rates. A variety of encoder-decoder structures have been implemented to obtain effective usage of midlevel and high-level extracted features [8, 9]. PSPNet [2] builds a novel pyramid pooling module to get multiscale contextual prior knowledge. DenseASPP [10] embeds multiscale features to expand the receptive field of convolution layers for segmentation task. All these methods effectively stack multiple convolution layers to collect multiscale information.

2.2. Global Context Reasoning. Many methods have been proposed to overcome the limitation that convolution layers are difficult to capture global context, such as self-attention mechanism and nonlocal networks. Self-attention mechanism is firstly proposed in [11] to model long-range dependencies for machine translation task and has been widely applied in many tasks in recent years [12]. PSANet [13] captures pixel-wise relation by applying attention module in spatial dimension. EncNet [14] and DFN [15] apply attention module along the channel dimension of the feature map to account for global context. DANet [16] uses attention module in both spatial and channel dimensions. Nonlocal networks [3, 17] aim to deliver long-range information from one position to another.

2.3. Graph-Based Reasoning. Graph-based reasoning provides an efficient idea of global context reasoning. Random walk and conditional random field (CRF) networks have been proposed based on graph for efficient image segmentation and classification. Recently, graph convolutional networks (GCNs) have been proposed for semisupervised image classification. Wang et al. [18] apply GCN to capture global contextual relation in video recognition task. Chen et al. [4] explore GCN to reason global relation in semantic segmentation task. Yan et al. introduce GCN to describe skeleton connections for action recognition [19, 20]. Following these methods, we propose a novel dual GCN module consists of spatial graph convolution and feature graph convolution to model global contextual information for iron slag segmentation. Our SFGCN module makes use of latent spatial and feature spaces to efficiently realize global relation reasoning, which alleviates the memory and computation burden of global context reasoning while improving the performance of segmentation.

3. Methods

In this section, we first review the graph convolution and then introduce the implementation of our SFGCN module. Finally, we detail the network architecture for slag segmentation.

3.1. Graph Convolution. Graph convolution is an efficient operation to reason global context information, which overcomes the limitation that convolution operation can only model local context information. Graph convolution defined in graph G with nodes N and edges E can effectively achieve global information interaction in a single operation. The specific operation can be defined as follows:

$$O = \sigma(AXW). \tag{1}$$

The specific implementation steps are shown as the following pseudocode, including (1) project the feature map from coordinate space to graph space, we employ the conventional convolution operations to project the feature map to graph space after the feature extraction operation, and the process is shown in Figure 3; (2) build adjacency matrix to describe intrabody connections of nodes within the graph; (3) update the weight matrix; and (4) reproject the graph to coordinate space. The feature map extracted by backbone networks contains spatial and channel dimensions. Assuming that spatial dimension abstracts the objects in the scene and channel dimension encapsulates the detailed object features, that means the graph established in spatial space is able to describe the relevance between objects

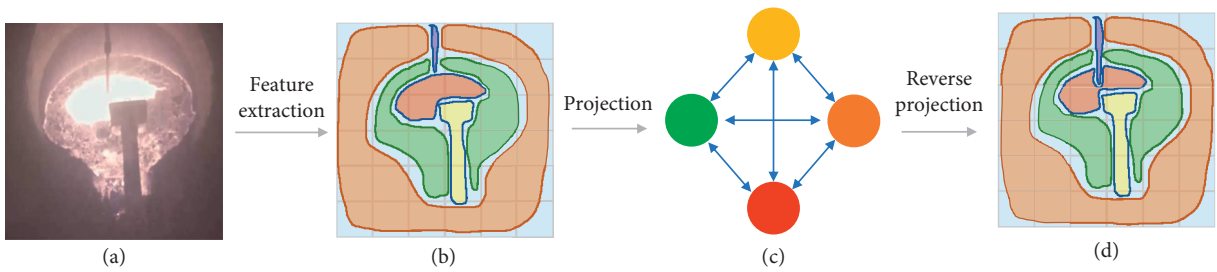


Figure 3: The illustration of graph convolution operation. (a) The input image. (b) The feature map obtained by backbones. (c) Projecting the feature map to graph space and reasoning the relation between all nodes. (d) Reprojecting the learned interaction to feature map for performance improvement. Here, we use nodes with different colors to represent different object regions.

in the scene and the graph established in feature space is able to express the relevance between object parts. Therefore, we conducted graph convolution on spatial graph and feature graph, respectively. The spatial branch is used to grasp thecinternal integrity of objects and the relationship between objects. The feature branch is used to characterize the details of objects and the relationship between features (Algorithm 1) [21].

3.2. Graph Convolution in Spatial Space

3.2.1. Spatial Space Projection. As shown in Figure 4, before conducting graph convolution operation, we first project the input feature map to latent spatial space to get the graph. In practice, spatial downsampling operation T_s is employed to transform the input feature $X \in R^{aH\times W\times C}$ to graph $G_s \in R^{(H\times W/d^2)\times C}$ in the latent spatial space S_s, where d represents the downsampling rate. We achieve T_s based on stacked depth-wise convolution operations in each layer with a stride of 2 and kernel size of 3×3. Then, G_s is obtained via

$$G_s = T_s(X). \tag{2}$$

3.2.2. Spatial Graph Convolution. After projecting the input feature X to graph G_s, the graph consists of $(H/d)\times(W/d)$ nodes. Each node of the graph integrates the information of a cluster of pixels in the feature map. To measure the correlation between nodes, we form an adjacency matrix $A_s \in R^{(HW/d^2)\times(HW/d^2)}$. The spatial graph convolution is implemented according to the following formulation to achieve global relation reasoning:

$$O_s = f\left(\delta_s(G_s)\cdot\psi_s(G_s)^T\right)\cdot G_s W_s, \tag{3}$$

where $f(\delta_s(G_s)\cdot\psi_s(G_s)^T)$ gives the adjacency matrix A_s and $f(\cdot)$ represents the softmax activation function, in which $\cdot$ is the dot-production operation. W_s is the weight matrix for updating information.

3.2.3. Reprojection. After relational reasoning, we reproject O_s back to the original coordinate space $(R^{H\times W\times C})$ for compatibility with later operations. Different from the downsampling operation T_s in graph projection, we directly employ nearest neighbour interpolation to upsample O_s to the original input size. Finally, the output feature map of spatial graph convolution branch is obtained according to $\widetilde{O}_s = \xi_s(\text{interp}\ (O_s))$.

3.3. Graph Convolution in Feature Space. Spatial graph convolution models the spatial correlation of pixel clusters in a scene, which enables the network to make correlation prediction based on all objects in the whole scene. Next, we consider projecting input feature map to feature space and reasoning correlation along the channel dimension. Assuming that the latter layers of the FCN are responsive to the object parts and high-level semantic features, conducting GCN in feature space can model the correlation of abstract features such as object parts. We first adopt a channel downsampling operation $\theta(\cdot)$ to reduce the channels of input feature from $X \in R^{H\times W\times C}$ to $H_f \in R^{H\times W\times C_1}$ and employ a linear combination function $\varphi(\cdot)$ to aggregate information along the channel dimension. Finally, we obtain the formulation of input feature $X \in R^{H\times W\times C}$ to feature space graph $G_f \in R^{C_1\times C_2}$:

$$G_f = \theta(X)^T\cdot\varphi(X) = H_f^T\cdot\varphi(X), \tag{4}$$

where C_1 represents nodes and C_2 denotes the states of each node. After feature space projection, the feature graph convolution and reprojection are conducted according to the following equations:

$$\begin{aligned} O_f &= \left(I + A_f\right)G_f W_f, \\ \widetilde{O}_f &= \xi_f\left(H_f\cdot O_f\right). \end{aligned} \tag{5}$$

Considering the low dimension of feature graph, we employ two 1D convolution layers as adjacent matrix A_f and trainable edge weights W_f. To alleviate the optimization difficulty, the adjacent matrix A_f is updated with a residual structure and reconstructed as $(I + A_f)$. Both A_f and W_f are randomly initialized and optimized with gradient descent during the training process.

3.4. SFGCNet. Finally, the output of SFGCN is computed as $\widetilde{X} = \omega X + \omega_s\widetilde{O}_s + \omega_f\widetilde{O}_f$, where "+" denotes point-wise summation and ω is the learnable weight coefficient. Now we can easily embed our SFGCN module into the existing network backbone (e.g., FCN and ResNet).

Input: Tensor extracted by convolutional network
Output: Tensor after graph convolution operation

```
function SFGCN(Tensor)
    Project coordinate input to graph space X ← Projecting(Tensor)
    δ ← Conv(X)
    ϕ ← Conv(X)
    Build adjacency matrix A ← Soft max(δ^T * ϕ)
    Update weight matrix AXW ← Conv(A, X)
    O ← Activation(AXW)
    Reproject graph O to coordinate space result ← Reprojecting(O)
    return result
end function
```

Algorithm 1: GCN 1: realization of graph convolution.

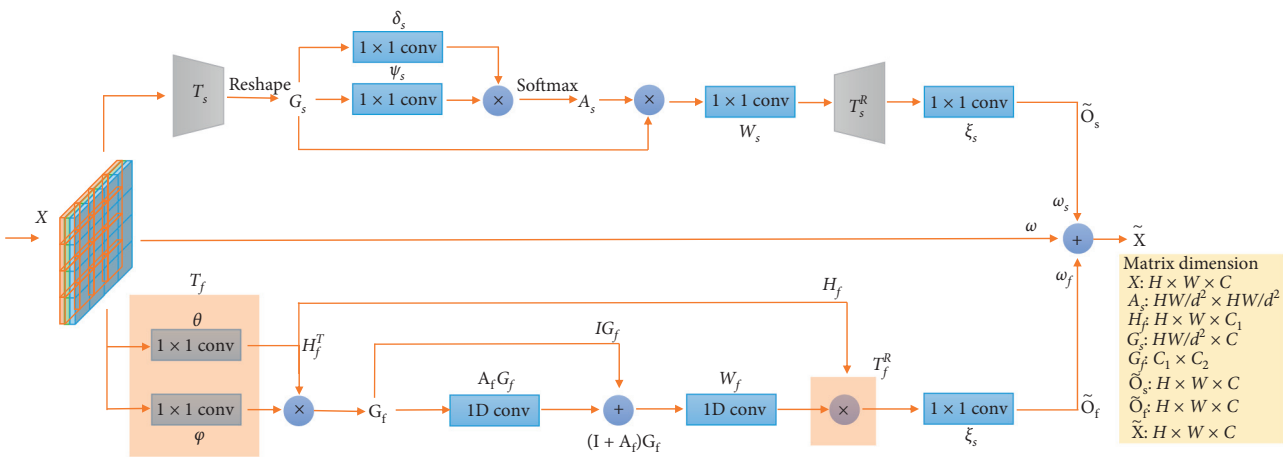


Figure 4: The design details of the SFGCN module. Our method contains two branches of graph convolution operation to model global contextual information along spatial and channel dimensions of feature map X.

3.4.1. Implementation of SFGCNet. As shown in Figure 5, we embed SFGCN module in the last stage of fully convolutional networks (FCNs) to achieve the segmentation of iron and slag. In order to verify the effectiveness of SFGCN module, we construct SFGCNet by adopting FCN [1], BiSeNet [22], ICNet [23], and ResNet-50 [24] as the network backbones, respectively. BiSeNet and ICNet are two light-weight networks to achieve real-time semantic segmentation.

4. Experiments

4.1. Dataset and Evaluation Metrics. As there has no public slagging-off dataset, we collect 7 videos from different industrial cameras. Due to the time-consuming and laborious segmentation labeling, we only select 24 clips from all 7 videos randomly. Each of the clips contains 64 frames. All of these clips are segmented with Photoshop software manually, by three raters, following the same annotation protocol, and their annotations are approved by experienced workers, and then, we split these images into training set and test set with a ratio of 3: 1. The annotation sample is presented in Figure 6. The training set is used to train models, and the test set is used to validate the performance of trained models.

The efficiency and accuracy of the model are mainly considered in industrial applications. The efficiency of the model can be evaluated by inference time, model parameters amount, and the total number of floating-point operations per second (FLOPs). To evaluate the accuracy of the model, we adopt the commonly used metrics in the segmentation task, including Mean Intersection over Union (MIoU) and pixel accuracy (PA). The two metrics are defined as follows:

$$\mathrm{MIoU} = \frac{1}{K+1}\sum_{i=0}^{K}\frac{P_{ii}}{\sum_{j=0}^{K}P_{ij} + \sum_{j=0}^{K}P_{ji} - P_{ii}},$$
$$\mathrm{PA} = \sum_{i=0}^{K}\sum_{j=0}^{K}\frac{P_{ii}}{P_{ij}}, \tag{6}$$

where P_{ii} represents the pixel predicted correctly (i.e., the true category of the pixel is class i, and the prediction is class i too). P_{ij}, P_{ji} mean the pixel prediction is wrong (i.e., the true category of the pixel is class (i/j), and the prediction is class (j/i).

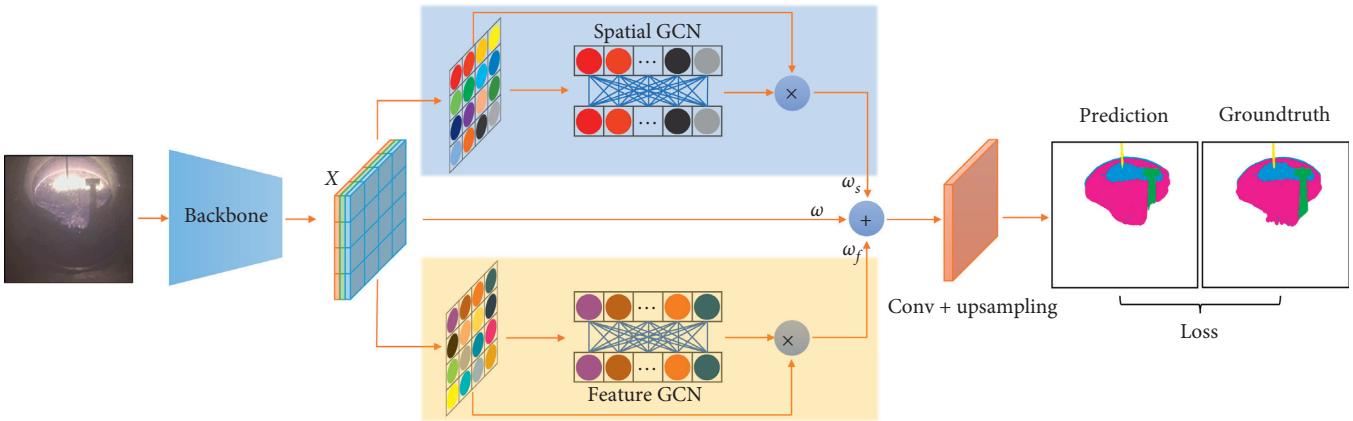


Figure 5: The architecture of the proposed network, i.e., SFGCNet. SFGCN module is inserted in the last stage of fully convolutional networks. The weights of SFGCNet are optimized by gradient descent algorithm and the cross entropy loss between prediction and groundtruth.

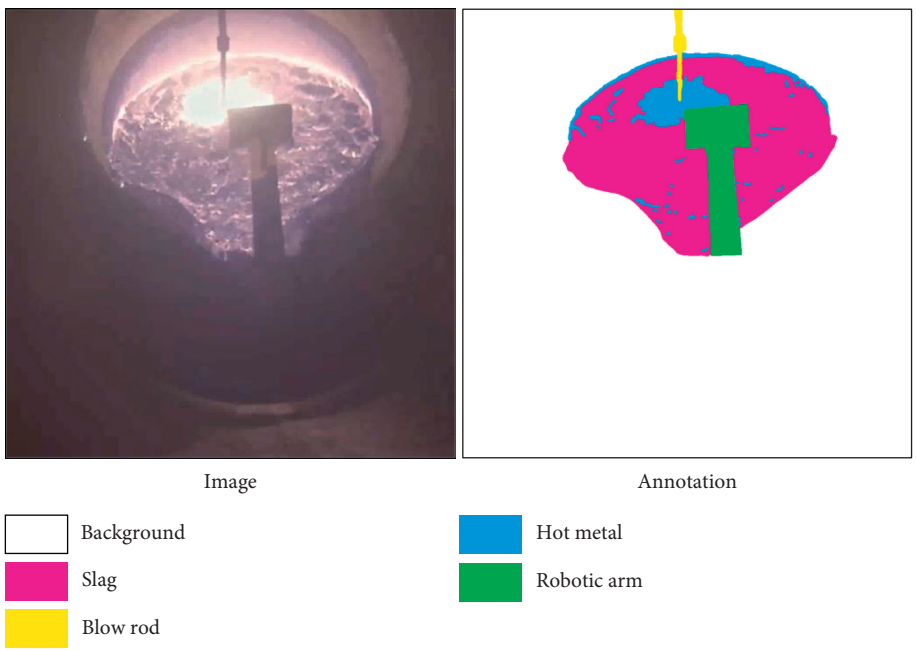


Figure 6: Image annotation for semantic segmentation. There are 5 categories in annotations, including white for background, pink for slag, yellow for blowing equipment, blue for iron, and green for robotic arm.

4.2. Preprocessing. The annotation of semantic segmentation is time-consuming and labor-intensive. Also, it is difficult to obtain a large number of labeled data in industrial applications. Thus, data augmentation is an effective method to expand the dataset, which is helpful for alleviating the overfitting problem and enhancing the robustness of the network. Considering that images acquired by the video camera contain a large number of background areas, which cannot benefit the accuracy, we firstly crop the raw image from 1920×1080 to 1024×1024 to reduce the proportion of background area. After that, we randomly apply the data augmentation methods with 50% probability, including the following:

(1) Random horizontal and vertical flips

(2) Random scaling between $[0.5, 2]$

(3) Random intensity shift between $[-0.1, 0.1]$

Table 1: The results of deep learning-based methods on the test set. The size of the input image is 1024×1024.

Models	Iron	Slag	Robotic arm	Blow pole	MIoU (%)	PA (%)	Inference time (ms)	Parameters (M)	FLOPs (G)
BiSeNet [22]	67.49	82.91	82.25	55.77	72.11	97.04	15.47	12.42	48.77
BiSeNet + SFGCN	64.55	82.98	80.66	71.16	**77.74**	97.26	18.28	13.4	60.35
ICNet [23]	60.74	69.54	67.17	72.92	67.59	94.61	44.62	28.29	147.68
ICNet + SFGCN	61.47	73.54	68.63	70.05	**68.42**	95.45	45.79	28.79	153.0
FCN [1]	68.22	83.08	83.62	64.63	74.89	97.15	66.67	18.64	321.78
FCN + SFGCN	68.24	83.76	84.92	71.23	**78.38**	97.31	67.46	21.86	324.52
ResNet-50 [24]	55.06	78.97	79.15	71.32	71.13	96.38	30.18	28.51	98.18
ResNet-50 + SFGCN	66.29	73.04	76.92	72.49	**75.00**	96.62	30.73	28.75	98.57

4.3. Experiments and Results

4.3.1. Experiment Setup. We implement our method with PyTorch. Cosine annealing learning rate policy is used with 30 warming-up epochs. The initial learning rate lr_0 is set to 0.001 and adjusted based on the following formulation:

$$\mathrm{lr} = \begin{cases} \mathrm{lr}_0 * \left(1 - \cos\left(\frac{\pi}{2} * \frac{e}{w}\right)\right), & e \le w, \\ \mathrm{lr}_0 * \frac{\cos\left(\pi^* (e - w)/t\right) + 1)}{2} & w < e < t. \end{cases} \tag{7}$$

Specifically, e represents the current training epoch, t is the total number of training epochs, and w denotes the warming-up epochs. We train our model with Adam optimizer and synchronized BN on four parallel Nvidia 2080Ti GPUs for 300 epochs. The batch size is set to 8 to guarantee the performance of batch normalization.

4.3.2. Experiment Results. We apply our SFGCN module to the last stage of typical backbones such as FCN, ResNet-50, BiSeNet, and ICNet to reason long-distance dependencies. Considering the distribution difference between industrial dataset and natural scene dataset, we train all the above backbones from scratch. As shown in Table 1, our SFGCN module widely improves the performance of different backbones. In terms of MIoU, SFGCN module brings 5.63%, 0.83%, 3.49%, and 3.87% improvements on BiSeNet, ICNet, FCN, and ResNet-50, respectively. Benefited from the global reasoning function of graph convolution, SFGCN module makes the isolated slag and molten iron region more easily to be identified. As shown in Figure 7, while the dispersed areas of slag and iron are easy to be segmented incorrectly, SFGCN alleviates the influence of neighbor regions on the classification of these regions. On the other hand, the introduction of SFGCN module only results in 2.81 ms, 1.17 ms, 0.79 ms, and 0.55 ms more inference time for BiSeNet, ICNet, FCN, and ResNet-50, respectively, as well as slight parameters and FLOPs increase, which demonstrates that our SFGCN module is efficient. Especially, taking lightweight BiSeNet as the backbone, our SFGCNet achieves real-time segmentation of iron and slag.

4.3.3. Ablation Studies and Discussion. Embedded location of SFGCN module: our SFGCN module can be flexibly embedded in any stage of the network backbone, and it is worth exploring where the embedding can achieve better results. Moreover, the embedding location will affect the accuracy and efficiency of the network at the same time. The feature map of shallow layers has high resolution, which directly increases the parameters and FLOPs of the SFGCN module. From the perspective of feature extraction, shallow layers cannot capture abundant semantic information due to the lacking of enough receptive fields, which will also limit the performance of SFGCN module. Experiments show that the SFGCN module achieves higher efficiency when it is embedded in the last stage of various backbones.

The effectiveness of each branch: to verify the effectiveness of SGCN branch and FGCN branch, we conduct experiments on BiSeNet and FCN with different settings in Table 2.

As shown in Table 2, both SGCN and FGCN boost the performance of BiSeNet and FCN. The introduction of SGCN and FGCN, respectively, yields 3.82% and 4.46% improvement in MIoU for baseline of BiSeNet. Meanwhile, SGCN and FGCN outperform the FCN baseline by 2.15% and 2.81%. After integrating SGCN and FGCN branches, our method achieves 5.63% and 3.49% performance boost for BiSeNet and FCN. Results show that SFGCN module brings benefits for the segmentation of iron and slag.

The effects of SGCN and FGCN branches are visualized in Figure 8. As shown in the third column, SGCN aggregates information of pixel cluster and delivers messages between nodes, thus guaranteeing the integrity of objects. However, spatial branch loses details of each node while aggregating node information. The FGCN branch focuses more on reasoning the details of objects to make up for the deficiency of the SGCN branch which focuses more on connection between objects. The refinement of segmentation is significantly improved as shown in the fourth column.

We compute the coefficients of SGCN and FGCN branches to objectively evaluate the contribution of these two branches. The shortcut connection weight ω is set to 1. ω_s and ω_f are initialized as 1 and learnable. The final coefficients of each branch of the SFGCN module in different backbones are shown in Table 3, and the results show that SGCN and FGCN branches do provide extra information for the segmentation. Moreover, the coefficients vary for different network backbones. Therefore, the learnable coefficients provide the flexibility of adjusting the contribution of each branch based on the information learned by the base network.

Effect of projection: as described in Section 3, we aggregate information along spatial and channel dimensions to

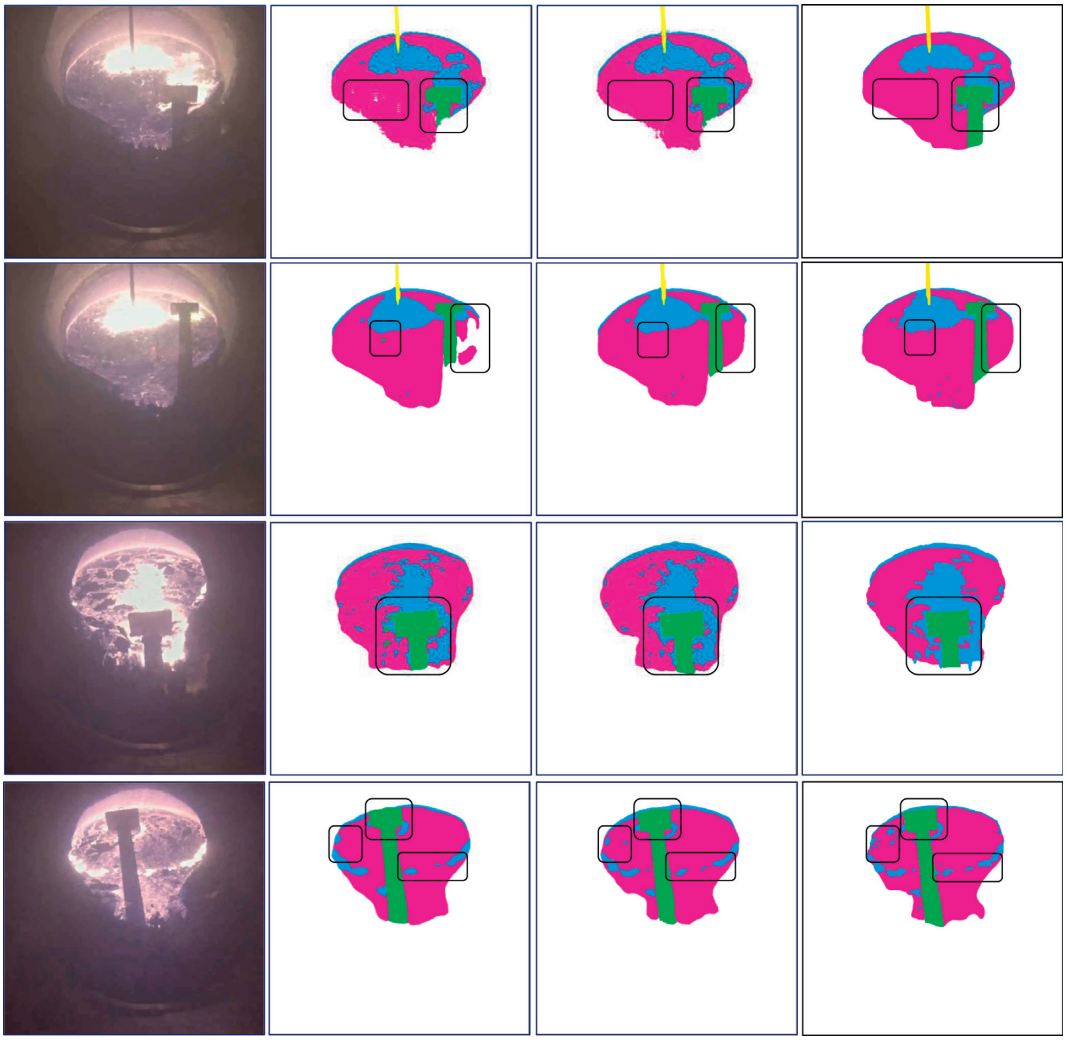


Figure 7: The visualization results of SFGCNet with different backbones. The results of FCN, BiSeNet, ICNet, and ResNet-50 are shown from the first row to the last row. More visual results are available at https://github.com/ustbzjf1/SFGCNet-for-hot-metal-slag-segmentation.

Table 2: The ablation study based on the network backbone of BiSeNet and FCN on the test set.

Backbone	SGCN	FGCN	MIoU
BiSeNet			72.11
BiSeNet	√		75.93
BiSeNet		√	76.57
BiSeNet	√	√	77.74
FCN			74.89
FCN	√		77.04
FCN		√	77.70
FCN	√	√	78.38

SGCN and FGCN represent spatial GCN and feature GCN, respectively.

project the input feature map to the graph space. The downsampling ratio of SGCN branch directly determines the degree of spatial information aggregation. Large ratio loses details while small ratio retains useless information. The number of nodes in the FGCN branch also affects the relation reasoning between the features of objects. Appropriate number of nodes is important for recovering the details of each object. After conducting extensive experiments on our dataset, we observe that SFGCN module brings more performance improvement when the size of G_s

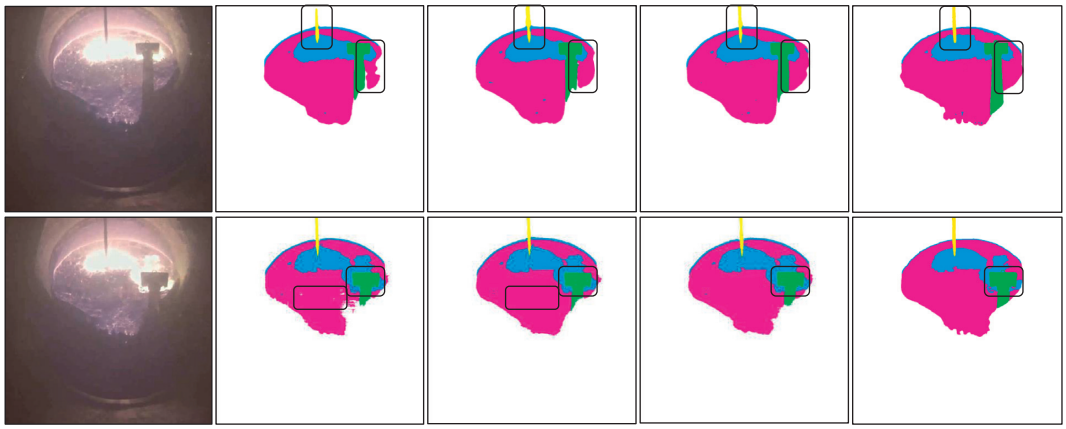


Figure 8: The visualization results of SGCN and FGCN on the test set.

Table 3: The adaptive coefficients of SGCN and FGCN branches.

SFGCNets	ω_s	ω_f	MIoU
BiSeNet + SFGCN	0.452	0.286	77.74
ICNet + SFGCN	0.143	0.712	68.42
FCN + SFGCN	0.320	0.438	78.38
ResNet-50 + SFGCN	0.897	0.526	75.00

is (1/64) of the input image size and the number of nodes for G_f is 32. We speculate that 64× downsampling to aggregate information is more suitable for the scale of objects and 32 nodes can better express the details of objects in the slagging-off scene.

5. Conclusion

In this work, we explore deep learning methods for iron and slag recognition. We formulate this problem as a semantic segmentation task and propose a SFGCN module to reason global contextual information according to the characteristic of the slagging-off task. Extensive experiments have verified that our method not only triumphs over traditional segmentation methods but also widely improves the performance of current mainstream deep learning models in the slagging-off task. Taking lightweight network as backbone, our SFGCNet can realize real-time and accurate recognition of iron and slag, which provides a significant reference for downstream automatic slagging-off operation.

Although our algorithm has achieved satisfactory results in view of accuracy and efficiency, we need to expand the dataset to improve the performance of the model in more scenarios. It is difficult to label industrial big data manually, in the future work, we will dedicate our efforts to the weakly supervised learning for quick annotation of big data stream to improve the generalization ability of current models.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by State Key Laboratory of Advanced Special Steel, Shanghai Key Laboratory of Advanced Ferrometallurgy and the Science and Technology Commission of Shanghai Municipality (No. 19DZ2270200), the Fundamental Research Funds for the China Central Universities of USTB (FR-DF-19-002), and Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB (BK20BE014).

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Seattle, WA, USA, June 2015.

[2] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890, Honolulu, HI, USA, July 2017.

[3] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, Salt Lake, UT, USA, October 2018.

[4] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 433–442, Long Beach, CL, USA, September 2019.

[5] X. Liang, Z. Hu, H. Zhang, L. Lin, and E. P. Xing, "Symbolic graph reasoning meets convolutions," *Advances in Neural Information Processing Systems*, vol. 17, pp. 1853–1863, 2018.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[7] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 834–848, 2017.

[8] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, July 2017.

[9] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5693–5703, Long Beach, CL, USA, July 2019.

[10] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Denseaspp for semantic segmentation in street scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3684–3692, Salt Lake, UT, USA, October 2018.

[11] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 5998–6008, Barcelona, Spain, May 2017.

[12] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "Disan: directional self-attention network for rnn/cnn-free language understanding," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, February 2018.

[13] H. Zhao, Y. Zhang, S. Liu et al., "Psanet: point-wise spatial attention network for scene parsing," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 267–283, Venice, Italy, October 2018.

[14] H. Zhang, K. Dana, J. Shi et al., "Context encoding for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 32, pp. 7151–7160, 2018.

[15] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1857–1866, Salt Lake, UT, USA, June 2018.

[16] H. Nam, J. W. Ha, and J. Kim, "Dual attention networks for multimodal reasoning and matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 299–307, Honolulu, HI, USA, July 2017.

[17] Y. Yuan and J. O. Wang, "Object Context Network for Scene Parsing," 2018, https://arxiv.org/abs/1809.00916.

[18] X. Wang and A. Gupta, "Videos as space-time region graphs," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 399–417, Glasgow, UK, August 2018.

[19] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proceedings of the Thirty-second AAAI conference on artificial intelligence*, New Orleans, LI, USA, February 2018.

[20] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12026–12035, Long Beach, CL, USA, September 2019.

[21] S. Zhang, H. Tong, J. Xu et al., "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019.

[22] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: bilateral segmentation network for real-time semantic segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 325–341, Munich, Germany, September 2018.

[23] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 405–420, Munich, Germany, September 2018.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

Hindawi
Complexity
Volume 2021, Article ID 6682871, 8 pages
https://doi.org/10.1155/2021/6682871

Research Article

End-to-End Speech Synthesis for Tibetan Multidialect

Xiaona Xu, Li Yang, Yue Zhao, and Hui Wang

School of Information Engineering, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Li Yang; 654577893@qq.com and Yue Zhao; zhaoyueso@muc.edu.cn

Received 30 October 2020; Revised 27 December 2020; Accepted 12 January 2021; Published 27 January 2021

Academic Editor: Ning Cai

The research on Tibetan speech synthesis technology has been mainly focusing on single dialect, and thus there is a lack of research on Tibetan multidialect speech synthesis technology. This paper presents an end-to-end Tibetan multidialect speech synthesis model to realize a speech synthesis system which can be used to synthesize different Tibetan dialects. Firstly, Wylie transliteration scheme is used to convert the Tibetan text into the corresponding Latin letters, which effectively reduces the size of training corpus and the workload of front-end text processing. Secondly, a shared feature prediction network with a cyclic sequence-to-sequence structure is built, which maps the Latin transliteration vector of Tibetan character to Mel spectrograms and learns the relevant features of multidialect speech data. Thirdly, two dialect-specific WaveNet vocoders are combined with the feature prediction network, which synthesizes the Mel spectrum of Lhasa-Ü-Tsang and Amdo pastoral dialect into time-domain waveform, respectively. The model avoids using a large number of Tibetan dialect expertise for processing some time-consuming tasks, such as phonetic analysis and phonological annotation. Additionally, it can directly synthesize Lhasa-Ü-Tsang and Amdo pastoral speech on the existing text annotation. The experimental results show that the synthesized speech of Lhasa-Ü-Tsang and Amdo pastoral dialect based on our proposed method has better clarity and naturalness than the Tibetan monolingual model.

1. Introduction

Speech synthesis, also known as text-to-speech (TTS) technology, mainly solves the problem of converting text information into audible sound information. Up to now, speech synthesis technology has become one of the most commonly used methods of human-computer interaction. It is gradually replacing traditional human-computer interaction methods, making human-computer interaction more convenient and faster. With the continuous development of speech synthesis technology, multilingual speech synthesis technology has become a research interest for researchers. This technology can realize the synthesis of different languages in a unified speech synthesis system [1–3].

There are lots of ethnic minorities in China. Many ethnic minorities have their own languages and scripts. Tibetan is one of the minority languages; it can be divided into three major dialects: Ü-Tsang dialect, Amdo dialect, and Kham dialect, which are mainly used in Tibet, Qinghai, Sichuan, Gansu, and Yunnan. All dialects use Tibetan characters as written text, but there are some differences in the pronunciation of each dialect, so it is difficult for the people who use different dialects to communicate with each other. There have been some research studies on Lhasa-Ü-Tsang dialect speech synthesis technology [4–12]. The end-to-end method [12] has more training advantages than the statistical parameter method, and the synthesis effect is better. There are few existing research studies on the speech synthesis of Amdo dialect, and only the work [13] applied the statistical parameter speech synthesis (SPSS) based on the hidden Markov model (HMM) for Tibetan Amdo dialect.

For the multilingual speech synthesis, the research works mainly use unit-selection concatenative synthesis technique, SPSS based on HMM, and deep learning technology. The unit-selection concatenative synthesis technique mainly includes selecting an unit scale, constructing a corpus and designing an algorithm of unit selection and splicing. This method relies on a large-scale corpus [14, 15]. Additionally, the synthesis effect is unstable and the connection of the splicing unit may have discontinuities. SPSS technology usually requires a complex text front-end to extract various

linguistic features from raw text, a duration model, an acoustic model, which is used to learn the transformation between linguistic features and acoustic features, and a complex signal-processing-based vocoder to reconstruct waveform from the predicted acoustic features. The work [16] proposes a framework for estimating HMM on data containing both multiple speakers and multiple languages, aiming to transfer a voice from one language to others. The works [2, 17, 18] propose a method to realize HMM-based cross-lingual SPSS using speaker adaptive training. For speech synthesis technology based on deep learning, the work [19] realizes a deep neural network- (DNN-) based Mandarin-Tibetan bilingual speech synthesis. The experimental results show that synthesized Tibetan speech is better than the HMM-based Mandarin-Tibetan cross-lingual speech synthesis. The work [20] trains the acoustic models with DNN, hybrid long short-term memory (LSTM), and hybrid bidirectional long short-term memory (BLSTM) and implements a deep learning-based Mandarin-Tibetan cross-lingual speech synthesis under a unique framework. Experiments demonstrated that the hybrid BLSTM-based cross-lingual speech synthesis framework was better than the Tibetan monolingual framework. Additionally, there are some research studies which reveal that multilingual speech synthesis using the end-to-end method gains a good performance. The work [21] presents an end-to-end multilingual speech synthesis model using a Unicode encoding "byte" input representation to train a model which outputs the corresponding audio of English, Spanish, or Mandarin. The work [22] proposes a multispeaker, multilingual TTS synthesis model based on Tacotron which is able to produce high-quality speech in multiple languages.

Taking into account that traditional methods require a lot of professional knowledge for phoneme analysis, tone, and prosody labelling, the work is time-consuming and costly, and the modules are usually trained separately, which will lead to the effect of error stacking [23] while the end-to-end speech synthesis system can automatically learn alignments and mapping from linguistic features to acoustic features. These systems can be trained on < text, audio > pairs without complex language-dependent text front-end. Inspired by above works, this paper proposes to use an end-to-end method to implement speech synthesis in Lhasa-Ü-Tsang and Amdo pastoral dialect, using a single sequence-to-sequence (seq2seq) architecture with attention mechanism as the shared feature prediction network for Tibetan multi-dialect and introducing two dialect-specific WaveNet networks to realize the generation of time-domain waveforms.

There are some similarities between this work and works [12, 24]. The WaveNet model is used in these works. However, in our work and [12], the WaveNet model is used for the generation of waveform sample with the input of predicted Mel spectrogram for speech synthesis. In the work [24] about speech recognition, the WaveNet model is used for the generation of text sequence and the input is MFCC features. The work [12] achieved the speech synthesis for Tibetan Lhasa-Ü-Tsang by using end-to-end model. In this paper, we improved the model of the work [12] to implement multidialect speech synthesis.

Our contributions can be summarized as follows. (1) We propose an end-to-end Tibetan multidialect speech synthesis model, which unifies all the modules into one model and realizes the speech synthesis for different Tibetan dialects using one speech synthesis system. (2) Joint learning is used to train the shared feature prediction network by learning the relevant features of multidialect speech data, and it is helpful to improve the speech synthesis performance of different dialects. (3) We use Wylie transliteration scheme to convert the Tibetan text into the corresponding Latin letters, which is used as the training units of the model. It effectively reduces the size of training corpus, reduces the workload of front-end text processing, and improves the modelling efficiency.

The rest of this paper is organized as follows. Section 2 introduces the end-to-end Tibetan multidialect speech synthesis model. The experiments are presented in detail in Section 3 and the results are discussed as well. Finally, we describe our conclusions in Section 4.

2. Model Architecture

The end-to-end speech synthesis model is mainly composed of two parts: the first part contains a seq2seq feature prediction network containing attention mechanism and the second part contains two dialect-specific WaveNet vocoders based on Mel spectrogram. The model adopts a synthesis method from text to intermediate representation and intermediate representation to speech waveform. The encoder and decoder implement the conversion from text to intermediate representation, and the WaveNet vocoders restore the intermediate representation into waveform samples. Figure 1 shows the end-to-end Tibetan multidialect speech synthesis model.

2.1. Front-End Processing. Although Tibetan pronunciation has evolved over thousands of years, the orthography of written language remains unchanged. It led to Tibetan spelling becoming very complicated. Tibetan sentence is written from left to right and consists of a series of single syllable. Single syllable is also called Tibetan character. The punctuation mark "·"means the "soundproof symbol" between the syllables, and the single hanging symbol "|" is used at the end of a phrase or sentence. Figure 2 shows a Tibetan sentence.

Each syllable in Tibetan has a root, which is the central consonant of the syllable. A vowel label can be added above or below the root to indicate different vowels. Sometimes, there is a superscript at the top of the root, one or two subscripts at the bottom, and a prescript at the front, indicating that the initials of the syllable are compound consonants. The sequence of connection of compound consonants is prescript, superscript, root, and subscript. Sometimes, there is one or two postscripts after the root, which means that the syllable has one or two consonant endings. The structure of Tibetan syllables is shown in Figure 3.

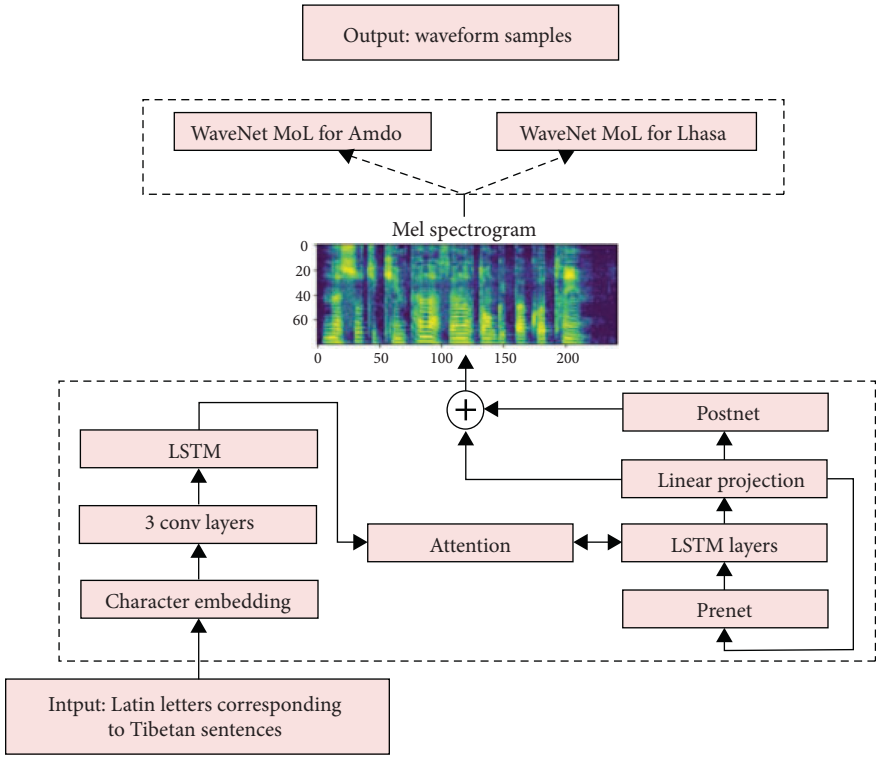


Figure 1: End-to-end Tibetan multidialect speech synthesis model.

Figure 2: A Tibetan sentence.

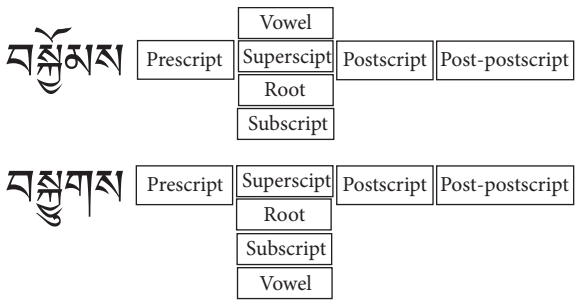


Figure 3: The structure of Tibetan syllables.

Due to the complexity of Tibetan spelling, a Tibetan syllable can have as many as 20450 possibilities. If a single syllable of a Tibetan character is used as the basic unit of speech synthesis, a large amount of speech data will need to be trained, and the corpus construction workload will be huge. The existing Tibetan speech synthesis system [10, 13] uses the initials and vowels of Tibetan characters as the input of the model, which requires a lot of professional knowledge of Tibetan linguistics and the front-end text processing. In this paper, we adopt the Wylie transliteration scheme, using only the basic 26 Latin letters, without adding letters and symbols, to convert the Tibetan text into the corresponding Latin letters. It effectively reduces the size of training corpus, reduces the workload of front-end text processing, and improves modelling efficiency. Figure 4 shows the converted Tibetan sentence obtained by using the Wylie transliteration scheme for the Tibetan sentence in Figure 2.

2.2. The Shared Feature Prediction Network. We use Lhasa-Ü-Tsang and Amdo dialect datasets to train the shared feature prediction network and capture the relevant features between two dialects speech data by joint learning. The shared feature prediction network is used to map the Latin transliteration vector of Tibetan character to Mel spectrograms. In this process, Lhasa-Ü-Tsang and Amdo dialect share the same feature prediction network. The shared feature prediction network consists of an encoder, an attention mechanism, and a decoder.

2.2.1. Encoder. The encoder module is used to extract the text sequence representation, including a character embedding layer, 3 convolutional layers, and a long short-term memory (LSTM) layer, as shown in the lower left part of Figure 1. Firstly, the input Tibetan characters are embedded into sentence vectors using character embedding layer and then input into 3 convolutional layers. These convolutional layers model longer-term context in the input character sequence, and the output of the final convolutional layer will

da res kyi don rkyen thog nas bzo pa'i gral rim ni blos 'gel chog pa zhig yin pa dang

Figure 4: A Tibetan sentence after Wylie transliteration.

be used as the input of a single bidirectional LSTM layer to generate intermediate representations.

2.2.2. Decoder. The decoder consists of a prenet layer, LSTM layers, and a linear projection layer, as shown in the lower right part of Figure 1. The decoder is an autoregressive recurrent neural network, which is used to predict the output spectrogram according to the encoded input sequence. The result of the previous prediction is first input to a prenet layer, and the output of the prenet layer and the output context vector of the attention mechanism network are concatenated and passed through 2 unidirectional LSTM layers. Then, the LSTM output and the attention context vector are concatenated and passed through a linear project layer to predict the target spectrogram frame. Finally, the predicted Mel spectrogram passes through a postnet, and the residual connection is made with the predicted spectrum to obtain the Mel spectrogram.

2.2.3. Attention Mechanism. The input sequence in the seq2seq structure will be compiled into a feature vector C of a certain dimension. The feature vector C always links the encoding and decoding stages of the encoder-decoder model. The encoder compresses the information of the entire sequence into a fixed-length vector. But with the continuous growth of the sequence, this will cause the feature vector to fail to fully represent the information of the entire sequence, and the latter input sequence will easily cover the first input sequence, which will cause the loss of many detailed information. To solve this problem, an attention mechanism is introduced. This mechanism will encode the encoder into different c_i, according to each time step of the sequence, that is, the original unified feature vector C will be replaced with a constantly changing c_i according to the current generated word. When decoding, combine each different c_i to decode the output so that when each output is generated, the information carried by the input sequence can be fully utilized, and the result will be more accurate.

The feature vector c_i is obtained by adding the hidden vector sequence $(h_1, h_2, \ldots, h_{T_x})$ during encoding according to the weight, as shown in equation (1). α_{ij} is the weight value, as in equation (2), which represents the matching degree between the jth input of the encoder and the ith output of the decoder.

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j, \tag{1}$$

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{T_x} \exp\left(e_{ik}\right)}. \tag{2}$$

2.3. WaveNet Vocoder. Tacotron [25, 26] launched by Google can convert phonetic characters or text data into frequency spectrum, and it also needs a vocoder to restore the frequency spectrum to waveforms to obtain synthesized speech. Tacotron's vocoder uses the Griffin–Lim algorithm for waveform reconstruction. The Griffin–Lim algorithm is an algorithm to reconstruct speech under the condition that only the amplitude spectrum is known and the phase spectrum is unknown. It is a relatively classic vocoder with simple and efficient algorithm. However, because the waveform generated by the Griffin–Lim vocoder is too smooth, the synthesized voice has a poor quality and sounds obviously "mechanical." WaveNet is a typical autoregressive generation model, which can improve the quality of synthetic speech. Therefore, this work uses the WaveNet model as a vocoder to cover the limitation of the Griffin–Lim algorithm. The sound waveform is a one-dimensional array in time domain, and the audio sampling points are usually relatively large. The waveform data at a sampling rate of 16 kHz will have 16000 elements per second, which requires a large amount of calculation using ordinary convolution. In this regard, WaveNet uses causal convolution, which can increase the receptive field of convolution. But causal convolution requires more convolutional layers, which is computationally complex and costly. Therefore, WaveNet has adopted the method of dilated causal convolution to expand the receptive field of convolution without significantly increasing the amount of calculation. The dilated convolution is shown in Figure 5. When the network generates the next element, it can use more previous element values. WaveNet is composed of stacked dilated causal convolutions and synthesizes speech by fitting the distribution of audio waveforms by the autoregressive method, that is, WaveNet predicts the next sampling point according to a number of input sampling points and synthesizes speech by predicting the value of the waveform at each time point waveform.

In the past, traditional acoustic and linguistic features were used as the input of the WaveNet model for speech synthesis. In this paper, we choose a low-level acoustic representation: Mel spectrogram, as the input of WaveNet for training. The Mel spectrogram emphasizes the details of low frequency, which is very important for the clarity of speech. And compared to the waveform samples, the phase of each frame in Mel spectrogram is unchanged; it is easier to train with the square error loss. We train WaveNet vocoders for Lhasa-Ü-Tsang dialect and Amdo pastoral dialect, and they can synthesize the corresponding Tibetan dialects with the corresponding WaveNet vocoder.

2.4. Training Process. Training process can be summarized into 2 steps: firstly, training the shared feature prediction network; secondly, training a dialect-specific WaveNet vocoder for Lhasa-Ü-Tsang dialect and Amdo pastoral dialect, respectively, based on the outputs generated by the network which was trained in step 1.

We trained the shared feature prediction network on the datasets of Lhasa-Ü-Tsang dialect and Amdo pastoral dialect. On a single GPU, we used the teacher-forcing method to train the feature prediction network, and the input of the

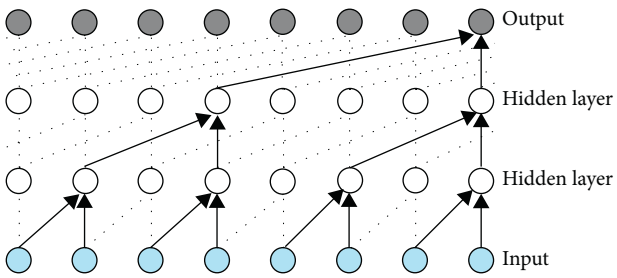


Figure 5: Dilated causal convolution [27].

decoder was the correct output, not the predicted output, with a batch size of 8. An Adam optimizer was used with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-6}$. The learning rate decreased from 10^{-3} to 10^{-4}, after 40000 iterations.

Then, the predicted outputs from the shared feature prediction network were aligned with the ground truth. We trained the WaveNet for Lhasa-Ü-Tsang dialect and Amdo pastoral dialect, respectively, by using the aligned predicted outputs. It means that these predicted data were generated in the teacher-forcing mode. Therefore, each spectrum frame is exactly aligned with a sample of the waveform. In the process of training the WaveNet network, we used an Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-6}$, and the learning rate was fixed at 10^{-3}.

3. Results and Analysis

3.1. Experimental Data. The training data consist of the Lhasa-Ü-Tsang dialect and Amdo pastoral dialect. The Lhasa-Ü-Tsang dialect speech data are about 1.43 hours with 2000 text sentences. The Amdo pastoral dialect speech data are about 2.68 hours with 2671 text sentences. Speech data files are converted to 16 kHz sampling rate, with 16 bit quantization accuracy.

3.2. Experimental Evaluation. In order to ensure the accuracy of the experimental results, we apply two methods, objective and subjective experiments, to evaluate the experimental results.

In objective experiment, the root mean square error (RMSE) of the time-domain sequences is calculated to measure the difference between the synthesized speech and the reference speech. The smaller the RMSE is, the closer the synthesized speech is to the reference and the better the effect of speech synthesis is. The formula of RMSE is shown in equation (3), where $x_{1,t}$ and $x_{2,t}$, respectively, represent the value of the time series of reference speech and synthesized speech at time t.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}\left(x_{1,t} - x_{2,t}\right)^2}{n}}. \tag{3}$$

For Lhasa-Ü-Tsang dialect and Amdo pastoral dialect, we randomly select 10 text sentences, use end-to-end Tibetan multi-dialect speech synthesis model for speech synthesis, and calculate the average RMSE to evaluate the closeness of the synthesized speech of the Lhasa-Ü-Tsang dialect and Amdo pastoral dialect to the reference speech. In order to evaluate the performance of the model, we compare it with the end-to-end Tibetan Lhasa-Ü-Tsang dialect speech synthesis model and end-to-end Tibetan Amdo pastoral dialect speech synthesis model. These two models were used to synthesize the same 10 text sentences, and the average RMSE was calculated. The results are shown in Table 1. For Lhasa-Ü-Tsang dialect, the RMSE of the multidialect speech synthesis model is 0.2126, which is less than the one of Lhasa-Ü-Tsang dialect speech synthesis model (0.2223). For Amdo pastoral dialect, the RMSE of the multidialect speech synthesis model is 0.1223, which is less than the one of Amdo pastoral dialect speech synthesis model (0.1253). It means that both Lhasa-Ü-Tsang dialect and Amdo pastoral dialect, which are synthesized by our model, are closer to their reference speech. The results show that our method has capability of the feature representation for both Lhasa-Ü-Tsang and Amdo pastoral dialect through the shared feature prediction network, so as to improve the multidialect speech synthesis performance against single dialect. Besides, the synthetic speech effect of Amdo pastoral dialect is better than that of Lhasa-Ü-Tsang dialect because the data scale of Amdo pastoral dialect is larger than that of Lhasa-Ü-Tsang dialect.

Figures 6 and 7, respectively, show the predicted Mel spectrogram and target Mel spectrogram output by the feature prediction network for Lhasa-Ü-Tsang dialect and Amdo pastoral dialect. It can be seen from the figures that the predicted mel spectrograms of Lhasa-Ü-Tsang dialect and Amdo pastoral dialect are both similar to the target Mel spectrograms.

In subjective experiment, the absolute category rating (ACR) measurement method was used to evaluate the synthesized speech of the Lhasa-Ü-Tsang and Amdo pastoral dialects mentioned above. In the ACR measurement, we selected 25 listeners. After listening to the synthesized speech, we used the original speech as a reference and scored the synthesized speech according to the grading standard in Table 2. After obtaining the scores given by all listeners, the mean opinion score (MOS) of the synthesized speech was calculated, and Table 3 shows the results. The MOS values of the synthesized speech in Lhasa-Ü-Tsang dialect and Amdo pastoral dialects are 3.95 and 4.18, respectively, which means that the synthesized speech has good clarity and naturalness.

3.3. Comparative Experiment. In order to verify the performance of the end-to-end Tibetan multidialect speech synthesis system, we have compared it with the "linear prediction amplitude spectrum + Griffin–Lim" and "Mel spectrogram + Griffin–Lim" speech synthesis system. The results of comparison experiment are shown in Table 4. According to Table 4, it can be seen that whether it is Lhasa-Ü-Tsang dialect or Amdo pastoral dialect, the MOS value of the synthesized speech of "Mel spectrogram + Griffin–Lim" speech synthesis system is higher than that of "linear prediction amplitude spectrum + Griffin–Lim" speech synthesis system. The results show that the Mel spectrogram is more effective as a predictive feature than the linear predictive

Table 1: Objective evaluation of the results.

Tibetan dialect	The RMSE of end-to-end Tibetan multidialect speech synthesis model	The RMSE of end-to-end Tibetan Lhasa-Ü-Tsang dialect speech synthesis model	The RMSE of end-to-end Tibetan Amdo pastoral dialect speech synthesis model
Lhasa-Ü-Tsang dialect	0.2126	0.2223	—
Amdo pastoral dialect	0.1223	—	0.1253

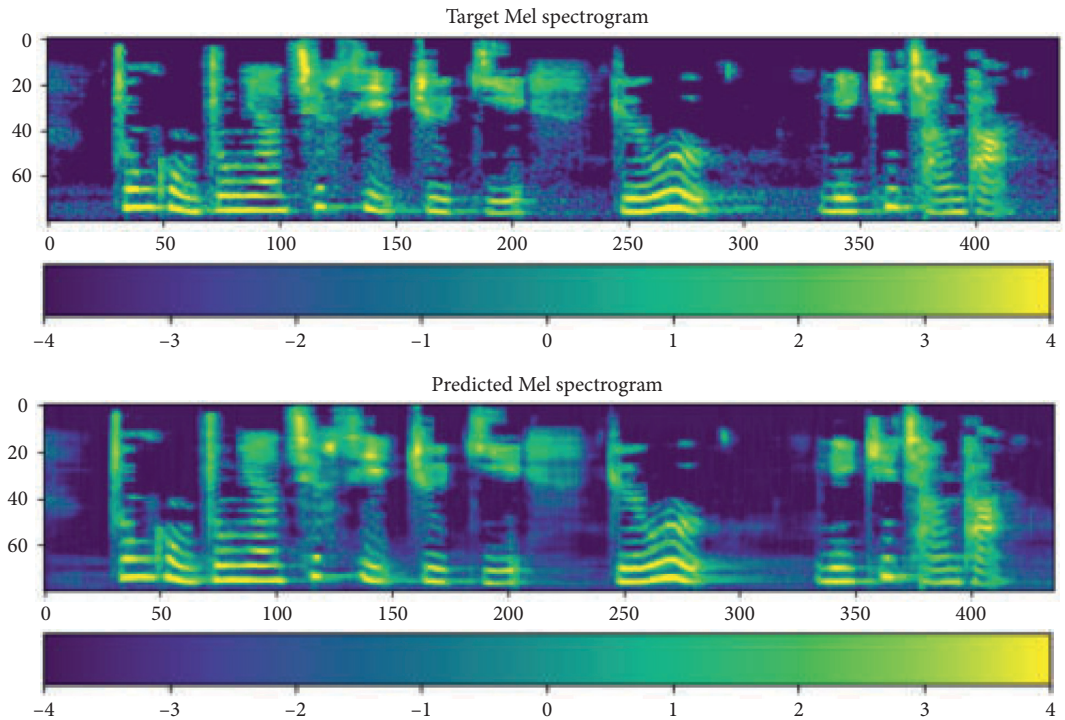


Figure 6: The comparison of the output Mel spectrogram and the target Mel spectrogram of Lhasa-Ü-Tsang dialect.

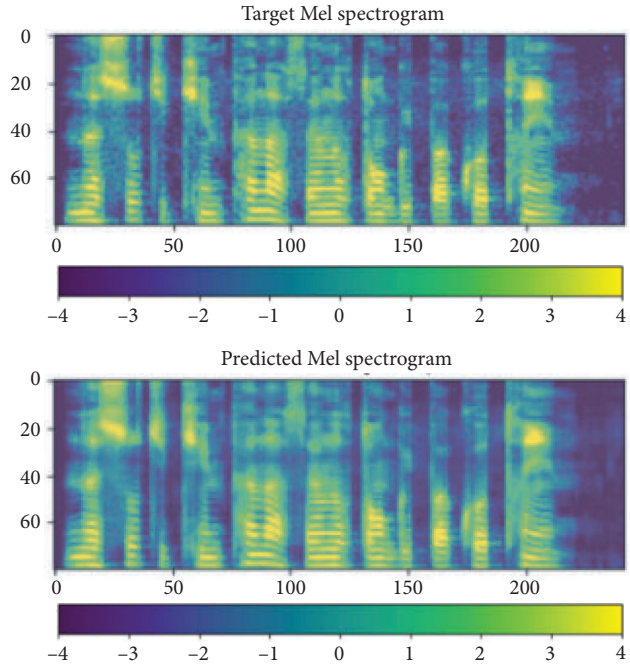


Figure 7: The comparison of the output Mel spectrogram and the target Mel spectrogram of Amdo pastoral dialect.

Table 2: Grading standards of ACR.

Grading value	Estimated quality
5	Very good
4	Good
3	Medium
2	Bad
1	Very bad

Table 3: The MOS comparison of speech synthesized by different synthesis primitive models.

Tibetan dialect	MOS
Lhasa-Ü-Tsang dialect	3.95
Amdo pastoral dialect	4.18

Table 4: The MOS comparison of speech synthesized by different models.

Model	MOS of Lhasa-Ü-Tsang dialect	MOS of Amdo pastoral dialect
Linear predictive amplitude spectrum + Griffin–Lim	3.30	3.52
Mel spectrogram + Griffin–Lim	3.55	3.70
Mel spectrogram + WaveNet	3.95	4.18

amplitude spectrum, and the quality of the generated speech is higher. The "Mel spectrogram + WaveNet" speech synthesis system outperforms the "Mel spectrogram + Griffin–Lim" speech synthesis system with the higher MOS value, which means that WaveNet has a better performance in recovering speech phase information and generating higher quality of the synthesis speech than the Griffin–Lim algorithm.

4. Conclusion

This paper builds an end-to-end Tibetan multidialect speech synthesis model, including a seq2seq feature prediction network, which maps the character vector to the Mel spectrogram, and a dialect-specific WaveNet vocoder for Lhasa-Ü-Tsang dialect and Amdo pastoral dialect, respectively, which synthesizes the Mel spectrogram into time-domain waveform. Our model can utilize dialect-specific WaveNet vocoders to synthesize corresponding Tibetan dialect. In the experiments, Wylie transcription scheme is used to convert Tibetan characters into Latin letters, which effectively reduces the number of composite primitives and the scale of training data. Both objective and subjective experimental results show that the synthesized speech of Lhasa-Ü-Tsang dialect and Amdo pastoral dialect has high qualities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China under grant no. 61976236.

References

[1] Y. J. Wu, Y. Nankaku, and K. Tokuda, "State mapping based method for cross-lingual speaker adaptation in HMM-based speech synthesis," in *Proceedings of the Interspeech, 10th Annual Conference of the International Speech Communication Association*, Brighton, UK, September 2009.

[2] H. Liu, *Research on HMM-Based Cross-Lingual Speech Synthesi*, University of Science and Technology of China, Hefei, China, 2011.

[3] R. Sproat, *Multilingual Text-To-Speech Synthesis: The Bell Labs Approach*, Kluwer Academic Publishers, Amsterdam, Netherland, 1998.

[4] Z. M. Cairang, *Research on Tibetan Speech Synthesis Technology Based on Mixed Primitives*, Shanxi Normal University, Xi'an, China, 2016.

[5] L. Gao, Z. H. Yu, and W. S. Zheng, "Research on HMM-based Tibetan Lhasa speech synthesis technology," *Journal of Northwest University for Nationalities*, vol. 32, no. 2, pp. 30–35, 2011.

[6] J. X. Zhang, *Research on Tibetan Lhasa Speech Synthesis Based on HMM*, Northwest University for Nationalities, Lanzhou, China, 2014.

[7] S. P. Xu, *Research on Speech Quality Evaluation for Tibetan Statistical Parametric Speech Synthesis*, Northwest Normal University, Lanzhou, China, 2015.

[8] X. J. Kong, *Research on Methods of Text Analysis for Tibetan Statistical Parametric Speech Synthesis*, Northwest Normal University, Lanzhou, China, 2017.

[9] Y. Zhou and D. C. Zhao, "Research on HMM-based Tibetan speech synthesis," *Computer Applications and Software*, vol. 32, no. 5, pp. 171–174, 2015.

[10] G. C. Du, Z. M. Cairang, Z. J. Nan et al., "Tibetan speech synthesis based on neural network," *Journal of Chinese Information Processing*, vol. 33, no. 2, pp. 75–80, 2019.

[11] L. S. Luo, G. Y. Li, C. W. Gong, and H. L. Ding, "End-to-end speech synthesis for Tibetan Lhasa dialect," *Journal of Physics: Conference Series*, vol. 1187, no. 5, 2019.

[12] Y. Zhao, P. Hu, X. Xu, L. Wu, and X. Li, "Lhasa-Tibetan speech synthesis using end-to-end model," *IEEE Access*, vol. 7, pp. 140305–140311, 2019.

[13] L. Su, *Research on the Speech Synthesis of Tibetan Amdo Dialect Based on HMM*, Northwest Normal University, Lanzhou, China, 2018.

[14] S. Quazza, L. Donetti, L. Moisa, and P. L. Salza, "Actor: a multilingual unit-selection speech synthesis syste," in *Proceedings of the 4th ISCA Workshop on Speech Synthesis*, Perth, Australia, 2001.

[15] F. Deprez, J. Odijk, and J. D. Moortel, "Introduction to multilingual corpus-based concatenative speech synthesis," in *Proceedings of the Interspeech, 8th Annual Conference of the International Speech Communication Association*, pp. 2129–2132, Antwerp, Belgium, August 2007.

[16] H. Zen, N. Braunschweiler, S. Buchholz et al., "Statistical parametric speech synthesis based on speaker and language

factorization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1713–1724, 2012.

[17] H. Y. Wang, *Research on Statistical Parametric Mandarin-Tibetan Cross-Lingual Speech Synthesis*, Northwest Normal University, Lanzhou, China, 2015.

[18] L. Z. Guo, *Research on Mandarin-Xingtai Dialect Cross-Lingual Speech Synthesis*, Northwest Normal University, Lanzhou, China, 2016.

[19] P. W. Wu, *Research on Mandarin-Tibetan Cross-Lingual Speech Synthesis*, Northwest Normal University, Lanzhou, China, 2018.

[20] W. Zhang, H. Yang, X. Bu, and L. Wang, "Deep learning for Mandarin-Tibetan cross-lingual speech synthesis," *IEEE Access*, vol. 7, pp. 167884–167894, 2019.

[21] B. Li, Y. Zhang, T. Sainath, Y. H. Wu, and W. Chan, "Bytes are all you need: end-to-end multilingual speech recognition and synthesis with bytes," in *Proceedings of the ICASSP*, Brighton, UK, May 2018.

[22] Y. Zhang, R. J. Weiss, H. Zen et al., "Learning to speak fluently in a foreign language: multilingual speech synthesis and cross-language voice cloning," 2019, https://arxiv.org/abs/1907.04448.

[23] Z. Y. Qiu, D. Qu, and L. H. Zhang, "End-to-end speech synthesis based on WaveNet," *Journal of Computer Applications*, vol. 39, no. 5, pp. 1325–1329, 2019.

[24] Y. Zhao, J. Yue, X. Xu, L. Wu, and X. Li, "End-to-end-based Tibetan multitask speech recognition," *IEEE Access*, vol. 7, pp. 162519–162529, 2019.

[25] R. Skerry-Ryan, E. Battenberg, Y. Xiao et al., "Towards end-to-end prosody transfer for expressive speech synthesis with tacotron," in *Proceedings of the International Conference on Machine Learning (ICML)*, Stockholm, Sweden, July 2018.

[26] Y. Wang, D. Stanton, Y. Zhang et al., "Style tokens: unsupervised style modeling, control and transfer in end-to-end speech synthesis," in *Proceedings of the International Conference on Machine Learning (ICML)*, Stockholm, Sweden, July 2018.

[27] A. V. D. Oord, S. Dieleman, H. Zen et al., "WaveNet: a generative model for raw audio," 2016, https://arxiv.org/abs/1609.03499.

Hindawi
Complexity
Volume 2021, Article ID 6657307, 13 pages
https://doi.org/10.1155/2021/6657307

Research Article

Optimizing Network Controllability with Minimum Cost

Xiao Wang and Linying Xiang

School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

Correspondence should be addressed to Linying Xiang; xianglinying@neuq.edu.cn

Received 8 December 2020; Revised 6 January 2021; Accepted 13 January 2021; Published 27 January 2021

Academic Editor: Ning Cai

In this paper, the issue of optimally modifying the structure of a directed network to guarantee its structural controllability is investigated. Given a directed network, in order to obtain a structurally controllable system, a framework for finding the minimum number of directed edges that need to be added to the network is proposed. After we get these edge-addition configurations, we further calculate the network cost of each optimization scheme and choose the one with the minimum cost. Our main contribution is twofold: first, we provide an algorithm able to find all optimal network modifications in polynomial time; second, we provide a way to calculate the cost of optimizing the network based on the node betweenness. Numerical simulations are given to illustrate the theoretical results.

1. Introduction

The ultimate goal of complex network research is to find effective means to control network behavior and make it serve human beings. Controllability is a basic concept in control theory, which quantifies the ability to control a dynamical system from any initial state to any final state in finite time [1]. In the past decade, the issue of network controllability for complex dynamical systems has attracted increasing attention and becomes a focal topic in interdisciplinary research [2–30]. Numerous works have been reported from rather diverse perspectives on such topics as structural controllability [2, 3]; exact controllability [18]; edge dynamics [19–21]; optimization [22–24]; control energy [25, 26]; and robustness [27, 28].

In the study of network controllability, we usually rely on the theory of structural controllability [31–37]. If there is a matrix pair that is controllable, all structurally equivalent matrix pairs are controllable except for special ill-conditioned cases [31]. Recently, those results have been applied to the controllability analysis of directed complex networks [2, 3, 16, 19, 22, 23] from a graph-theoretic perspective. Note that it is very effective to analyze network controllability by using tools developed under the background of structural control theory [31].

Optimization of the network controllability is of prime importance in real applications. Generally speaking, given a network which is structurally uncontrollable, we can make it structurally controllable through two strategies: (i) add external input signals to the original network [16] and (ii) add new edges to the network topology [23]. Wang et al. provided a method to change the structure of a complex network to make the system structurally controllable when only a single driver node was considered [22]. Zhang and Zhou considered three related problems on determining the minimal cost structural perturbations, including edge additions, edge deletions, and input deletions to make a networked system structurally controllable/uncontrollable [24]. Chen et al. proposed an approach to adding minimum directed edges to the original network so as to ensure structural controllability [23].

Motivated by the above discussions, a minimum-cost optimization method to guarantee structural controllability is investigated in this paper. It should be emphasized that, differing from [23], in this work, a new method is proposed to optimize network topology and thus to ensure the network controllability. Moreover, it also provides a way to calculate the total cost of optimizing the network. However, in [23], it only gives a method to optimize the network

topology without considering the optimization cost. Note that calculating the optimization cost is exactly the major point in this work. In [27], Zhang et al. considered the problem of network cost. Although the measurement index of edge cost was given therein, it did not provide a simple and effective method to calculate the total network cost. Compared with the previous works, we not only address the problem of optimizing network controllability but also propose a way to calculate the cost of optimizing the network. The main contributions of this article are as follows. (i) We propose a new method to optimize the network topology so as to ensure the network controllability. (ii) We propose an algorithm to solve the optimal edge-addition configuration problem. (iii) After getting all the edge-addition configurations, we introduce network cost measurement indexes to calculate the cost of optimizing the network. Based on which, we can determine the optimal edge-addition configuration with minimum-cost. The results of this paper can provide both theoretical and technical guidance for the analysis and control of real complex networks. The obtained results shed some lights on the transformation of a structurally uncontrollable network to a structurally controllable one with a low cost. For example, in the power network, transmission lines with the lowest cost can be set up among substations to safely and efficiently control the entire power network.

The rest of the paper is organized as follows. Section 2 introduces the notation and terminology used in this paper. Problem formulation and preliminaries on graph theory are introduced in Section 3. The main results are given in Section 4. In Section 5, a network cost index is given to determine the minimum-cost edge-addition configuration. Finally, the summary of this paper and the prospect of future research are presented in Section 6.

2. Notations

In this paper, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}^m$ is the space of real m-vectors, and $\mathbb{R}^{m\times n}$ is the space of $m \times n$ real matrices. For a set $\mathcal{S}$, its cardinality is denoted by $|\mathcal{S}|$.

A directed graph $G = (V, E)$ consists of a node set $V = \{1, 2, \ldots, n\}$ and an edge set $E = \{(i, j)\}$. Here, $(i, j) \in E$ implies that there exists a directed edge from node i to node j, and i and j are called the parent node and the child node, respectively. We can also say that the tail node i is pointing toward the head node j. For a digraph G, a directed path of length $k + 1$ from node i to node j is defined as a sequence of distinct edges of the form $(i, i_1), (i_1, i_2), \ldots, (i_k, j)$, in which all nodes $i, i_1, \ldots, i_k, j$ are distinct. Here, node i is called the beginning node and j the end node of the directed path. A node $i_2 \in V$ is reachable from $i_1 \in V$ if there exists a directed path in G from i_1 to i_2. A directed graph $G_s = (V_s, E_s)$ is a subgraph of G if $V_s \subseteq V$ and $E_s \subseteq E$. A directed graph is said to be strongly connected if there exists a directed path between any two nodes. A strongly connected component (SCC) is a maximal subgraph G_s that is strongly connected. Particularly, a source SCC has no incoming edges from another SCC.

A digraph G contains a dilation if there is a subset of nodes $S \subset V$ such that the common-neighbor set of S, denoted by $T(S)$, has fewer nodes than S itself, i.e., $|T(S)| < |S|$. Here, $T(S)$ is the set of nodes j, in which there is a directed edge from node j to some other node in S. Notice that a digraph G contains no dilation if each node has its own independent parent node. It is intuitively plausible that a dilation is a subgraph containing a relatively large number of nodes that are "dominated" by a small number of other nodes.

3. Problem Statement and Preliminaries

Consider a linear time-invariant (LTI) networked dynamical system described by

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{1}$$

where $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T \in \mathbb{R}^n$ is the state vector of all nodes; $u(t) = [u_1(t), u_2(t), \ldots, u_m(t)]^T \in \mathbb{R}^m$ is the input vector; $B = (b_{ij}) \in \mathbb{R}^{n\times m}$ is the input matrix identifying the nodes that are directly controlled, and $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ is the adjacency matrix of the underlying network. The overall networked system described by (1) can be denoted by the matrix pair (A, B).

Definition 1. Linear network (1) is said to be state controllable if, for any initial state $x(t_0) \in \mathbb{R}^n$ and any final state $x(t_f) \in \mathbb{R}^n$, there exist a finite time t_1 and an input $u(t) \in \mathbb{R}^m$, $t \in [t_0, t_1]$, such that $x(t_1; x(t_0), u) = x(t_f)$.

If networked system (1) is state controllable, we can say that the matrix pair (A, B) is state controllable.

Definition 2 (see [16, 31]). A linear control system (A, B) is a structured system if the elements in A and B are either fixed zeros or independent nonzero parameters. Both the two matrices A and B are called structured matrices.

In this paper, it is assumed that we only know the structure of the matrices A and B. This means that we know which elements in the matrices are fixed to zero and consequently which elements are nonzero free parameters.

Definition 3. A linear control system (A, B) is structurally controllable if we can set some values to the nonzero parameters in A and B such that the resulting system is state controllable in the sense of Kalman defined in Definition 1.

A structured system can be represented by a directed graph whose nodes denote the (state and input) variables and edges indicate the connections between some variables [31]. In this paper, a structured system (A, B) is denoted by a directed graph $G(A, B) = (V, E)$, in which $V = V_A \cup V_B$ is the node set and $E = E_{V_A,V_A} \cup E_{V_B,V_A}$ is the edge set. In particular, $V_A = \{x_1, x_2, \ldots, x_n\}$ is the set of state nodes, corresponding to the n nodes in the original network; $V_B = \{u_1, u_2, \ldots, u_m\}$ is the set of input nodes corresponding to the m inputs; $E_{V_A,V_A} = \{(x_i, x_j) | a_{ji} \neq 0\}$ is the set of edges

between state nodes; and $E_{V_B,V_A} = \{(u_i, x_j) \mid b_{ji} \neq 0\}$ is the set of edges between input nodes and state nodes. In the whole paper, suppose that any input signal is applied to only one node, referred to as a driver node. A state node being reachable means that there is a directed path from some input node to this state node. Similarly, a node set is reachable if each node in the set is reachable. Notice that, in the remaining of the paper, unless otherwise specified, the reachability is only used for the state nodes.

In a digraph, an edge subset $\widetilde{M}$ is a matching if no two edges in $\widetilde{M}$ share a common parent node or a common child node. A matching of maximum size is called a maximum matching. The maximum matching of a digraph can be denoted by mapping the digraph to its bipartite representation. Consider a directed network $G(A, B)$, whose bipartite representation can be described by $\mathscr{B}(A, B) = \mathscr{B}(V_A^+ \cup V_B, V_A^-, E_{V_A^+,V_A^-} \cup E_{V_B,V_A^-})$, in which $V_A^+ = \{x_1^+, x_2^+, \ldots, x_n^+\}$ and $V_A^- = \{x_1^-, x_2^-, \ldots, x_n^-\}$. That is, each state node x_i of the original digraph is split into two nodes x_i^+ and x_i^-. Here, $\{x_i^+, x_j^-\} \in E_{V_A^+,V_A^-}$ if $(x_i, x_j) \in E_{V_A,V_A}$ and $\{u_i, x_j^-\} \in E_{V_B,V_A^-}$ if $(u_i, x_j) \in E_{V_B,V_A}$. To describe the relationship between the digraph and its bipartite graph, we use a signal-notation mapping $f\colon E_{V_A,V_A} \cup E_{V_B,V_A} \longrightarrow E_{V_A^+,V_A^-} \cup E_{V_B,V_A^-}$ to map directed edges from the system digraph into undirected edges of the system bipartite graph as follows: $f((u_i, x_j)) = \{u_i, x_j^-\}$ and $f((x_i, x_j)) = \{x_i^+, x_j^-\}$. Also, we have that $f^{-1}(\{u_i, x_j^-\}) = (u_i, x_j)$ and $f^{-1}(\{x_i^+, x_j^-\}) = (x_i, x_j)$.

Definition 4. The element $r_{ij} = 1$ in the matrix $R \in \mathbb{R}^{n\times n}$ if there is a directed path from node i to node j $(i \neq j)$. Set $r_{ii} = 1$, $i = 1, 2, \ldots, n$. The matrix R is called reachable matrix.

If only one external input is applied to node 1, then the first row of the matrix R can be used to determine which nodes are unreachable.

Definition 5. The element $p_{ij} = 1$ in the matrix P if edge (i, j) is one of the matching edges of a maximum matching about a bipartite graph. The matrix P is called maximum matching matrix.

The maximum matching of a directed graph is not unique. Therefore, the corresponding maximum matching matrix P is not unique. It can be found from the matrix P that the number of nonzero elements in the matrix P is the number of matching edges in the maximum matching, and each row and each column have at most one nonzero element. The jth column is full of zero elements, indicating that node j in the network does not have its own independent parent node.

Definition 6. Consider a directed network, in which only one external input signal is applied to node 1. If $n = \sum_{j=1}^{n} r_{1j}$, $r_{1j} \in R$, then such reachable matrix R is called $1 - R$ matrix. For example,

$$1 - R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{2}$$

Obviously, if the reachable matrix R of a network is a $1 - R$ matrix, then all the state nodes in the network are reachable.

Definition 7. Consider a directed network, in which only one external input signal is applied to node 1. If the maximum matching matrix P has a unique nonzero element in each column except for the first column, then such maximum matching matrix P is called $1 - P$ matrix. For example,

$$1 - P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3}$$

Obviously, if the maximum matching matrix P of a network is $1 - P$ matrix, then there is no dilation in the network.

A necessary and sufficient condition for the structural controllability of an LTI system is given as follows [31].

Lemma 1 (see [31]). *The pair (A, B) is structurally controllable if and only if the following two conditions are satisfied simultaneously:*

(1) Every state node $x \in V_A$ in the digraph $G(A, B) = (V_A \cup V_B, E_{V_A,V_A} \cup E_{V_B,V_A})$ is reachable from some input node $u \in V_B$

(2) The digraph $G(A, B)$ contains no dilations

Then, we have the following controllability criterion.

Theorem 1. *A directed network $G(A, B)$ with $B = [b_1, 0, \ldots, 0]^T$ is structurally controllable if and only if the following two conditions are satisfied simultaneously:*

(1) The reachable matrix of $G(A, B)$ is a $1 - R$ matrix

(2) The maximum matching matrix of $G(A, B)$ is a $1 - P$ matrix

In this paper, given a structurally uncontrollable directed network, we study the problem of adding the least edges to improve the topology so as to obtain a structurally controllable system. After we get these optimal edge-addition configurations, we need to calculate the network cost of each optimization scheme and choose the one with the minimum cost. In summary, the problem is given as follows.

Problem 1. Given the pair (A, B) with $B = [b_1, 0, \ldots, 0]^T$, find

$$\widetilde{A}^* = \arg \min_{\widetilde{A} \in \{0,1\}^{n \times n}} \|\widetilde{A}\|_0, \tag{4}$$

s.t. the reachable matrix of digraph $G(A + \widetilde{A}, B)$ is a $1 - R$ matrix and the maximum matching matrix is a $1 - P$ matrix,

where $\|\widetilde{A}\|_0$ denotes the number of nonzero elements in a matrix $\widetilde{A}$.

If $(A + \widetilde{A}, B)$ is structurally controllable, we refer to the matrix $\widetilde{A}$ as an effective perturbed matrix and to $\widetilde{A}^*$ in (4) as the modified matrix. The aim of this paper is to provide a characterization of all possible modified matrices by using graph-theoretical tools and design an algorithm to obtain such a solution.

4. Network Topology Optimization to Ensure Structural Controllability

Note that the system digraph is denoted by $G(A, B) = (V_A \cup V_B, E_{V_A,V_A} \cup E_{V_B,V_A})$. Therefore, given an effective perturbed matrix $\widetilde{A}$, we can relate a digraph to the perturbed structured system $(A + \widetilde{A}, B)$, which we denote by $G(A + \widetilde{A}, B) = (V_A \cup V_B, E_{V_A,V_A} \cup E_{V_B,V_A} \cup \widetilde{E})$, where the edge set $\widetilde{E} \subseteq V_A \times V_A$ is such that $(x_i, x_j) \in \widetilde{E}$ if and only if $\widetilde{a}_{ji} = 1$. Since the matrix $\widetilde{A}$ is closely related to the $\widetilde{E}$, we can rewrite Problem 1 in a different way.

Problem 2. Given the system digraph $G(A, B) = (V_A \cup V_B, E_{V_A,V_A} \cup E_{V_B,V_A})$ with $B = [b_1, 0, \ldots, 0]^T$, find

$$\widetilde{E}^* = \arg \min_{\widetilde{E} \subseteq V_A \times V_A} |\widetilde{E}|, \tag{5}$$

s.t. the reachable matrix of the digraph $G(A + \widetilde{A}, B) = (V_A \cup V_B, E_{V_A,V_A} \cup E_{V_B,V_A} \cup \widetilde{E})$ is a $1 - R$ matrix and the maximum matching matrix is a $1 - P$ matrix.

Additionally, define a feasible edge-addition configuration as a set of directed edges that is a feasible solution of Problem 2.

The solutions to Problem 2 are given in this section. First, a definition is introduced to describe the smallest set of edges needed to achieve reachability, i.e., satisfy condition (1) in Lemma 1. Let $G(A, B) = (V_A \cup V_B, E_{V_A,V_A} \cup E_{V_B,V_A})$ be the system digraph. The set of state nodes V_A can be divided into two sets based on their reachability, namely, $V_A = \overline{R} \cup \overline{N}$, where $\overline{R}$ is the set of reachable nodes and $\overline{N}$ is the set of unreachable nodes. In addition, assume that there are r source SCCs that are unreachable, whose node sets are denoted by $N_1, N_2, \ldots, N_r \subseteq \overline{N}$. In order to make the nodes in these unreachable source SCCs reachable, we need to add a new edge between the reachable node and the node in the source SCC so that all the nodes in the source SCC are reachable. Moreover, since the source SCC has outgoing edges pointing to other nodes, the unreachable nodes that are connected to the source SCC will also become reachable.

Definition 8. A set S_E is made up of connected edges, then the set S_E is called the connected edge set. Here, the connected edge refers to the connecting edge between the reachable node and the unreachable node.

Algorithm 1 is illustrated in Figure 1. The connected edge set contains the minimum number of added edges required to ensure that all the state nodes are reachable. Obviously, the connected edge set can only satisfy condition (1) in Lemma 1 and cannot guarantee the structural controllability of the networked system. To ensure structural controllability of the system, these edge additions must satisfy two conditions: (i) a set of connected edges and (ii) the "tail" node of the new edge is not used as an independent parent node in the maximum matching. It is the "head" node of the edge that has no independent parent node.

Theorem 2. *Consider a directed network $G(A, B)$, whose bipartite representation is denoted by $\mathscr{B}(A, B)$. Let M be a maximum matching, $U_o(M) = \{v_i^o\colon i \in \{1, 2, \ldots, n_o\}\}$ be a node set in which each node is not used as independent parent node, and $U_r(M) = \{v_i^r\colon i \in \{1, 2, \ldots, n_r\}\}$ be a node set with no independent parent nodes. A set $\widetilde{E}$ is a feasible edge-addition configuration if and only if it contains the union of the following two sets:*

(1) S_E is the set of connected edges

(2) $S_M = \{f^{-1}(\{v_i^o, v_i^r\})\colon v_i^o \in U_o(M), v_i^r \in U_r(M), i = \{1, 2, \ldots, n_r\}\}$

Theorem 2 provides some feasible edge-addition configurations, but we need to find the optimal one from these configurations. Therefore, the first task is to select the optimal solution from these feasible solutions. From the above discussion, it can be found that, after determining the maximum matching of a bipartite graph, if those unmatched nodes (nodes without independent parent nodes) happen to be distributed in different source SCCs, then the added edges just meet both conditions in Lemma 1, which is exactly what is needed. To explore this situation, we introduce the following concepts.

Definition 9. Consider a directed network $G(A, B)$, whose bipartite representation is denoted by $\mathscr{B}(A, B)$. Let M be a maximum matching associated with $\mathscr{B}(A, B)$. Moreover, let $U_r(M)$ be the set of nodes in which each node has no independent parent nodes. If there is at least one node i, $i \in U_r(M)$ in an unreachable source SCC, then such an unreachable source SCC is called an ideal source SCC.

Whether an unreachable source SCC is an ideal source SCC depends mainly on the specific maximum matching. Because there may be more than one maximum matching corresponding to a directed network, it is not possible to determine whether a node has an independent parent node in the maximum matching.

Definition 10. The N_s of the directed network $G(A, B)$ is defined as the maximum number of ideal source SCCs in all the maximum matchings.

Input: reachable nodes sets $\overline{R}$ and unreachable nodes sets $\overline{N}$
(1) Order the unreachable source SCCs: $N_1, N_2, \ldots, N_r$
(2) Select any edge (i, j) in which i is in the set of reachable nodes and j is in the first source SCC
(3) Merge all reachable state nodes into a larger set (we can do it using either BFS/DFS or union-find)
(4) Call Steps 2-3 recursively until all unreachable source SCCs become reachable

Algorithm 1: Set of connected edges.

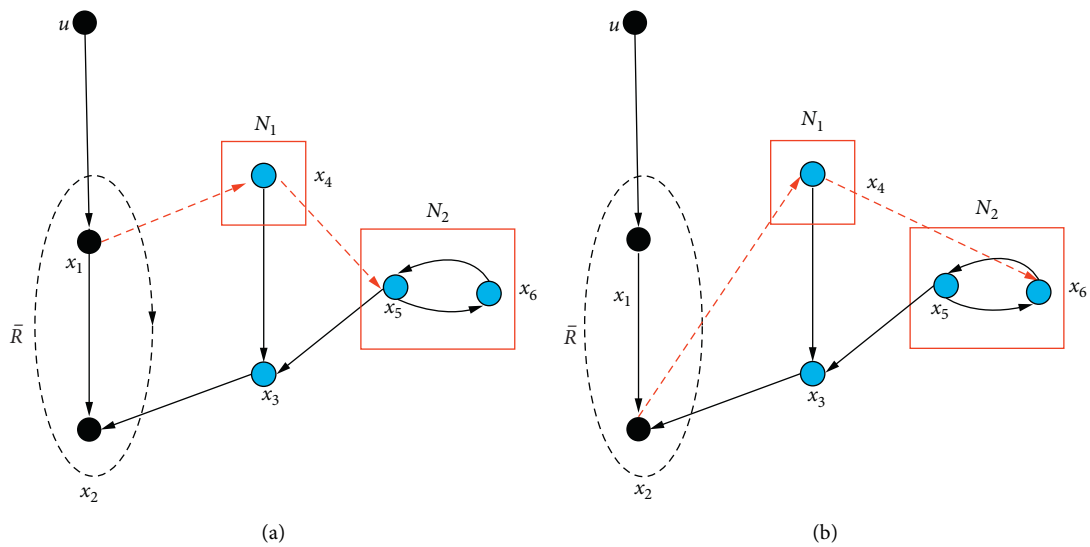


Figure 1: Illustration of Algorithm 1. The black and blue nodes and all the black edges consist of the original directed network $G(A, B)$. The black nodes, except for the input node u, constitute the set of reachable state nodes $\overline{R} = \{x_1, x_2\}$. Blue nodes constitute the set of unreachable state nodes $\overline{N} = \{x_3, x_4, x_5, x_6\}$. There are two unreachable state source SCCs, N_1 and N_2. In (a), we give a possible edge-addition configuration for Algorithm 1. First, we add edge (x_1, x_4) to S_E. Then, the state node x_4 from N_1 becomes reachable, and thus the state node x_3 becomes reachable. Next, we add edge (x_4, x_5) to S_E, and then the state nodes x_5 and x_6 from N_2 become reachable, i.e., $S_E = \{(x_1, x_4), (x_4, x_5)\}$. In (b), we add edges (x_2, x_4) and (x_4, x_6) to S_E', i.e., $S_E' = \{(x_2, x_4), (x_4, x_6)\}$. Therefore, S_E and S_E' are two possible sets of connected edges.

We can determine a maximum matching attaining N_s using Algorithm 2.

We take Figure 2, for example, to illustrate Algorithm 2.

The reachable matrix corresponding to the digraph in Figure 2(a) is expressed as follows:

$$R = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{6}$$

The unreachable node set can be determined as $\overline{N} = \{x_3, x_5, x_6\}$ by the position of the 0 element in the first row of R. Moreover, there are two unreachable source SCCs (red box), whose node sets are $N_1 = \{x_3\}$ and $N_2 = \{x_5, x_6\}$, respectively. Then, we can label columns 3, 5, and 6 of R as follows:

$$R = \begin{bmatrix} & & * & & * & * \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{7}$$

Figure 2(b) shows the bipartite representation of the original directed network (Figure 2(a)). In order to make the column ordinals corresponding to all 0 columns in the maximum matching matrix P coincide with the marked column ordinals as much as possible, an ideal maximum matching M is determined in Figure 2(c), and its corresponding maximum matching matrix is expressed as follows:

Input: A directed network $G(A, B)$ with $B = [b_1, 0, \ldots, 0]^T$;
(1) Write the reachable matrix R of the directed network, and determine the unreachable node set $\overline{N}$ in the network by the position (column ordinal) of the 0 element in the first row.
(2) Find the unreachable source SCCs.
(3) Select the nodes located in the source SCCs from the unreachable nodes set $\overline{N}$ and mark their column ordinals.
(4) By using the marked column ordinals to identify an ideal maximum matching M. Its corresponding maximum matching matrix is P^*. The column ordinals corresponding to all 0 columns in the matrix P^* need to match the marked column ordinals as much as possible.
(5) According to Step 3, an ideal maximum matching matrix P^* can be obtained. From the matrix P^*, the nodes corresponding to the matching column ordinals can be found.
(6) Based on the distribution of the nodes found in Step 5 in the source SCCs, N_s can be calculated.

Algorithm 2: Determine the ideal maximum matching to get N_s.

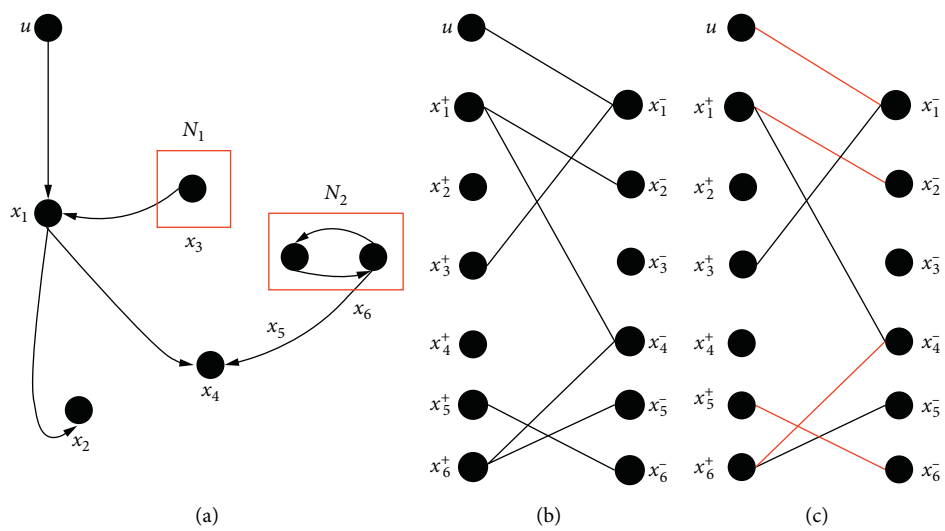


Figure 2: Example illustrating Algorithm 2.

$$P^* = \begin{bmatrix} & & * & & * & \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \quad (8)$$

There are at most two 0 columns in P^* that are consistent with the marked column ordinals, and the corresponding node x_3 is located in N_1, and node x_5 is located in N_2, so $N_s = 2$.

If all the state nodes that are not used as independent parent nodes are unreachable, then additional edges are needed to satisfy condition (1) in Lemma 1. Therefore, in this case, calculating N_s according to Algorithm 2 does not necessarily lead to an optimal configuration of added edges. To illustrate this statement, we take Figure 3 for example.

Next, we will propose Algorithm 3 to solve Problem 2. Algorithm 3 is mainly divided into the following four steps:

Step 1. All the state nodes in the directed network are classified into a reachable node set and an unreachable node set, respectively, based on the node reachability.

Step 2. Determine the ideal maximum matching to get N_s. If there exist some unreachable nodes that are not used as independent parent nodes in the ideal maximum matching, then we alter the matching by finding a directed path rooted at the input node.

Step 3. Add some edges to satisfy Lemma 1. These edges start at reachable nodes that are not used as independent parent nodes and end at nodes that have no independent parent nodes in unreachable source SCCs.

Step 4. If there are unreachable nodes that are not used as independent parent nodes, then we need to add a set of connected edges to ensure that both two conditions of Lemma 1 are satisfied.

Given a structurally uncontrollable system (A, B) that contains unreachable nodes and/or dilations. Therefore, we need to optimize the network topology to ensure structural controllability by adding edges. Algorithm 3 is given to obtain optimal edge-addition configuration to solve Problem 2.

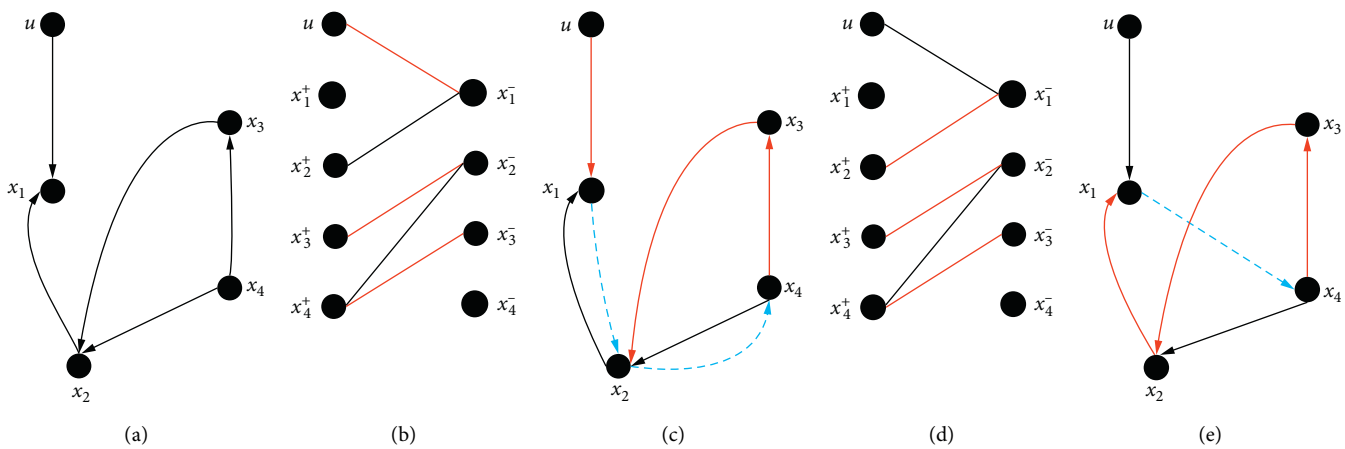


FIGURE 3: The maximum matching of a directed graph is not unique, and different maximum matchings will result in different feasible edge-addition configurations. In (a), the initial system digraph $G(A, B)$ is given. The red edges in (b) and (d) form two different maximum matchings. The red edges in (c) and (e) are determined by the maximum matchings in (b) and (d), respectively. In (c), after determining the maximum matching, node x_4 has no independent parent node and node x_2 has not been used as the parent node. So, we need to add the edge (x_2, x_4) to satisfy condition (2) of Lemma 1. Since node x_2 is unreachable, we also need to add the edge (x_1, x_2) to satisfy condition (1) of Lemma 1. Then, we have $\widetilde{E}_1 = \{(x_1, x_2), (x_2, x_4)\}$. In (e), after determining the maximum matching, node x_4 has no independent parent node and node x_1 has not been used as the parent node. So, we can add edge (x_1, x_4) to satisfy both two conditions of Lemma 1, i.e., $\widetilde{E}_2 = \{(x_1, x_4)\}$. Therefore, $\widetilde{E}_2$ is an optimal edge-addition configuration but $\widetilde{E}_1$ is not.

Input: A directed network $G(A, B)$;
(1) All the state nodes in the network are classified into a reachable node set $\overline{R}$ and an unreachable node set $\overline{N}$. Then, determine the unreachable source SCCs in the directed network $G(A, B)$.
(2) Using Algorithm 2 to get M' and N_s.
(3) if $U_o(M') \cap \overline{R} = \varnothing$, then
(4) Find an unreachable node x_j, and thus add the edge $(x_i, x_j), x_i \in \overline{R}$;
(5) $M \leftarrow M' \cup \{(x_i, x_j)\}$;
(6) else
(7) Set $M = M'$;
(8) end if
(9) Obtain the unique set of disjoint directed paths $L = \cup_{i=1}^{q} L_i$ in M, where the beginning node of each L_i is in some unreachable source SCCs and the end node is not used as a separate parent node;
(10) Let $Q = \{q_1, q_2, \ldots, q_n\}$ and $Z = \{z_1, z_2, \ldots, z_n\}$, q_i, z_i are the beginning and end nodes of each path L_i, respectively;
(11) Let $\widetilde{E}^* \leftarrow \varnothing, k \leftarrow 1$;
(12) if $Z \cap \overline{R} = \varnothing$, then
(13) Find a reachable node $v_o, v_o \in U_o(M)$;
(14) for $k \leq q$ do
(15) $\widetilde{E}^* \leftarrow \widetilde{E} \cup \{v_{k-1}, q_k\}; k \leftarrow k + 1$;
(16) $Z \subseteq \overline{R}$
(17) end for
(18) if $x_i^+ \in U_o(M), U_r(M) \neq \varnothing$, then
(19) $\widetilde{E}^* \leftarrow \widetilde{E} \cup \{(x_i, x_j)\}, x_j^- \in U_r(M)$;
(20) $U_o(M) \leftarrow U_o(M) \backslash x_i^+; U_r(M) \longleftarrow U_r(M) \backslash x_j^-$;
(21) when $U_r(M) = \varnothing$
(22) end if

ALGORITHM 3: Minimal edge addition.

Next, an example in Figure 4 is given to illustrate Algorithm 3.

5. Network Optimization Cost

We have solved the optimal edge-addition configuration problem; however, there are multiple potential edge-addition configurations to ensure structural controllability. From the application perspective, the lowest cost configuration is usually selected as the final optimization solution. Therefore, we present Problem 3 based on Problem 2, taking the network cost into account. In order to solve Problem 3, we introduce an edge cost measurement index to calculate the edge cost and thus obtain the cost of the whole network.

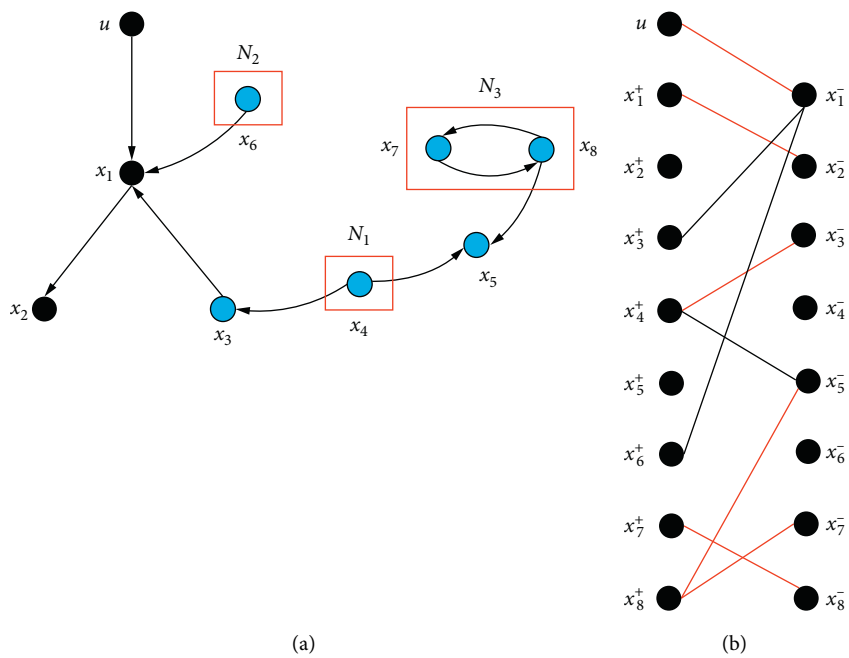


Figure 4: In (a), the directed network $G(A, B)$ contains a single input node u and eight state nodes $x_1, x_2, \ldots, x_8$. We will first decompose the directed graph according to the first step of Algorithm 3, $\overline{R} = \{x_1, x_2\}$, $\overline{N} = \{x_3, x_4, x_5, x_6, x_7, x_8\}$. There are three unreachable source SCCs in the digraph, $N_1 = \{x_4\}, N_2 = \{x_6\}, N_3 = \{x_7, x_8\}$. In (b), we provide $\mathscr{B}(A, B)$ the bipartite graph of the directed graph to attain a maximum matching M' (red edges) according to Step 2 of Algorithm 3, i.e., $M' = \{(u, x_1), (x_1, x_2), (x_4, x_3), (x_8, x_5), (x_7, x_8)\}$. According to the maximum matching, nodes x_4, x_6, and x_7 have no independent parent node, $U_r(M') = \{x_4, x_6, x_7\}$. The nodes x_2, x_3, x_5, and x_6 are not used as the independent parent node, $U_o(M') = \{x_2, x_3, x_5, x_6\}$. According to Step 3 of Algorithm 3, reachable node x_2 is not used as the parent node, $x_2 \in \overline{R}, x_3, x_5, x_6 \in \overline{N}$. Therefore, we need to pick nodes in $U_o(M')$ and $U_r(M')$, respectively, and they form edges that make nodes x_3, x_5, and x_6 become reachable. The added edge (x_2, x_4) can satisfy the above conditions $(U_o(M) \subseteq \overline{R})$. After the edge (x_2, x_4) is added, a new maximum matching $M = M' \cup \{(x_2, x_4)\}$ is formed. Then, the remaining set in which each node has not been used as an independent parent node is $U_o(M) = \{x_5, x_3, x_6\}$. The set with no independent parent node is $U_r(M) = \{x_6, x_7\}$. According to Step 18 of Algorithm 3, $U_r(M) = \{x_6, x_7\} \neq \varnothing$. So, we need to keep adding edges until the two conditions of Lemma 1 are satisfied. The other two edges added have four choices: $\{(x_5, x_6), (x_3, x_7)\}$, $\{(x_5, x_7), (x_3, x_6)\}$, $\{(x_5, x_6), (x_6, x_7)\}$, and $\{(x_3, x_6), (x_6, x_7)\}$. Finally, after adding three edges to the graph, both two conditions of Lemma 1 are satisfied, and thus a new directed graph $G(A + \widetilde{A}, B)$ is obtained. In summary, we can get four optimal edge-addition configurations as follows: $\widetilde{E}_1^* = \{(x_2, x_4), (x_5, x_6), (x_3, x_7)\}$, $\widetilde{E}_2^* = \{(x_2, x_4), (x_5, x_7), (x_3, x_6)\}$, $\widetilde{E}_3^* = \{(x_2, x_4), (x_5, x_6), (x_6, x_7)\}$, and $\widetilde{E}_4^* = \{(x_2, x_4), (x_6, x_7), (x_3, x_6)\}$ (a solution to Problem 2).

In addition, we need to adopt a simple and practical method to calculate the cost of the network and determine a minimum-cost configuration to ensure the controllability based on the optimal edge-addition configuration.

Problem 3. Consider a directed network $G(A, B)$, find

$$\widetilde{E}^* = \arg \min_{\widetilde{E} \subseteq V_A \times V_A} |\widetilde{E}|, \tag{9}$$

s.t. the new directed network $G(A + \widetilde{A}, B)$ contains neither unreachable nodes nor dilations. Also, the cost of the new directed network must be the lowest one.

5.1. Main Idea. Given a structurally uncontrollable directed network $G(A, B)$. The optimal edge-addition configuration is obtained by using Algorithm 3. The first step of calculating the network optimization cost is to obtain the load of each node in the network. Note that the nature of node load is exactly consistent with the betweenness centrality of the node. Betweenness centrality of a node refers to the proportion of the number of paths passing through the node in the total number of shortest paths. Intuitively, the betweenness centrality reflects the importance of the node as a "bridge." Therefore, the initial load on each node can be denoted by its betweenness centrality [27]. We can calculate the betweenness centrality of each node by "pajek" software after importing a directed network. There is a nonlinear relationship between the load of a node and its capacity [38, 39], so we can determine the node capacity by this nonlinear relation. The cost of a node can be measured by its node capacity in the network. We take the larger one of the two node capacities as the cost of the edge that connects these two nodes [40]. In this paper, we calculate the network costs of all optimal edge-addition configurations and then choose the one with the lowest network cost as the optimal edge-addition configuration.

The specific calculation process of network cost is given as follows:

Step 1. Node load can be measured by the betweenness centrality

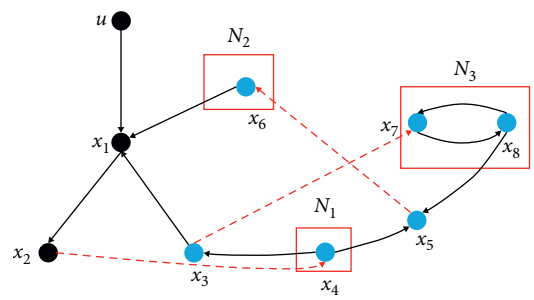


FIGURE 5: The new directed network resulting from the first configuration scheme $\widetilde{E}_1^* = \{(x_2, x_4), (x_5, x_6), (x_3, x_7)\}$.

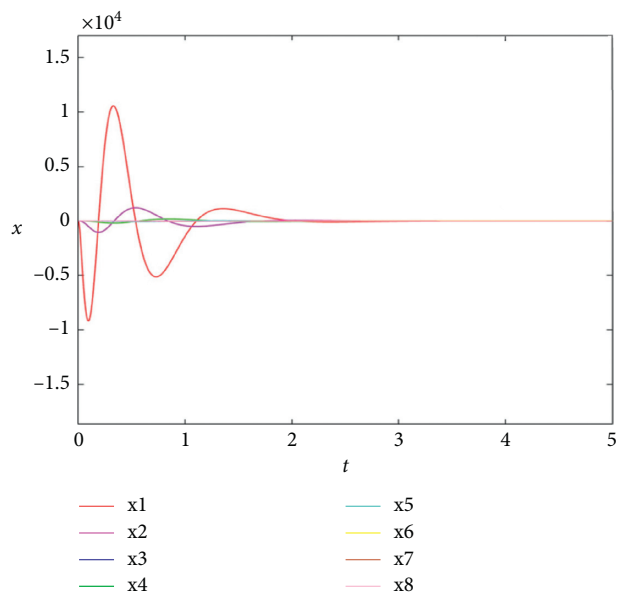


FIGURE 6: The different color curves in the figure represent the trend that the states of the 8 nodes in the new directed network change over time.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{10}$$

where $C_B(v)$ denotes the betweenness centrality of node v, $\sigma_{st}(v)$ denotes the number of the shortest directed paths $(s \longrightarrow t)$ that passes through node v, and σ_{st} means the number of the shortest directed paths from node s to node t.

Step 2. There is a nonlinear relationship between node load and node capacity described by

$$\mathrm{Cap}(v) = C_B(v) + \beta(C_B(v))^{\alpha}, \quad v = 1, 2, \ldots, n, \tag{11}$$

where $\mathrm{Cap}(v)$ is the capacity of node v, $\alpha > 0, \beta > 0$. Since there is a positive correlation between node load and capacity, set $\alpha = \beta = 1$. Thus, the node capacity is determined by

$$\mathrm{Cap}(v) = 2C_B(v). \tag{12}$$

Step 3. Use the index of node capacity to measure the node cost

$$\mathrm{Cost}(v) = \mathrm{Cap}(v) = 2C_B(v), \tag{13}$$

where $\mathrm{Cost}(v)$ denotes the cost of node v.

Step 4. Compare the capacities of two nodes of an edge, and take the larger one as the capacity of the edge (edge cost)

$$\mathrm{Cost}(l_{ij}) = \max\{\mathrm{Cap}(v_i), \mathrm{Cap}(v_j)\}, \tag{14}$$

where $\mathrm{Cost}(l_{ij})$ is the cost of edge l_{ij}.

Step 5. Calculate the network cost of each configuration according to Step 4

$$\mathrm{Cost}(\mathrm{Net}) = \sum \mathrm{Cost}(l_{ij}), \tag{15}$$

where $\mathrm{Cost}(\mathrm{Net})$ denotes the cost of the whole network.

5.2. Data Processing. In Figure 4(a), the initial directed network $G(A, B)$ is given. Get the optimal edge-addition configuration by Algorithm 3, $\widetilde{E}_1^* = \{(x_2, x_4), (x_5, x_6), (x_3,$

Table 1: The original data of betweenness centrality of each node in Figure 5.

Node	Val	Label
1	0.47619051	1
2	0.4761905	2
3	0.3095238	3
4	0.4761905	4
5	0.3333333	5
6	0.3333333	6
7	0.1904762	7
8	0.1904762	8

We calculate the betweenness centrality value of eight nodes in Figure 5 by pajek software.

Table 2: The data of node load, node capacity, edge cost, and network cost for the first configuration scheme.

Edge	Node load	Node capacity	Edge cost
(1, 2)	(0.48, 0.48)	(0.96, 0.96)	0.96
(3, 1)	(0.31, 0.48)	(0.62, 0.96)	0.96
(2, 4)	(0.48, 0.48)	(0.96, 0.96)	0.96
(4, 3)	(0.48, 0.31)	(0.96, 0.62)	0.96
(4, 5)	(0.48, 0.33)	(0.96, 0.66)	0.96
(6, 1)	(0.33, 0.48)	(0.66,0 .96)	0.96
(5, 6)	(0.33, 0.33)	(0.66, 0.66)	0.66
(3, 7)	(0.31, 0.19)	(0.62, 0.38)	0.62
(7, 8)	(0.19, 0.19)	(0.38, 0.38)	0.38
(8, 7)	(0.19, 0.19)	(0.38, 0.38)	0.38
(8, 5)	(0.19, 0.33)	(0.38, 0.66)	0.66
	Network cost		8.46

$x_7)\}$ and $\widetilde{E}_2^* = \{(x_2, x_4), (x_5, x_7), (x_3, x_6)\}$ and $\widetilde{E}_3^* = \{(x_2, x_4), (x_5, x_6), (x_6, x_7)\}$ and $\widetilde{E}_4^* = \{(x_2, x_4), (x_6, x_7), (x_3, x_6)\}$.

The new directed network resulting from the first configuration scheme is shown in Figure 5. Figure 6 shows the curve of the state of each node over time.

We import this new directed network $G(A + \widetilde{A}, B)$ into pajek software to calculate the betweenness centrality of each node. The original data of betweenness centrality of each node are shown in Table 1. In Table 2, we collate the data of node load, node capacity, edge cost, and network cost according to each step described in Section 5.1. Then, we get the network cost of the first configuration scheme.

The new directed network resulting from the second configuration scheme is shown in Figure 7. The original data of betweenness centrality of each node are shown in Table 3. Similarly, we can obtain the data of node load, node capacity, edge cost, and network cost, as shown in Table 4.

The new directed network resulting from the third configuration scheme is shown in Figure 8. The original data of betweenness centrality of each node are shown in Table 5. Similarly, we can obtain the data of node load, node capacity, edge cost, and network cost, as shown in Table 6.

The new directed network resulting from the fourth configuration scheme is shown in Figure 9. The original data of betweenness centrality of each node are shown in Table 7. Furthermore, we can obtain the data of node load, node capacity, edge cost, and network cost, as shown in Table 8.

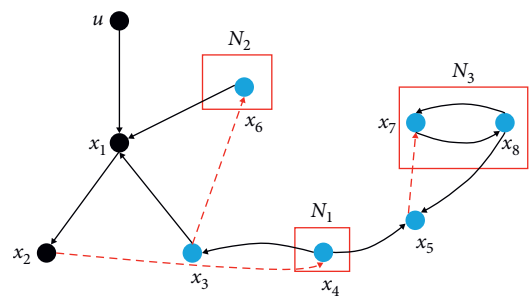


Figure 7: The new directed network resulting from the second configuration scheme $\widetilde{E}_2^* = \{(x_2, x_4), (x_5, x_7), (x_3, x_6)\}$.

Table 3: The original data of betweenness centrality of each node in Figure 7.

Node	Val	Label
1	0.2857143	1
2	0.3571429	2
3	0.1428571	3
4	0.4285714	4
5	0.2380952	5
6	0.0000000	6
7	0.1428571	7
8	0.0238095	8

We calculate the betweenness centrality value of eight nodes in Figure 7 by pajek software

Table 4: The data of node load, node capacity, edge cost, and network cost for the second configuration scheme.

Edge	Node load	Node capacity	Edge cost
(1, 2)	(0.29, 0.36)	(0.58, 0.72)	0.72
(3, 1)	(0.14, 0.29)	(0.28, 0.58)	0.58
(2, 4)	(0.36, 0.43)	(0.72, 0.86)	0.86
(4, 3)	(0.43, 0.14)	(0.86, 0.28)	0.86
(4, 5)	(0.43, 0.24)	(0.86, 0.48)	0.86
(6, 1)	(0.00, 0.29)	(0.00, 0.58)	0.58
(5, 7)	(0.24, 0.14)	(0.48, 0.24)	0.48
(3, 6)	(0.14, 0.00)	(0.28, 0.00)	0.28
(7, 8)	(0.14, 0.02)	(0.28, 0.04)	0.28
(8, 7)	(0.02, 0.14)	(0.04, 0.28)	0.28
(8, 5)	(0.02, 0.24)	(0.04, 0.48)	0.48
	Network cost		6.26

Comparing the network costs of the above four configuration schemes, we choose the fourth scheme as the optimal edge-addition configuration so as to get the solution of Problem 3.

5.3. Illustrative Example. In [23], a directed network as shown in Figure 10 is considered. The authors proposed 14 edge-addition configurations, i.e., $\widetilde{E}^* = \{(x_2, x_{10}), (x_9, x_5), (x_i, x_j)\}$, $i \in \{1, \ldots, 6, 10\}$, $j \in \{7, 8\}$. However, they did not tell us which one is the optimal edge-addition configuration with the lowest cost. Using the results of our work, the cost of each optimization scheme can be calculated, and finally a scheme $\widetilde{E}_8^* = \{(x_2, x_{10}), (x_9, x_5), (x_4, x_8)\}$ with the lowest

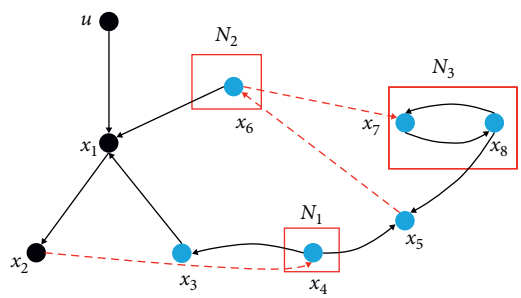


Figure 8: The new directed network resulting from the third configuration scheme $\widetilde{E}_3^* = \{(x_2, x_4), (x_5, x_6), (x_6, x_7)\}$.

Table 5: The original data of betweenness centrality of each node in Figure 8.

Node	Val	Label
1	0.4523810	1
2	0.4523810	2
3	0.0714286	3
4	0.4523810	4
5	0.5238095	5
6	0.5238095	6
7	0.1666667	7
8	0.1666667	8

We calculate the betweenness centrality value of eight nodes in Figure 8 by pajek software.

Table 6: The data of node load, node capacity, edge cost, and network cost for the third configuration scheme.

Edge	Node load	Node capacity	Edge cost
(1, 2)	(0.45, 0.45)	(0.90, 0.90)	0.90
(3, 1)	(0.07, 0.45)	(0.14, 0.90)	0.90
(2, 4)	(0.45, 0.45)	(0.90, 0.90)	0.90
(4, 3)	(0.45, 0.07)	(0.90, 0.14)	0.90
(4, 5)	(0.45, 0.52)	(0.90, 1.04)	1.04
(6, 1)	(0.52, 0.45)	(1.04, 0.90)	1.04
(5, 6)	(0.52, 0.52)	(1.04, 1.04)	1.04
(6, 7)	(0.52, 0.17)	(1.04, 0.34)	1.04
(7, 8)	(0.17, 0.17)	(0.34, 0.34)	0.34
(8, 7)	(0.17, 0.17)	(0.34, 0.34)	0.34
(8, 5)	(0.17, 0.52)	(0.34, 1.04)	1.04
	Network cost		9.48

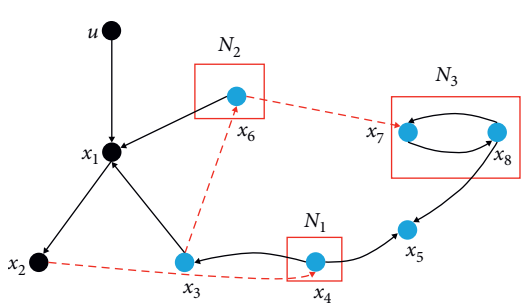


Figure 9: The new directed network resulting from the fourth configuration scheme $\widetilde{E}_4^* = \{(x_2, x_4), (x_6, x_7), (x_3, x_6)\}$.

Table 7: The original data of betweenness centrality of each node in Figure 9.

Node	Val	Label
1	0.1547619	1
2	0.2261905	2
3	0.2857143	3
4	0.2976190	4
5	0.0000000	5
6	0.2023810	6
7	0.1547619	7
8	0.0595238	8

We calculate the betweenness centrality value of eight nodes in Figure 9 by pajek software.

Table 8: The data of node load, node capacity, edge cost, and network cost for the fourth configuration scheme.

Edge	Node load	Node capacity	Edge cost
(1, 2)	(0.15, 0.23)	(0.30, 0.46)	0.46
(3, 1)	(0.29, 0.15)	(0.58, 0.30)	0.58
(2, 4)	(0.23, 0.30)	(0.46, 0.60)	0.60
(4, 3)	(0.30, 0.29)	(0.60, 0.58)	0.60
(4, 5)	(0.30, 0.00)	(0.60, 0.00)	0.60
(6, 1)	(0.20, 0.15)	(0.40, 0.30)	0.40
(3, 6)	(0.29, 0.20)	(0.58, 0.40)	0.58
(6, 7)	(0.20, 0.15)	(0.40, 0.30)	0.40
(7, 8)	(0.15, 0.06)	(0.30, 0.12)	0.30
(8, 7)	(0.06, 0.15)	(0.12, 0.30)	0.30
(8, 5)	(0.06, 0.00)	(0.12, 0.00)	0.12
	Network cost		4.94

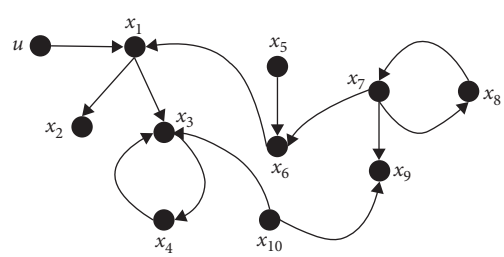


Figure 10: A directed network.

cost can be selected to ensure the structural controllability of the network.

6. Conclusions

In this paper, we have solved the problem of how to optimize the network topology to ensure structural controllability. Given a structurally uncontrollable directed network, Algorithm 3 presents all possible edge-addition configurations. After determining the optimal edge-addition configuration, a network cost index is given to choose the lowest cost configuration.

In future, we can combine these two strategies of adding edges and adding external input signals to ensure the network controllability and choose the scheme with the highest benefit by comparing the costs of several strategies. In

addition, we can extend a single directed network to the topology design of a multiplex network [29, 41] so as to ensure the structural controllability of the multiplex network.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grant no. 61973064, the Natural Science Foundation of Hebei Province of China under grant no. F2019501126, the Natural Science Foundation of Liaoning Province of China under grant no. 2020-KF-11-03, and the Fundamental Research Funds for the Central Universities under grant no. N182304013.

References

[1] R. E. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial and Applied Mathematics Series A Control*, vol. 1, no. 2, pp. 152–192, 1963.

[2] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.

[3] Y.-Y. Liu and A.-L. Barabási, "Control principles of complex systems," *Reviews of Modern Physics*, vol. 88, Article ID 35006, 2016.

[4] N. Cai and Y.-S. Zhong, "Formation controllability of high-order linear time-invariant swarm systems," *IET Control Theory & Applications*, vol. 4, no. 4, pp. 646–654, 2010.

[5] C. Commault, "Structural controllability of networks with dynamical structured nodes," *IEEE Transactions on Automatic Control*, vol. 65, no. 6, pp. 2736–2742, 2020.

[6] Y. Y. Yang, "Research progress in enhancing the controllability of complex networks," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 5759264, 8 pages, 2020.

[7] B. Liu, H. Su, L. Wu, and X. Shen, "Observability of leader-based discrete-time multi-agent systems over signed networks," *IEEE Transactions on Network Science and Engineering*, vol. 1, 2020.

[8] W. Wang, F. Chen, L. Xiang, and G. Chen, "A distributed algorithm for tracking general functions of multiple signals not-necessarily having steady states," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 1, 2020.

[9] I. Rajapakse, M. Groudine, and M. Mesbahi, "Dynamics and control of state-dependent networks for probing genomic organization," *Proceedings of the National Academy of Sciences*, vol. 108, no. 42, pp. 17257–17262, 2011.

[10] B. Liu, X. Shen, L. Wu, and H. Su, "Observability of leader-based discrete-time multi-agent systems with switching topology," *IET Control Theory & Applications*, vol. 14, no. 16, pp. 2462–2471, 2020.

[11] F. Chen, X. Z. Chen, L. Y. Xiang, and W. Ren, "Distributed economic dispatch via a predictive scheme: heterogeneous delays and privacy preservation," *Automatica*, vol. 123, Article ID 109356, 2020.

[12] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[13] S. Liu, Z. J. Ji, and H. Z. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, Article ID 7685460, 18 pages, 2020.

[14] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, vol. 51, no. 1, pp. 64–76, 2021.

[15] F. Chen and W. Ren, "Sign projected gradient flow: a continuous-time approach to convex optimization with linear equality constraints," *Automatica*, vol. 120, Article ID 109156, 2020.

[16] L. Xiang, F. Chen, W. Ren, and G. Chen, "Advances in network controllability," *IEEE Circuits and Systems Magazine*, vol. 19, no. 2, pp. 8–32, 2019.

[17] L. Xiang, P. Wang, F. Chen, and G. Chen, "Controllability of directed networked MIMO systems with heterogeneous dynamics," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 2, pp. 807–817, 2020.

[18] Z. Z. Yuan, C. Zhao, Z. R. Di, W.-X. Wang, and Y.-C. Lai, "Exact controllability of complex networks," *Nature Communications*, vol. 4, p. 2447, 2013.

[19] T. Nepusz and T. Vicsek, "Controlling edge dynamics in complex networks," *Nature Physics*, vol. 8, no. 7, pp. 568–573, 2012.

[20] S.-P. Pang, W.-X. Wang, F. Hao, and Y.-C. Lai, "Universal framework for edge controllability of complex networks," *Scientific Reports*, vol. 7, p. 4224, 2017.

[21] L. Xiang and G. Chen, "Minimal edge controllability of directed networks," *Advances in Complex Systems*, vol. 22, Article ID 1950017, 2019.

[22] W.-X. Wang, X. Ni, Y.-C. Lai, and C. Grebogi, "Optimizing controllability of complex networks by minimum structural perturbations," *Physical Review E*, vol. 85, no. 2, Article ID 26115, 2012.

[23] X. Chen, S. Pequito, G. J. Pappas, and V. M. Preciado, "Minimal edge addition for network controllability," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 1, pp. 312–323, 2019.

[24] Y. Zhang and T. Zhou, "Minimal structural perturbations for controllability of a networked system: complexities and approximations," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 7291, pp. 4191–4208, 2019.

[25] G. Yan, J. Ren, Y.-C. Lai, C.-H. Lai, and B. Li, "Controlling complex networks: how much energy is needed?" *Physical Review Letters*, vol. 108, Article ID 218703, 2002.

[26] F. Chen and J. Chen, "Minimum-energy distributed consensus control of multiagent systems: a network approximation approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 1144–1159, 2020.

[27] Z. H. Zhang, Y. F. Yin, and X. Zhang, "Optimization of robustness of interdependent network controllability by redundant design," *PLoS One*, vol. 13, no. 2, Article ID e0192874, 2018.

[28] Y. Lou, L. Wang, and G. Chen, "Enhancing controllability robustness of q-snapback networks through redirecting edges," *Research*, vol. 11, Article ID 7857534, 2019.

[29] M. Posfai, J. Gao, and S. P. Cornelius, "Controllability of multiplex, multitime-scale networks," *Physical Review E*, vol. 94, no. 3, Article ID 32316, 2016.

[30] L. Z. Wang, Y. Z. Chen, W.-X. Wang, and Y.-C. Lai, "Physical controllability of complex networks," *Scientific Reports*, vol. 7, Article ID 40198, 2017.

[31] C.-T. Lin, "Structural controllability," *IEEE Transactions on Automatic Control*, vol. 19, no. 3, pp. 201–208, 1974.

[32] R. Shields and J. Pearson, "Structural controllability of multiinput linear systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 2, pp. 203–212, 1976.

[33] K. Glover and L. Silverman, "Characterization of structural controllability," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 534–537, 1976.

[34] J.-M. Dion, C. Commault, and J. van der Woude, "Generic properties and control of linear structured systems: a survey," *Automatica*, vol. 39, no. 7, pp. 1125–1144, 2003.

[35] C. Sueur and G. Dauphin-Tanguy, "Bond-graph approach for structural analysis of MIMO linear systems," *Journal of the Franklin Institute*, vol. 328, no. 1, pp. 55–70, 1991.

[36] K. Murota and S. Poljak, "Note on a graph-theoretic criterion for structural output controllability," *IEEE Transactions on Automatic Control*, vol. 35, no. 8, pp. 939–942, 1990.

[37] C. Commault, J. M. Dion, and J. W. V. D. Woude, "Characterization of generic properties of linear structured systems for efficient computations," *Kybernetika*, vol. 38, no. 5, pp. 503–520, 2002.

[38] D.-H. Kim and A. E. Motter, "Resource allocation pattern in infrastructure networks," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, no. 22, Article ID 224019, 2008.

[39] B.-L. Dou, X.-G. Wang, and S.-Y. Zhang, "Robustness of networks against cascading failures," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 11, pp. 2310–2317, 2010.

[40] L. J. Liu, Y. F. Yin, Z. H. Zhang, and Y. K. Malaiya, "Redundant design in interdependent networks," *PLoS One*, vol. 11, no. 10, Article ID e0164777, 2016.

[41] Z. Wang, C. Xia, Z. Chen, and G. Chen, "Epidemic propagation with positive and negative preventive information in multiplex networks," *IEEE Transactions on Cybernetics*, vol. 1, 2020.

Hindawi
Complexity
Volume 2020, Article ID 8894566, 10 pages
https://doi.org/10.1155/2020/8894566

Research Article

Multitask Learning with Local Attention for Tibetan Speech Recognition

Hui Wang, Fei Gao, Yue Zhao, Li Yang, Jianjian Yue, and Huilin Ma

School of Information Engineering, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Fei Gao; 18301390@muc.edu.cn

Received 22 September 2020; Revised 12 November 2020; Accepted 26 November 2020; Published 18 December 2020

Academic Editor: Ning Cai

In this paper, we propose to incorporate the local attention in WaveNet-CTC to improve the performance of Tibetan speech recognition in multitask learning. With an increase in task number, such as simultaneous Tibetan speech content recognition, dialect identification, and speaker recognition, the accuracy rate of a single WaveNet-CTC decreases on speech recognition. Inspired by the attention mechanism, we introduce the local attention to automatically tune the weights of feature frames in a window and pay different attention on context information for multitask learning. The experimental results show that our method improves the accuracies of speech recognition for all Tibetan dialects in three-task learning, compared with the baseline model. Furthermore, our method significantly improves the accuracy for low-resource dialect by 5.11% against the specific-dialect model.

1. Introduction

Multitask learning has been applied successfully for speech recognition to improve the generalization performance of the model on the original task by sharing the information between related tasks [1–9]. Chen and Mak [6] used the multitask framework to conduct joint training of multiple low-resource languages, exploring the universal phoneme set as a secondary task to improve the effect of the phoneme model of each language. Krishna et al. [7] proposed a hierarchical multitask model, and the performance differences between high-resource language and low-resource language were compared. Li et al. [8] and Toshniwal et al. [9] introduced additional information of language ID to improve the performance of end-to-end multidialect speech recognition systems.

Tibetan is one of minority languages in China. It has three major dialects in China, i.e., Ü-Tsang, Kham, and Amdo. There are also several local subdialects in each dialect. Tibetan dialects pronounce very differently, but the written characters are unified across dialects. In our previous work [10], Tibetan multidialect multitask speech recognition was conducted based on the WaveNet-CTC, which performed simultaneous Tibetan multidialect speech content recognition, dialect identification, and speaker recognition in a single model. WaveNet is a deep generative model with very large receptive fields, and it can model the long-term dependency of speech data. It is very effective to learn the shared representation from speech data of different tasks. Thus, WaveNet-CTC was trained on three Tibetan dialect data sets and learned the shared representations and model parameters for speech recognition, speaker identification, and dialect recognition. Since the Lhasa of Ü-Tsang dialect is a standard Tibetan speech, there are more corpora available for training than Changdu-Kham and Amdo pastoral dialect. Although two-task WaveNet-CTC improved the performance on speech recognition for Lhasa of Ü-Tsang dialect and Changdu-Kham dialect, the three-task model did not improve performance for all dialects. With an increase in task number, the speech recognition performance degraded.

To obtain a better performance, attention mechanism is introduced into WaveNet-CTC for multitask learning in this paper. Attention mechanism can learn to set larger weight to more relevant frames at each time step. Considering the computation complexity, we conduct a local attention using a sliding window on the whole of speech feature frames to create the weighted context vectors for different recognition tasks. Moreover, we explore to place a local attention at the

different positions within WaveNet, i.e., in the input layer and high layer, respectively.

The contribution of this work is three-fold. For one, we propose the WaveNet-CTC with local attention to perform multitask learning for Tibetan speech recognition, which can automatically capture the context information among different tasks. This model improves the performance of the Tibetan multidialect speech recognition task. Moreover, we compared the performance of local attention inserted at different positions in the multitask model. The attention component embedded in the high layer of WaveNet obtains better performance than the one in the input layer of WaveNet for speech recognition. Finally, we conduct a sliding window on the speech frames for efficiently computing the local attention.

The rest of this paper is organized as follows: Section 2 introduces the related work. Section 3 presents our method and gives the description of the baseline model, local attention mechanism, and the WaveNet-CTC with local attention. In Section 4, the Tibetan multidialect data set and experiments are explained in detail. Section 5 describes our conclusions.

2. Related Work

Connectionist temporal classification (CTC) for end-to-end has its advantage of training simplicity and is one of the most popular methods used in speech recognition. Das et al. [11] directly incorporated attention modelling within the CTC framework to address high word error rates (WERs) for a character-based end-to-end model. But, in Tibetan speech recognition scenarios, the Tibetan character is a two-dimensional planar character, which is written in Tibetan letters from left to right, besides there is a vertical superposition in syllables, so a word-based CTC is more suitable for the end-to-end model. In our work, we try to introduce attention mechanism in WaveNet as an encoder for the CTC-based end-to-end model. The attention is used in WaveNet to capture the context information among different tasks for distinguishing dialect content, dialect identity, and speakers.

In multitask settings, there are some recent works focusing on incorporating attention mechanism in multitask training. Zhang et al. [12] proposed an attention mechanism for the hybrid acoustic modelling framework based on LSTM, which weighted different speech frames in the input layer and automatically tuned its attention to the spliced context input. The experimental results showed that attention mechanism improved the ability to model speech. Liu et al. [13] incorporated the attention mechanism in multitask learning for computer vision tasks, in which the multitask attention network consisted of a shared network and task-specific soft-attention modules to learn the task-specific features from the global pool, whilst simultaneously allowing for features to be shared across different tasks. Zhang et al. [14] proposed an attention layer on the top of the layers for each task in the end-to-end multitask framework to relieve the overfitting problem in speech emotion recognition. Different from the works of Liu et al. and Zhang et al. [13, 14], which distributed many attention modules in the network, our method merely uses one sliding attention window in the multitask network and has its advantage of training simplicity.

3. Methods

3.1. Baseline Model. We take the Tibetan multitask learning model in our previous work [10] as the baseline model as shown in Figure 1, which was initially proposed for Chinese and Korean speech recognition from the work of Xu [15] and Kim and Park [16]. The work [10] integrates WaveNet [17] with CTC loss [18] to realize Tibetan multidialect end-to-end speech recognition.

WaveNet contains the stacks of dilated causal convolutional layers as shown in Figure 2. In the baseline model, the WaveNet network consists of 15 layers, which are grouped into 3 dilated residual blocks of 5 layers. In every stack, the dilation rate increases by a factor of 2 in every layer. The filter length of causal dilated convolutions is 2. According to equations (1) and (2), the respective field of WaveNet is 46:

$$\text{Receptive\_field}_{\text{block}} = \sum_{i=1}^{n} \left(\text{Filter}_{\text{length}} - 1\right) \times \text{Dilation}_{\text{rate}_i} + 1. \tag{1}$$

$$\text{Receptive field}_{\text{stacks}} = S \times \text{Receptive field}_{\text{block}} - S + 1. \tag{2}$$

In equations (1) and (2), S refers to the number of stacks, $\text{Receptive\_field}_{\text{block}}$ refers to the receptive field of a stack of dilated CNN, $\text{Receptive field}_{\text{stacks}}$ refers to the receptive field of some stacks of dilated CNN, and $\text{Dilation}_{\text{rate}_i}$ refers to the dilation rate of the i-th layer in a block.

WaveNet also uses residual and parameterized skip connections [19] to speed up convergence and enable training of much deeper models. More details about WaveNet can be found in [17].

Connectionist temporal classification (CTC) is an algorithm that trains a deep neural network [20] for the end-to-end learning task. It can make the sequence label predictions at any point in the input sequence [18]. In the baseline model, since the Tibetan character is a two-dimensional planar character as shown in Figure 3, the CTC modeling unit for Tibetan speech recognition is Tibetan single syllable, otherwise a Tibetan letter sequence from left to right is unreadable.

3.2. Local Attention Mechanism. Since the effect of each speech feature frame is different for the target label output at current time, considering the computational complexity, we introduce the local attention [21] into WaveNet to create a weighted context vector for each time i. The local attention places a sliding window with the length $2n$ centered around the current speech feature frame on the input layer and before the softmax layer in WaveNet, respectively, and repeatedly produces a context vector C_i for the current input (or hidden) feature frame $x\,(h)_i$. The formula for C_i is shown

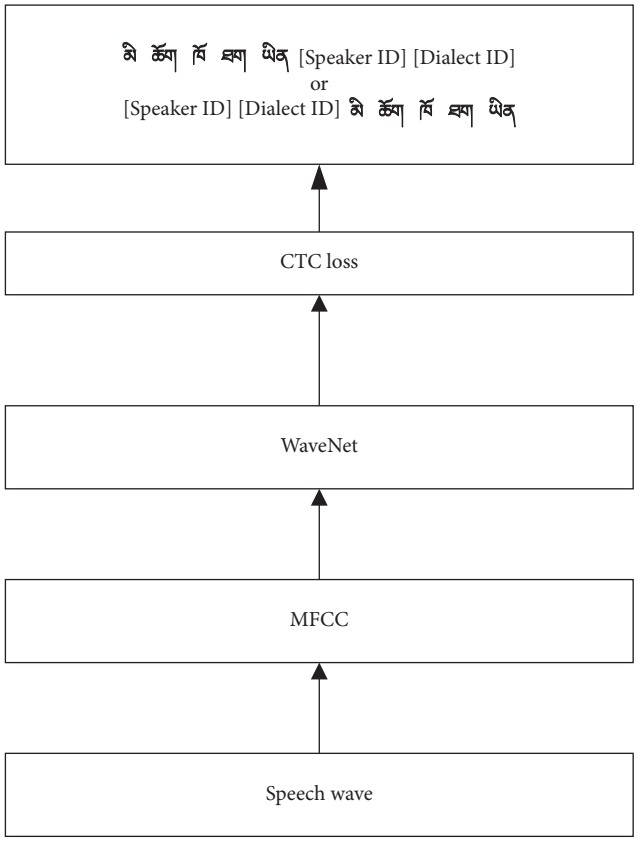


Figure 1: The baseline model.

in equation (3), and the schematic diagram is shown in Figure 4:

$$C_i = \sum_{j=i-n, j\neq i}^{i+n} \alpha_{i,j} \cdot x(h)_j, \tag{3}$$

where $\alpha_{i,j}$ is the attention weight, subject to $\alpha \geq 0$ and $\sum_j \alpha_{i,j} = 1$ through softmax normalization. The $\alpha_{i,j}$ calculation method is as follows:

$$\alpha_{i,j} = \frac{\exp\left(\text{Score}\left(x(h)_i, x(h)_j\right)\right)}{\sum_j \exp\left(\text{Score}\left(x(h)_i, x(h)_j\right)\right)}. \tag{4}$$

It captures the correlation of speech frame pair $(x(h)_i, x(h)_j, j \neq i)$. The attention operates on n frames before and after the current frame. Score (.) is an energy function, whose value is computed as equation (5) by the MLP which is jointly trained with all the other components in an end-to-end network. Those $x(h)_j, j \neq i$ that get larger scores would have more weights in context vector C_i.

$$\text{Score}\left(x_i, x_j\right) = v_a^T \tanh\left(W_a\left[x(h)_i; x(h)_j\right]\right). \tag{5}$$

Finally, $x(h)_i$ is concatenated with C_i as the extended feature frame and fed into the next layer of WaveNet as shown in Figures 5 and 6. The attention module is inserted in the input layer in Figure 5 referred as Attention-WaveNet-CTC. The attention module is embedded before the softmax layer in Figure 6 referred as WaveNet-Attention-CTC.

4. Experiments

4.1. Data. Our experimental data are from an open and free Tibetan multidialect speech data set TIBMD@MUC [10], in which the text corpus consists of two parts: one is 1396 spoken language sentences selected from the book "Tibetan Spoken Language" [22] written by La Bazelen and the other part contains 8,000 sentences from online news, electronic novels, and poetry of Tibetan on internet. All text corpora in TIBMD@MUC include a total of 3497 Tibetan syllables.

There are 40 recorders who are from Lhasa City in Tibet, Yushu City in Qinghai Province, Changdu City in Tibet, and Tibetan Qiang Autonomous Prefecture of Ngawa. They used different dialects to speak out the same text for 1396 spoken sentences, and other 8000 sentences are read loudly in Lhasa dialect. Speech data files are converted to 16K Hz sampling frequency, 16 bit quantization accuracy, and wav format.

Our experimental data for multitask speech recognition are shown in Table 1, which consists of 4.4 hours Lhasa-Ü-Tsang, 1.90 hours Changdu-Kham, and 3.28 hours Amdo pastoral dialect, and their corresponding texts contain 1205 syllables for training. We collect 0.49 hours Lhasa-Ü-Tsang, 0.19 hours Changdu-Kham, and 0.37 hours Amdo pastoral dialect, respectively, to test.

39 MFCC features of each observation frame are extracted from speech data using a 128 ms window with 96 ms overlaps.

The experiments are divided into two parts: two-task experiments and three-task experiments. Three dialect-specific models and a multi-dialect model without attention are trained on WaveNet-CTC.

In WaveNet, the number of hidden units in the gating layers is 128. The learning rate is 2×10^{-4}. The number of hidden units in the residual connection is 128.

4.2. Two-task Experiment. For two-task joint recognition, the performances of the dialect ID or speaker ID at the beginning and at the end of output sequence were evaluated, respectively. We set $n = 5$ frames before and after the current frame to calculate the attention coefficients for attention-based WaveNet-CTC, which are referred to as Attention (5)-WaveNet-CTC and WaveNet-Attention (5)-CTC, respectively, for the two architectures in Figures 5 and 6. Compared with the calculation of the attention coefficient of all frames, the calculation speed of local attention has been improved quickly, which is convenient for the training of models.

The speech recognition result is summarized in Table 2. The best model is the proposed WaveNet-Attention-CTC with the attention embedded before the softmax layer in WaveNet and dialect ID at the beginning of label sequence. It outperforms the dialect-specific model by 7.39% and 2.4%, respectively, for Lhasa-Ü-Tsang and Changdu-Kham and gets the SER close to the dialect-specific model for Amdo Pastoral, which has the highest ARSER (average relative syllable error rate) for three dialects. The model of dialectID-speech (D-S) in the framework of WaveNet-Attention-CTC is effective to improve multilinguistic speech content recognition. Speech content recognition is more sensitive to the

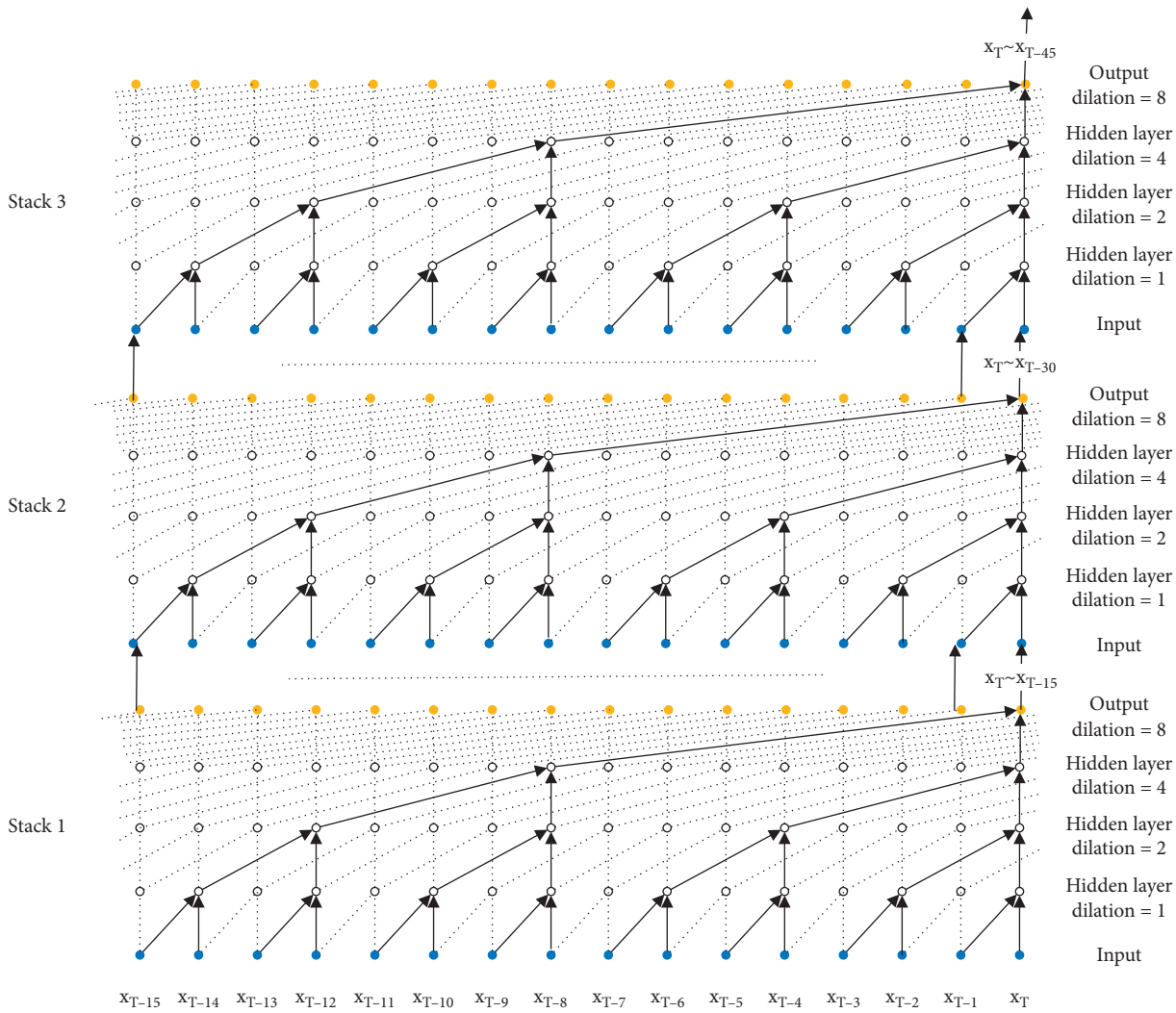


Figure 2: 3 stacks of 5 dilated causal convolutional layers with filter length 2.

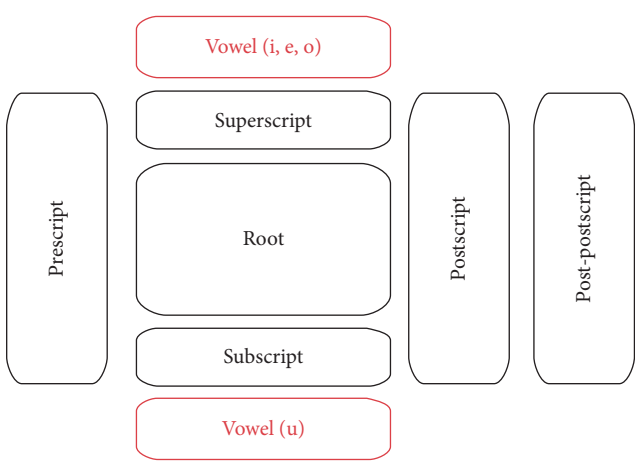


Figure 3: The structure of a Tibetan syllable.

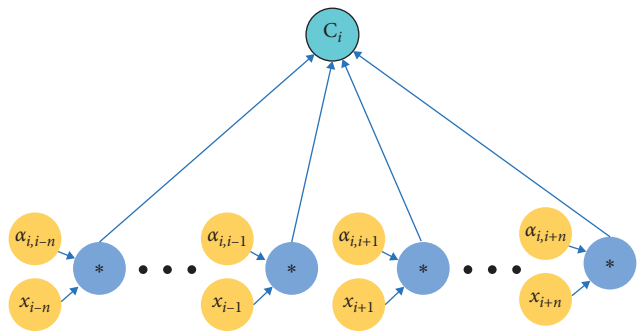


Figure 4: Local attention.

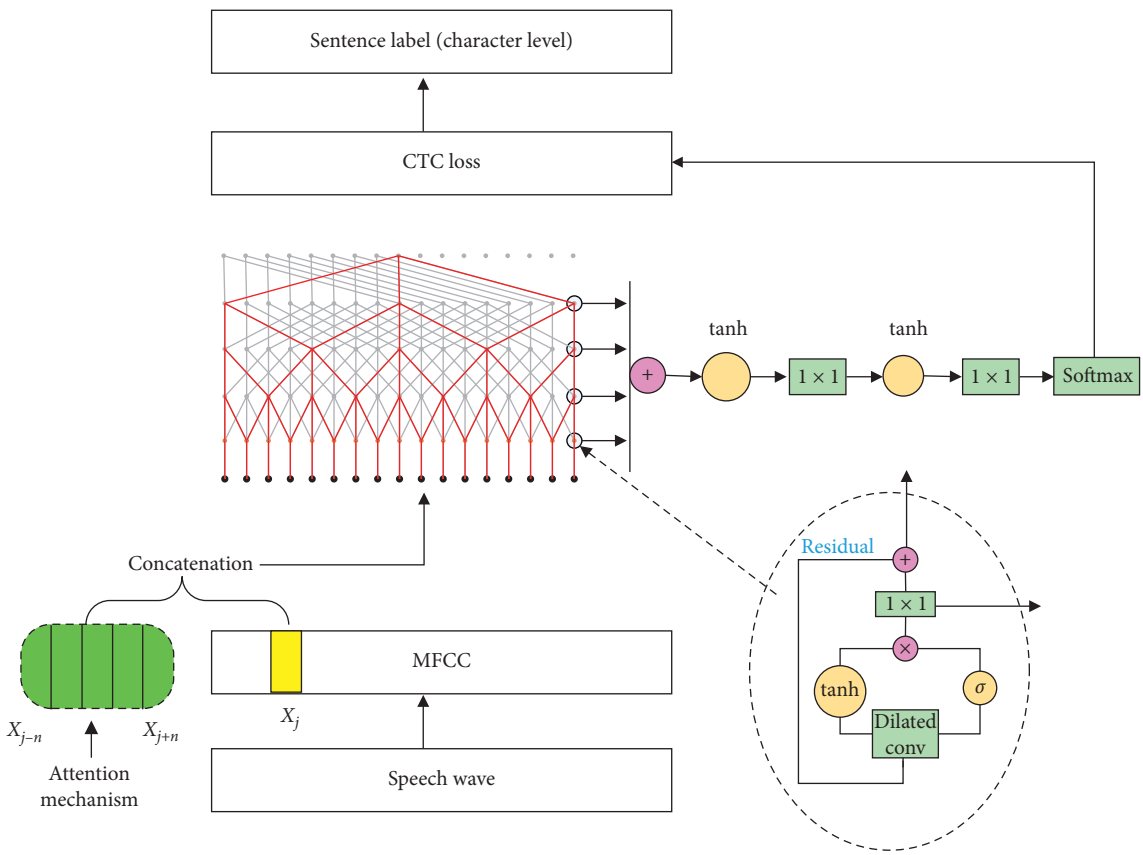


Figure 5: The architecture of attention-WaveNet-CTC.

recognition of dialect ID than speaker ID. The recognition of dialect ID helps to identify the speech content. However, the attention inserted before the input layer in WaveNet resulted in the worst recognition, which shows that raw speech feature cannot provide much information to distinguish the multitask.

For dialect ID recognition, in Table 3, we can see that the model with attention mechanism added before the softmax layer performs better than which is added in the input layer, and the dialect ID at the beginning is better than that at the end. From Table 2 and Table 3, it can be seen that the dialect ID recognition influences the speech content recognition.

We also test the speaker ID recognition accuracy for the two-task models. Results are listed in Table 4. It is worth noting that the Attention-WaveNet-CTC model performs poorly on both tasks of the speaker and speech content recognition. Especially in the speaker identification task, the recognition rate of the speakerID-speech model in all three dialects is very poor. Among the Attention-WaveNet-CTC models, it can be seen that the modelling ability of two models of the dialectID-speech and speakerID-speech model shows big gap, which means the Attention-WaveNet-CTC architecture cannot learn effectively the correlation among multiple frames of acoustic feature for multiple classification tasks. In contrast, the WaveNet-Attention-CTC model has a

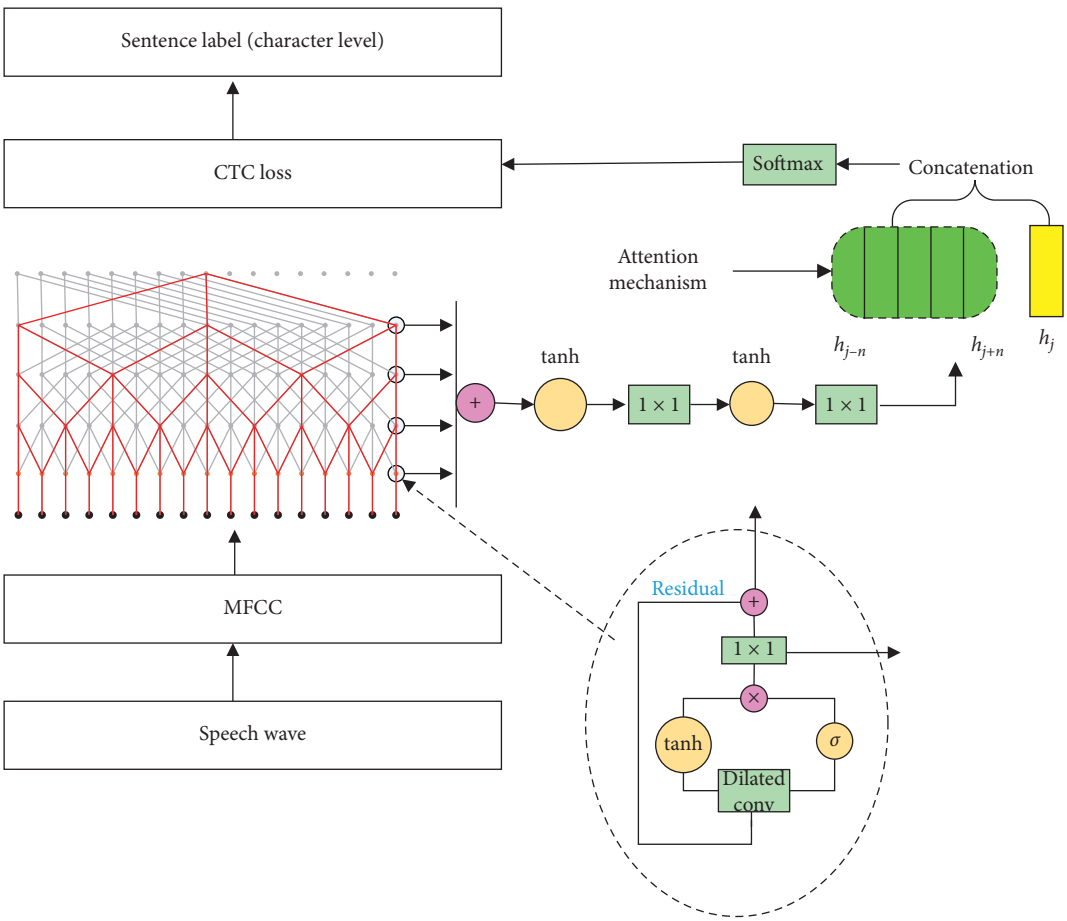


Figure 6: The architecture of WaveNet-attention-CTC.

Table 1: The experimental data statistics.

Dialect	Training data (hours)	Training utterances	Test data (hours)	Test utterances	Speaker
Lhasa-Ü-Tsang	4.40	6678	0.49	742	20
Changdu-Kham	1.90	3004	0.19	336	6
Amdo pastoral	3.28	4649	0.37	516	14
Total	9.58	14331	1.05	2110	40

much better performance on the two tasks. The attention embedded before the softmax layer can find the related and important frames to lead to high recognition accuracy.

4.3. Three-task Experiment. We compared the performances of two architectures, namely, Attention-WaveNet-CTC and WaveNet-Attention-CTC on three-task learning with the dialect-specific model and WaveNet-CTC, where we evaluated $n = 5$, $n = 7$, and $n = 10$, respectively, for the attention mechanism. The results are shown in Table 5.

We can see that the three-task models have worse performance compared with the two-task model, and WaveNet-Attention-CTC has lower SERs for Lhasa-Ü-Tsang and Amdo Pastoral against the dialect-specific model, but for Changdu-Kham, a relative low-resource Tibetan dialect, the model of dialectID-speech-speakerID (D-S-S2) based on the framework of WaveNet-Attention (10)-CTC achieved the highest recognition rate in all models, which outperforms the dialect-specific model by 5.11%. We analyzed the reason that maybe is the reduction of generalization error of the multitask model with the number of learning tasks increasing. It improves the recognition rate for small-data dialect, however not for big-data dialects. Since ASER reflects the generalization error of the model, D-S-S2 of WaveNet-Attention (10)-CTC has highest ASER in all models, which shows it has better generalization capacity. Meanwhile, WaveNet-Attention (10)-CTC achieved the better performance than WaveNet-Attention (5)-CTC and WaveNet-Attention (7)-CTC for speech content recognition as shown in Figure 7, where the syllable error rates declined with the number of n increasing for three dialects, and Changdu-Kham's SER has a quickest descent. We can conclude that attention mechanism needs a longer range to distinguish more tasks, and it pays more attention on the

Table 2: Syllable error rate (%) of two-task models on speech content recognition.

Architecture	Model	Lhasa-Ü-Tsang		Changdu-Kham		Amdo Pastoral		
		SER[1]	RSER[2]	SER	RSER	SER	RSER	ASER[3]
Dialect-specific model		28.83		62.56		**17.6**		
WaveNet-CTC		29.55	−0.72	62.83	−0.27	33.52	−15.92	−5.63
WaveNet-CTC with dialect ID or speaker ID (baseline model)	D-S[4]	32.84	−4.01	68.58	−6.02	33.00	−15.40	−8.48
	S-D[5]	26.80	2.03	64.03	−1.47	30.79	−13.09	−4.21
	S-S1[6]	27.21	1.62	64.17	−1.61	29.68	−12.08	−4.02
	S-S2[7]	28.13	0.7	62.43	0.13	28.04	-10.44	−3.20
Attention (5)-WaveNet-CTC	D-S	52.19	−23.36	65.24	−2.68	50.22	-32.62	−19.55
	S-D	55.16	−26.33	67.78	−5.22	55.23	-37.63	−23.06
	S-S1	77.42	−48.59	85.44	−22.88	82.08	-64.48	−45.32
	S-S2	83.32	−54.49	89.15	−26.94	81.47	-63.87	−48.43
WaveNet-Attention (5)-CTC	D-S	**21.44**	**7.39**	**60.16**	**2.40**	20.46	**−2.86**	**2.31**
	S-D	23.79	5.04	62.96	−0.4	24.15	−6.55	−0.64
	S-S1	34.86	−6.03	63.36	−0.8	40.10	−22.50	−9.78
	S-S2	34.83	−6.00	62.70	−0.14	37.63	−20.03	−8.72

[1]SER: syllable error rate, [2]RSER: relative syllable error rate, [3]ARSER: average relative syllable error rate, [4]D-S: the model trained using the transcription with dialect ID at the beginning of target label sequence, like "A ཐུང་གསར་ཆ," [5]S-D: the model trained using the transcription with dialect ID at the end of target label sequence, [6]S-S1: the model trained using the transcription with speaker ID at the beginning of target label sequence, and [7]S-S2: the model trained using the transcription with speaker ID at the end of target label sequence.

Table 3: Dialect ID recognition accuracy (%) of two-task models.

Architecture	Model	Lhasa-Ü-Tsang	Changdu-Kham	Amdo Pastoral
DialectID model		97.88	92.24	97.9
WaveNet-CTC with dialect ID	D-S	98.57	95.23	**99.6**
	S-D	99.01	97.61	99.41
Attention (5)-WaveNet-CTC	D-S	**100**	89.28	94.52
	S-D	0	0	0
WaveNet-Attention (5)-CTC	D-S	**100**	**98.8**	99.41
	S-D	**100**	94.04	98.06

Table 4: Speaker ID recognition accuracy (%) of two-task models.

Architecture	Model	Lhasa-Ü-Tsang	Changdu-Kham	Amdo Pastoral
SpeakerID model		67.75	93.13	95.31
WaveNet-CTC with speaker ID	S-S1	68.32	92.85	**97.48**
	S-S2	**71.15**	95.23	96.12
Attention (5)-WaveNet-CTC	S-S1	0	0	0
	S-S2	60.64	77.38	85.85
WaveNet-Attention (5)-CTC	S-S1	70.35	92.85	**97.48**
	S-S2	69.40	**100**	96.70

low-resource task. It is also observed that WaveNet-Attention (5)-CTC has better performance than Attention (5)-WaveNet-CTC, which demonstrates again that the attention mechanism placed in the high layer can find the related and important information which leads to more accurate speech recognition than when it is put in the input layer.

From Tables 6 and 7, we can observe that models with attention have worse performance than the ones without attention for dialect ID recognition and speaker ID recognition, and longer attention achieved the worse recognition for the language with large data. It also shows that in the case of more tasks, the attention mechanism tends towards the low-resource task, such as speech content recognition.

In summary, combining the results of the above experiments, whether two task or three task, the multitask model can make a significant improvement on the performance of the low-resource task by incorporating the attention mechanism, especially when the attention is applied to the high-level abstract features. The attention-based multitask model can achieve the improvements on speech recognition for all dialects compared with the baseline model. With an increase in the task number, the multitask model needs to increase the range for attention to distinguish multiple dialects.

Table 5: Syllable error rate (%) of three-task models on speech content recognition.

Architecture	Model	Lhasa-Ü-Tsang		Changdu-Kham		Amdo Pastoral		
		SER	RSER	SER	RSER	SER	RSER	ASER
Dialect-specific model		**28.83**		62.56		**17.60**		
WaveNet-CTC with dialect ID and speaker ID (baseline model)	S-D-S	30.64	−1.81	64.17	−1.61	34.06	−16.46	−6.62
	D-S-S1	39.64	−10.81	65.10	−2.54	45.15	−27.55	−13.63
	D-S-S2	33.43	−4.60	64.83	−2.27	37.56	−19.96	−8.94
Attention (5)-WaveNet-CTC	S-D-S	48.69	−19.86	68.31	−5.75	63.22	−45.62	−23.74
	D-S-S1	52.57	−23.74	69.38	−6.82	71.42	−53.82	−28.13
	D-S-S2	49.10	−20.27	79.41	−16.85	61.09	−43.49	−26.87
WaveNet-Attention (5)-CTC	S-D-S	30.75	−1.92	69.51	−6.95	34.21	−16.61	−8.49
	D-S-S1	33.17	−4.34	69.51	-6.95	38.49	−20.89	−10.73
	D-S-S2	31.16	−2.33	69.25	−6.69	34.14	−16.54	−8.52
WaveNet-Attention (7)-CTC	S-D-S	30.39	−1.56	70.05	−7.49	32.7	−15.1	−8.05
	D-S-S1	35.28	−6.45	68.12	−5.56	38.03	−20.73	−10.81
	D-S-S2	32.58	−3.75	62.74	−0.18	37.16	−19.56	−7.83
WaveNet-Attention (10)-CTC	S-D-S	30.25	**−1.42**	69.25	−6.69	32.01	**−14.41**	−7.51
	D-S-S1	34.06	−5.23	70.05	−7.49	40.10	−22.50	−11.74
	D-S-S2	31.85	−3.02	**57.45**	**5.11**	33.65	−16.05	**−4.65**

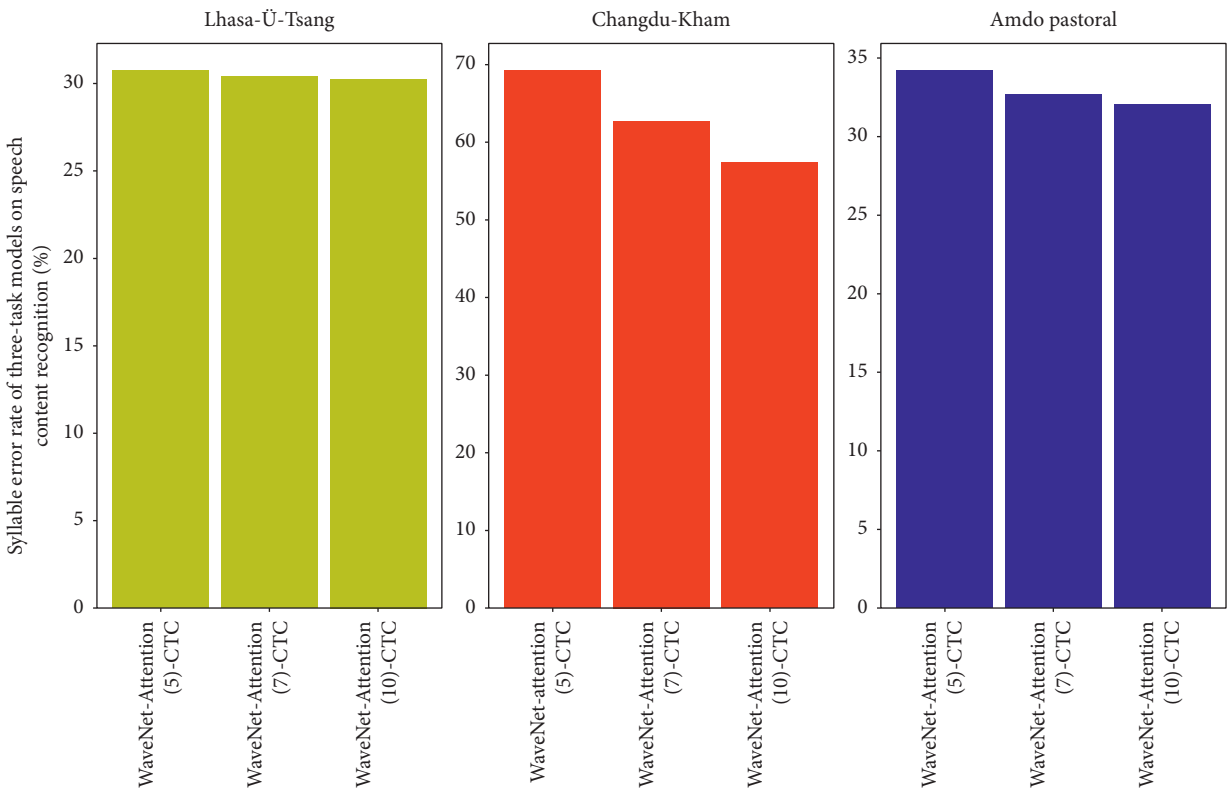


Figure 7: Syllable error rate of WaveNet-Attention-CTC for different lengths of the attention window.

Table 6: Dialect ID recognition accuracy (%) of three-task models.

Architecture	Model	Lhasa-Ü-Tsang	Changdu-Kham	Amdo Pastoral
DialectID model		97.88	92.24	97.9
WaveNet-CTC with dialect ID and speaker ID	D-S-S1	98.01	**98.8**	99.41
	D-S-S2	99.73	96.42	**99.61**
	S-D-S	99.25	95.23	99.03
Attention (5)-WaveNet-CTC	S-D-S	**100**	76.19	91.27
	D-S-S1	**100**	90.47	94.18
	D-S-S2	**100**	82.14	93.02
WaveNet-Attention (5)-CTC	S-D-S	**100**	89.28	93.79
	D-S-S1	**100**	85.71	93.79
	D-S-S2	**100**	95.23	94.18
WaveNet-Attention (7)-CTC	S-D-S	0	85.71	91.66
	D-S-S1	0	89.98	93.88
	D-S-S2	0	89.28	95.34
WaveNet-Attention (10)-CTC	S-D-S	0	85.71	95.54
	D-S-S1	0	94.04	93.99
	D-S-S2	0	0	0

Table 7: Speaker ID recognition accuracy (%) of three-task models.

Architecture	Model	Lhasa-Ü-Tsang	Changdu-Kham	Amdo pastoral
SpeakerID model		67.75	93.13	95.31
WaveNet-CTC with dialect ID and speaker ID	S-D-S	**72.91**	**98.8**	96.12
	D-S-S1	70.21	95.23	93.6
	D-S-S2	70.35	96.42	96.89
Attention (5)-WaveNet-CTC	S-D-S	61.08	83.33	89.53
	D-S-S1	62.12	83.33	87.01
	D-S-S2	61.99	84.52	90.11
WaveNet-Attention (5)-CTC	S-D-S	61.99	85.71	92.05
	D-S-S1	62.53	82.14	91.08
	D-S-S2	61.18	89.28	92.44
WaveNet-Attention (7)-CTC	S-D-S	60.91	85.71	91.66
	D-S-S1	62.04	84.31	92.01
	D-S-S2	58.49	86.90	90.69
WaveNet-Attention (10)-CTC	S-D-S	58.49	84.52	92.05
	D-S-S1	59.43	83.33	91.27
	D-S-S2	63.47	92.85	**97.86**

5. Conclusions

This paper proposes a multitask learning mechanism with local attention based on WaveNet to improve the performance for low-resource language. We integrate Tibetan multidialect speech recognition, speaker ID recognition, and dialect identification into a unified neural network and compare the attention effects on the different places in architectures. The experimental results show that our method is effective for Tibetan multitask processing scenarios. The WaveNet-CTC model with attention added into the high layer obtains the best performance for unbalance-resource multitask processing. In the future works, we will evaluate the proposed method on larger Tibetan data set or on different languages.

Data Availability

The data used to support the findings of this study are available from the corresponding author (1009540871@qq.com) upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Hui Wang and Yue Zhao contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation under grant no. 61976236.

References

[1] Z. Tang, L. Li, and D. Wang, "Multi-task recurrent model for speech and speaker recognition," in *Proceedings of the 2016 Asia-Pacific signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–4, Jeju, South Korea, December 2016.

[2] O. Siohan and D. Rybach, "Multitask learning and system combination for automatic speech recognition," in *Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 589–595, Scottsdale, AZ, USA, December 2015.

[3] Y. Qian, M. Yin, Y. You, and K. Yu, "Multi-task joint-learning of deep neural networks for robust speech recognition," in *Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 310–316, Scottsdale, AZ, USA, December 2015.

[4] A. Thanda and S. M. Venkatesan, "Multi-task learning of deep neural networks for audio visual automatic speech recognition," 2020, http://arxiv.org/abs/1701.02477.

[5] X. Yang, K. Audhkhasi, A. Rosenberg, S. Thomas, B. Ramabhadran, and M. Hasegawa-Johnson, "Joint modeling of accents and acoustics for multi-accent speech recognition," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, April 2018.

[6] D. Chen and B. K.-W. Mak, "Multitask learning of deep neural networks for low-resource speech recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 7, pp. 1172–1183, 2015.

[7] K. Krishna, S. Toshniwal, and K. Livescu, "Hierarchical multitask learning for ctc-based speech recognition," 2020, http://arxiv.org/abs/1807.06234.

[8] B. Li, T. N. Sainath, Z. Chen et al., "Multi-dialect speech recognition with a single sequence-to-sequence model," 2017, http://arxiv.org/abs/1712.01541.

[9] S. Toshniwal, T. N. Sainath, B. Li et al., "Multilingual speech recognition with a single end-to-end model," 2018, http://arxiv.org/abs/1711.01694.

[10] Y. Zhao, J. Yue, X. Xu, L. Wu, and X. Li, "End-to-end-based Tibetan multitask speech recognition," *IEEE Access*, vol. 7, pp. 162519–162529, 2019.

[11] A. Das, J. Li, R. Zhao, and Y. F. Gong, "Advancing connectionist temporal classification with attention," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, April 2018.

[12] Y. Zhang, P. Y. Zhang, and H. Y. Yan, "Long short-term memory with attention and multitask learning for distant speech recognition," *Journal of Tsinghua University (Science and Technology)*, vol. 58, no. 3, p. 249, 2018, in Chinese.

[13] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June2019.

[14] Z. Zhang, B. Wu, and B. Schuller, "Attention-augmented end-to-end multi-task learning for emotion prediction from speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[15] S. Xu, "Speech-to-text-wavenet: end-to-end sentence level Chinese speech recognition using deepmind's wavenet," 2020, https://github.com/CynthiaSuwi/Wavenet-demo.

[16] Kim and Park, "Speech-to-text-WaveNet," 2016, https://github.com/buriburisuri/ GitHub repository.

[17] A. van den Oord, A. Graves, H. Zen et al., "WaveNet: a generative model for raw audio," 2016, http://arxiv.org/abs/1609.03499.

[18] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, New York, NY, USA, 2012.

[19] M. Wei, "A novel face recognition in uncontrolled environment based on block 2D-CS-LBP features and deep residual network," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 2, pp. 207–221, 2020.

[20] A. S. Jadhav, P. B. Patil, and S. Biradar, "Computer-aided diabetic retinopathy diagnostic model using optimal thresholding merged with neural network," *International Journal of Intelligent Computing & Cybernetics*, vol. 13, no. 3, pp. 283–310, 2020.

[21] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2020, http://arxiv.org/abs/1508.04025.

[22] B. La, "Tibetan spoken language," in Chinese, 2005.

Hindawi
Complexity
Volume 2020, Article ID 8830359, 10 pages
https://doi.org/10.1155/2020/8830359

Research Article

Online Supervised Learning with Distributed Features over Multiagent System

Xibin An, Bing He, Chen Hu, and Bingqi Liu

High-Tech Institute of Xi'an, Xi'an 710025, China

Correspondence should be addressed to Bing He; 861427055@qq.com

Received 31 August 2020; Revised 27 September 2020; Accepted 7 October 2020; Published 16 November 2020

Academic Editor: Ning Cai

Most current online distributed machine learning algorithms have been studied in a data-parallel architecture among agents in networks. We study online distributed machine learning from a different perspective, where the features about the same samples are observed by multiple agents that wish to collaborate but do not exchange the raw data with each other. We propose a distributed feature online gradient descent algorithm and prove that local solution converges to the global minimizer with a sublinear rate $O(\sqrt{2T})$. Our algorithm does not require exchange of the primal data or even the model parameters between agents. Firstly, we design an auxiliary variable, which implies the information of the global features, and estimate at each agent by dynamic consensus method. Then, local parameters are updated by online gradient descent method based on local data stream. Simulations illustrate the performance of the proposed algorithm.

1. Introduction

With the development of multiagent system, the observed data are being generated at anywhere, anytime, using different devices and technologies [1–3]. There is a lot of interest in extracting knowledge from this massive amount of data and using it to choose a suitable business strategy [4–6], to generate control command [7–9] or to make a decision [10–13]. Many applications are required to process incoming data in online way, e.g., a bank monitors the transactions of its clients to detect frauds [2], wireless sensor networks makes inference [14], and sensor network tracks the uncooperative target [15]. The study of online learning is becoming an important topic of research itself [16–18].

The success of online machine learning often depends on the entire data stream. In some applications, the observed data may be generated on and held by multiple agents [1, 13]. Collecting data to a central site for training incurs extra management and privacy concerns [1]. As a result, some distributed machine learning algorithms have been proposed to train a model by letting each agent perform local model updates and exchange some information between neighbors [19–22]. Most of the existing algorithms fall into the data-parallel computation [1], where each agent has its local data stream with the entire features. However, in network applications, multiple agents are used to monitor an environment, where agents are distributed over space and are used to collect different measurements. For example, the observation is generated by different observed models [8, 9]. It is urgent to develop some applicable algorithm to deal with data streams with distributed features over networks.

In batch learning settings, some algorithms have been proposed for distributed features, such as variance-reduced dynamic diffusion (VRD^2) [12], feature distributed machine learning (FDML) [1], and the ADMM (alternating direction method of multipliers) sharing [23]. VRD^2 and FDML obtain the optimal solution in primal domain, and the local model is trained in a distributed manner based on the local features. The ADMM sharing algorithm formulates distributed feature learning as a distributed primal-dual problem and then obtains the optimal solution by ADMM algorithm. These algorithms in [1, 12, 23] effectively deal with the batch distributed feature learning in a distributed form. However, these algorithms in [1, 12, 23] need to access the entire dataset and cannot be applied in online settings. As the observation is continuously arriving very fast in

networks, it is important to study online feature distributed machine learning.

In this paper, we consider the situation where the features are split across agents in online settings either due to privacy consideration or because they are already physically collected in a distributed manner by means of a networked architecture. We propose a distributed feature online gradient algorithm. Online supervised learning over networks is formulated as a "cost of sum" form. The procedure of the proposed algorithm requires two-scales: one scale is used to update the parameters by gradient descent and a second faster scale for running the consistency step multiple times to track an auxiliary term. The main contributions of this paper are summarized as follows.

(1) We propose a distributed feature online gradient (DFOG) descent algorithm. By exchanging some information between neighbors, local solution can approximate the global solution. Compared with VRD^2 [12], FDML [1], and the sharing ADMM algorithm [23], DFOG is applicable to online supervised learning with distributed features over networks.

(2) We firstly formulate the centralized cost as a "cost of sum" form. By dynamic consensus algorithm, each node can track the sum term, which implies the entire features of the sample at each round time. Then, with the help of online gradient descent algorithm, each node locally updates the parameters based on its data stream.

(3) We prove that the proposed algorithm achieves an $O(\sqrt{2T})$ regret bound. That is, local solution can approach to the global solution, which is the best decision trained based on the entire dataset. The only transmitted message is some parameters' information, and the proposed algorithm does not require the data of the total number times and does not exchange the raw data between neighbors.

The rest of this paper is organized as follows: the problem formulation is discussed in Section 2. In Section 3, we focus on our online optimization algorithm with distributed features over multiagent system, followed by the theoretical results in Section 4. In Section 5, simulations illustrate the effectiveness of our algorithm. Finally, we conclude the paper in Section 6.

Notation and terminology: let x be the feature space and y be the corresponding label. We denote the (i, j)th element of a matrix A by $a_{i,j}$. For $t \in \mathbb{N}^+$, the set $\{1, 2, \ldots, T\}$ is denoted by $[T]$. For a convex function f, its gradient at a point ω is denoted as $\nabla_\omega f(\omega)$. We denote N as the number of agents in the network. Let $\mathbb{R}^d$ be the d-dimensional vector space and $\|\omega\|_2^2$ is the Euclidean norm of a vector $\omega \in \mathbb{R}^d$.

2. Problem Formulation

We consider a multiagent system with N agents. The communication between agents is described by a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [24], consisting of a set of nodes $\mathcal{V} = \{1, 2, \ldots, N\}$, a set of edges $\mathcal{E}$, and an adjacent matrix A [19]. For each agent $i \in \mathcal{V}$, we denote $\mathcal{E}_i = \{j \mid (j, i) \in \mathcal{E}\}$ as a set of neighbors of agent i (including agent i itself).

Assumption 1. The graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ and the adjacent weighted matrix A satisfy the following [25]:

(i) A is a doubly stochastic matrix with positive diagonal, that is, $a_{ii} > 0$, $\sum_{j=1}^N a_{j,i} = 1$, and $\sum_{j=1}^N a_{i,j} = 1$;

(ii) There exists a scalar $\zeta > 0$ such that $a_{j,i} > \zeta$ if $(j, i) \in \mathcal{E}$;

(iii) There exists an integer $B \geq 1$ such that the graph $(\mathcal{V}, \mathcal{E}_{i(B+1)} \cup \cdots \cup \mathcal{E}_{(j+1)B})$ is strongly connected.

In this work, we focus on a binary online supervised learning with distributed features. The features are distributed over a collection of K agents, as illustrated in Figure 1.

At each time $t = 1, 2, \ldots, T$, network receives a labeled sample (x_t, y_t). For all the time T, we consider an empirical risk as follows:

$$L(\omega) = \frac{1}{T} \sum_{t=1}^{T} f\left(\omega^T x_t, y_t\right) + r(\omega), \tag{1}$$

where the parameters are denoted as $\omega \in \mathbb{R}^{d\times 1}$, d is the dimension of the features, and $y_t \in \{-1, +1\}$ is the corresponding scalar label of x_t at time t. Moreover, the cost $f(\omega)$ is convex and differentiable. In most problem of interest, the cost function is dependent on the inner product $\omega^T x$, such as the linear SVM cost $f = \max(0, 1 - y_t(w^T x_t))$ and the logistic regression cost $f = \log(1 + \exp(-y_t(\omega^T x_t)))$. The factor $r(\omega)$ represents the regularization term. Since the features of x_t are distributed across agents, we set ω and x_t to be column vector and formulate ω and x_t into N subvectors denoted by ω_i and $x_{t,i}$, respectively, that is,

$$x_t = \begin{bmatrix} x_{t,1} \\ x_{t,2} \\ \vdots \\ x_{t,N} \end{bmatrix}, \qquad \omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{bmatrix}. \tag{2}$$

Each subfeature $x_{t,i}$ vector and subvector ω_i is located at agent i. Then, cost function (1) can be rewritten as

$$L = \frac{1}{T} \sum_{t=1}^{T} f\left(\sum_{i=1}^{N} \omega_i^T x_{t,i}; y_t\right) + \sum_{i=1}^{N} r(\omega_i), \tag{3}$$

where the regularization term is assumed to satisfy an additive form as

$$r(\omega) = \sum_{i=1}^{N} r(\omega_i). \tag{4}$$

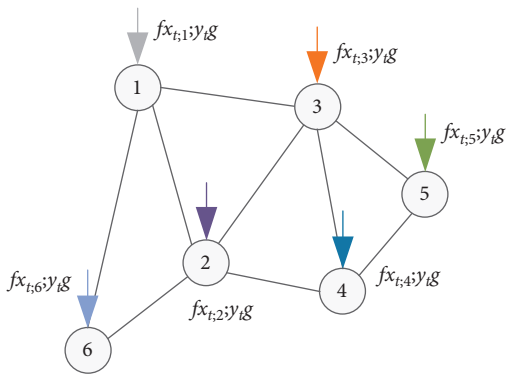


Figure 1: Distributing the features across agents.

This property holds for many popular regularization choices, such as l_2, l_1, and KL-divergence. Problem of this type has been studied before in the literature by using distributed optimization methods in [20, 21]. One common way is to formulate problem (3) into a constrained problem, that is,

$$L(\omega) = \min_{\omega} \frac{1}{T}\sum_{t=1}^{T} f(z_t; y_t) + \sum_{i=1}^{N} r(\omega_i) \text{s.t. } z_t = \sum_{i=1}^{N} \omega_i^T x_{t,i}, \quad t = 1, 2, \ldots, T. \tag{5}$$

For all the time T, problem (5) is a classical "cost of sum" form [20]. An effective way is to design the Lagrangian function by introducing the dual variable γ [23], namely,

$$L(\gamma, z, \omega) = \frac{1}{T}\sum_{t=1}^{T} f(z_t; y_t) + \frac{1}{T}\sum_{t=1}^{T} \gamma_t z_t - \frac{1}{T}\sum_{t=1}^{T} \gamma_t \sum_{i=1}^{N} \omega_i^T x_{i,t} + \sum_{i=1}^{N} r(\omega_i). \tag{6}$$

Problem (6) can be solved in a number of distributed primal-dual methods, such as alternating direction method of multipliers (ADMM) [4, 22, 26] and primal-dual methods [27–29]. These techniques have good convergence properties but suffer from high computational costs and two-time scale communications.

The other way is studied in primal domain [12]. The algorithm in [12] requires a two-time scale operation: a faster time-scale for the consensus iterations and a slower time-scale for the data sampling and the gradient computing. First, we use a consensus strategy to obtain the sum term $\sum_{i=1}^{N} \omega_i^T x_{t,i}$, namely,

$$\widehat{z}_{n_k,i} = \sum_{j \in \mathscr{E}_i} a_{ij} N \omega_{k,j}^T x_{n_k,j}, \tag{7}$$

where n_k denotes the index of the sample selected uniformly at random from $\{1, 2, \ldots, T\}$. After sufficient iterations, it is well-known that $\widehat{z}_{n_k,i} \longrightarrow (1/N)\sum_{i=1}^{N} z_{n_k,i}$. Then, the stochastic-gradient step is used to update the parameters ω, where the gradient is evaluated by the gradient vector of the cost evaluated at some random data (x_{n_k}, y_{n_k}).

In online settings, since the data (x_t, y_t) is observed one by one, we cannot access to the total dataset $\{(x_t, y_t)\}_{t=1}^{T}$. These algorithms in [1, 12, 23] cannot be applied for data stream with distributed feature over networks. For each time $t = 1, 2, \ldots, T$, the multiagent system is endowed with a sequence of cost function $\{L_t\}_{t=1}^{T}$, and the goal is to minimize the sum of the cost function. Specifically, we want to minimize the difference between the total cost multiagent system has incurred and that of the best fixed decision in hindsight, which is called regret, and its definition is given as follows:

$$\text{Reg}^T = \sum_{t=1}^{T} L_t(\omega_t) - \sum_{t=1}^{T} L_t(\omega^*), \tag{8}$$

where ω^* is the best decision of problem (1), that is,

$$\omega^* = \arg\min_{\omega} \sum_{t=1}^{T} L_t(\omega). \tag{9}$$

Moreover, we consider the time-varying cost function L_t as

$$L_t(\omega_t) = Q\left(\sum_{i=1}^{N} \omega_{t,i}^T x_{t,i}; y_t\right) + \sum_{i=1}^{N} r(\omega_{t,i}). \tag{10}$$

Generally speaking, the cost $Q(\sum_{i=1}^{N} \omega_i^T x_{t,i}; y_t)$ satisfies Assumption 2.

Assumption 2. The loss function $Q(\cdot)$ is convex and differentiable, and the gradient $\nabla_\omega Q(\omega)$ is uniform boundedness, that is, $\|\nabla_\omega Q(\omega)\| \le C$ for some scalar $C > 0$.

Regret is the standard measure of the performance of online optimization algorithm [19]. An algorithm attains good performance if the regret is sublinear as a function of the total time T.

Remark 1. In the multiagent system, since the entries of the feature x_t are distributed over N agents, each agent just observes its own data stream. We face the following two challenges in solving problem (8):

(1) Distributed challenge: each agent only receives local data stream $(x_{t,i}, y_t)$ and does not access to the entire features (x_t, y_t). Under the condition that we do not exchange the raw data between neighbors, each agent needs to obtain some information on the entire features.

(2) Online challenge: at any time t_1, we only have observation for $t \le t_1$ and do not know L_t for $t_1 \le t \le T$. It is difficult to store all the observations due to the high-dimensional and high-velocity data stream. We need to update the parameters based on the current sample and the previous parameters and pursue a solution approximating to the global solution ω^*,

which is the best decision based on all the data $\{(x_t, y_t)\}_{t=1}^{t}$ as a prior in offline settings.

3. Distributed Feature Online Gradient Descent Algorithm

In this section, we first analyse a dynamic average consensus method for approximating the sum of $\omega_i^T x_{t,i}$ at agent i and propose an online convex optimization to update the parameters ω. The detailed framework is summarised in Figure 2.

Now, we consider the problem of minimizing (5) by means of an online convex optimization. Let $z_t = \sum_{i=1}^{N} \omega_{t,i}^T x_{t,i}$ denote the inner product that is available at time $t \in [T]$. The cost function L_t can be described as

$$L_t(\omega_t) = Q(z_t; y_t) + \sum_{i=1}^{N} r(\omega_{t,i}). \tag{11}$$

If each agent i can obtain the auxiliary variable z_t at any time t, the parameters $\omega_{t,i}$ can be obtained by minimizing the local cost $L_{t,i}$, which is defined as

$$L_{t,i} = Q(z_t; y_t) + r(\omega_{t,i}). \tag{12}$$

However, the computation of z_t needs to access to all the subfeatures $x_{t,i}$ and the subvectors $\omega_{t,i}$ over N agents. We denote the average of the local inner products as

$$\overline{z}_t = \frac{1}{N} \sum_{i=1}^{N} \omega_{t,i}^T x_{t,i}. \tag{13}$$

Motivated by works in [30–34], $\overline{z}_t$ can be approximated by a diffusion-based algorithm. Since the desired variable z_t is proportional to the average value $\overline{z}_t$, $z_t = N\overline{z}_t$, the consensus strategy can be used to approximate z_t. Specifically, for the total number of iterations M, each agent would repeat the following steps M times:

$$\widehat{z}_{t,i}^{m+1} = \sum_{j \in \mathrm{E}_i} a_{ij} \widehat{z}_{t,i}^{m}, \quad m = 0, 1, \ldots M-1, \tag{14}$$

where $\widehat{z}_{t,i}^{0} = N\omega_{t,i}^T x_{t,i}$. After each agent obtains the estimator of z_t denoted as $\widehat{z}_{t,i}$, problem (12) is converted into a differentiable dynamic problem. For online convex optimization problem, online gradient descent and its variants have been achieving optimal dynamic regret in many applications [35]. Recalling that ω_t and x_t are partitioned into N blocks, the gradient step can be performed in parallel over N agents. Specifically,

$$\omega_{t,i} = \omega_{t-1,i} - \mu_t \nabla_z Q(\widehat{z}_{t,i}; y_t) x_{t,i} - \mu_t \nabla_\omega r(\omega_{t,i}), \tag{15}$$

where the step-size μ_t should satisfy $\mu_t > 0$, $\sum_{t=1}^{\infty} \mu_t = \infty$, and $\sum_{t=1}^{\infty} \mu_t^2 < \infty$.

The full algorithm is summarized in Algorithm 1.

Remark 2. Compared with FDML [1], VRD^2 [12], and the ADMM sharing algorithm [23], DFOG is applicable for data stream with distributed features over multiagent system. At each round time, agents observe the same sample from different features. Each agent can obtain an auxiliary term, which implies the information on the entire features. Then, each agent locally runs a gradient descent step to update its local parameters. The procedure of Algorithm 1 is designed to update the parameters $\omega_{t,i}$ locally.

4. Algorithm Analysis

4.1. Convergency Analysis. In this section, we analyse the convergence of the proposed algorithm. We first show that the distance between $\widehat{z}_{t,i}$ and z_t is upper bounded by the difference between P^M and $1/N$, which is shown in Lemma 1 and proved in [25].

Lemma 1. *Let Assumption 1 holds, for all agents i, j; we have*

$$\left| \left[P^M\right]_{ij} - \frac{1}{N} \right| \le \left(1 - \frac{\zeta}{4N^2}\right)^{(M/B)-2}, \tag{16}$$

where N is total number of agents and M is the number of consensus steps in(14).

Then, we show that the regret of online gradient descent (OGD) is upper bounded by the cumulative difference between the loss of ω_t and ω_{t+1}, which is present in Lemma 2 and proved in [18].

Lemma 2. *Let $\{\omega_{t,i}\}_{t=1}^{T}$ denotes the sequence of parameters produced by OGD. Then, for any u, we have*

$$\mathrm{Reg}_i^T = \sum_{t=1}^{T} \left(L_{t,i}(\omega_{t,i}) - L_{t,i}(u)\right) \le r(u) - r(\omega_{1,i}) + \sum_{t=1}^{T} \left(Q(\omega_{t,i}) - Q(\omega_{t+1,i})\right). \tag{17}$$

Because the features are distributed across agents, Reg_i^T mainly illustrates the difference between local parameters ω_i and the corresponding parameters ω_i^* in global solution. Based on the above lemma, we derive a regret bound of ω_i for DFOG with the regularization term $r(\omega_i) = (1/2)\mu\|\omega_i\|_2^2$.

Theorem 1. *Let Assumptions 1 and 2 hold, and consider running DFOG on a sequence of convex function, $Q(\omega_{t,i})$ for all t, with the regularization term $r(\omega_i) = (1/2)\mu\|\omega_i\|_2^2$. Let $\{\omega_{t,i}\}_{t=1}^{T}$ be the sequence of vectors produced by DFOG. If $\|u\| \le U$ and $\mu = (U/C)\sqrt{2T}$, the regret of ω_i satisfies*

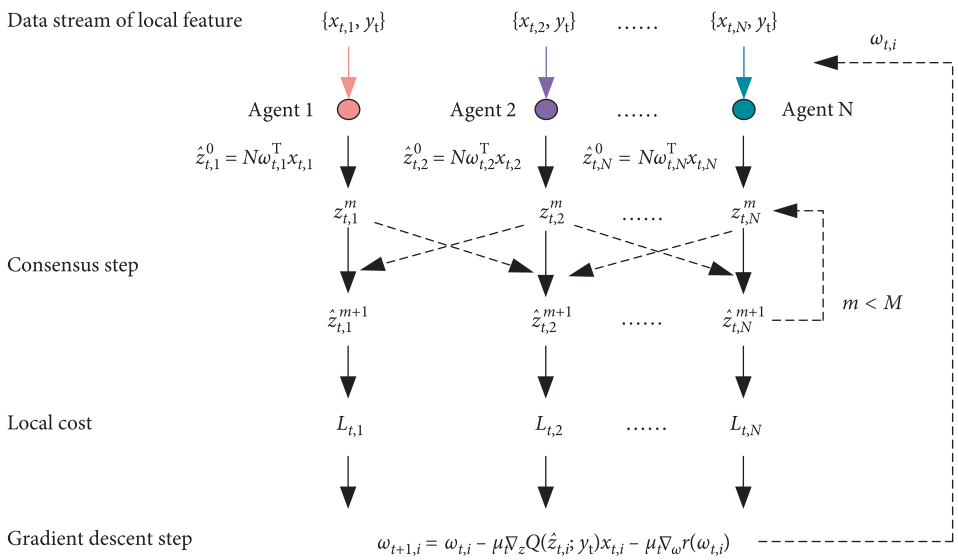


Figure 2: The framework of the proposed algorithm.

(1) Initialization: set $\omega_{0,i} = 0$.
(2) Repeat for $i = 1, 2, \ldots, N$:
(3) $\widehat{z}_{0,i} = N\omega^T_{0,i}x_{t,i}$
(4) For $m = 0, 1, 2, \ldots, M-1$
(5) $\widehat{z}^{m+1}_{t,i} = \sum_{j\in\mathscr{E}_i} a_{ij}\widehat{z}^m_{t,i}$
(6) End
(7) $\omega_{t,i} = \omega_{t-1,i} - \mu_t\nabla_z Q(\widehat{z}_{t,i}; y_t)x_{t,i} - \mu_t\nabla_\omega r(\omega_{t,i})$
(8) End

Algorithm 1: Distributed feature online gradient (DFOG) descent for agent i.

$$\mathrm{Reg}_i^T \le \mathrm{CU}\sqrt{2T} + C_2U\sqrt{2T} + U^2\sqrt{2T}, \tag{18}$$

where $C_2 = (C/2)(1 - (\zeta/4N^2))^{(M/B)-2}\|z_t\|_*\|x\|_*$. *The proof is presented in Appendix.*

Remark 3. This theorem indicates that the convergence rate of DFOG depends on the network topology through B and the number of consensus steps M. The larger the M is or the smaller the B is, the faster the convergence speed is. The theorem presents that the proposed algorithm converges to the global solution with sublinear rate. When the number of data samples increases, the difference between $\omega_{t,i}$ with ω_i^* will become closer.

4.2. Complexity Analysis

4.2.1. Time Complexity. There are two primary operations associated with learning for DFOG: (1) estimating the inner product $\widehat{z}_t$ for each sample at time t and (2) updating the parameters at gradient descent step. At any time t, each estimator $\widehat{z}_t$ computation requires $O(M)$ arithmetic operations. There is one gradient descent step to update the parameters, which requires $O(1)$ arithmetic operations. As for each time, each node will require $O(M)$ arithmetic operations. Hence, single node requires $O(\mathrm{TM})$ arithmetic operations for DFOG.

4.2.2. Space Complexity. At any time t, DFOG needs to store the parameters $\widehat{z}_t$ and ω_t, which are updated and time-varying. Hence, space complexity for DFOG is $O(1)$.

4.2.3. Communication Complexity. We denote the average degree of the communication graph as k. At each consensus step, each node requires to exchange $\widehat{z}_t$ (float type, 4 bytes) with its neighbors. Since the network topology is an undirected graph, it requires 8 kM bytes at any time t. Hence, DFOG requires communication traffic of DFOG is 8 kMT bytes for all the time T.

5. Simulation

In this section, we test our algorithm by minimizing norm regularized logistic regression on two public datasets, a9a and bank from UCI. Here, a multiagent system with 6 agents is considered, and the network is generated by the random geometric graph model. a9a dataset consists of 32561

Table 1: Parameter settings.

Parameter	Value
λ	0.1
N	5
M	30
B	10

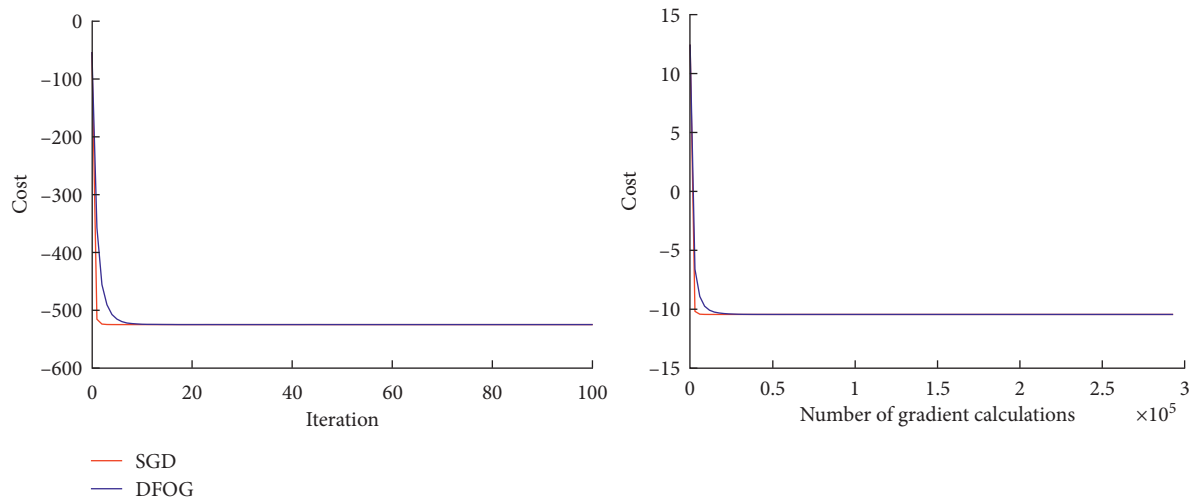


Figure 3: The evolution of cost for a9a dataset.

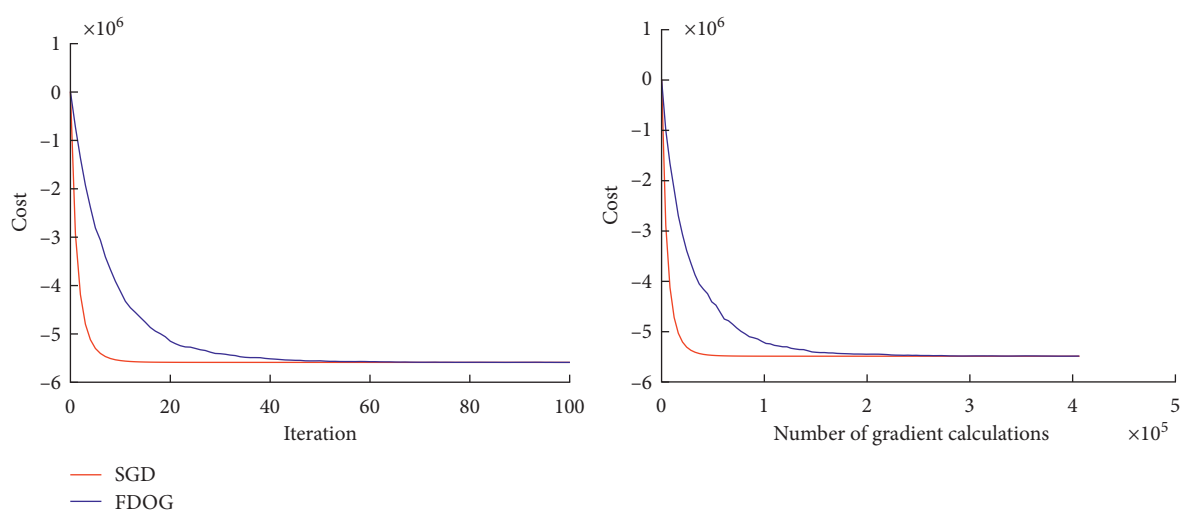


Figure 4: The evolution of cost for bank dataset.

training samples, 16281 testing samples, and 123 features. We separate the features into 6 parts, and each node obtains one part with 21, 21, 21, 20, 20, and 20 features as the local data, respectively. On the other hand, the bank dataset contains 4068 training samples, 453 testing samples, and 17 features. Similarly, we divide the features into 6 parts, each agent gets one part with 3, 3, 3, 3, 3, and 2 features as the local data, respectively. The loss function we use is

$$L(\omega) = \frac{1}{T}\sum_{t=1}^{T}\log\left(1+\exp\left(-y_t\sum_{i=1}^{N}\omega_i^T x_{t,i}\right)\right)+\lambda\sum_{i=1}^{N}\|\omega_i\|^2. \tag{19}$$

Generally, λ is a positive constant and $\lambda > 0$. The simulations are implemented in MATLAB to verify the proposed algorithm. Specifically, we use two optimization libraries, SGDLibrary [36] and DADAM [37], to minimize

Table 2: Testing error and the relative error of parameters.

Dataset	Testing error of SGD	Testing error of DFOG	$\sum_{i=1}^{N} \|\omega_i - \omega_i^*\| / \|\omega^*\|$
9a	0.7581	0.7581	0.0026
Bank	0.8455	0.8455	0.0064

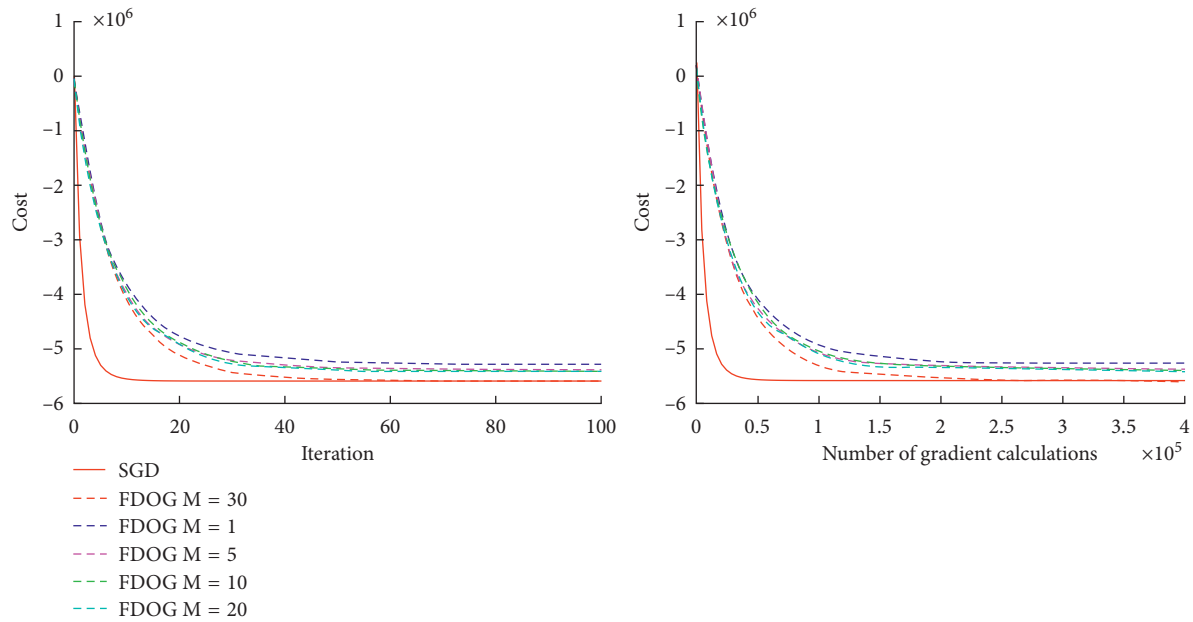


Figure 5: The convergence behavior with different M.

(19). In our simulation, the parameters are set as summarised in Table 1.

We adopt dynamic consensus method to obtain z_t in (14) and use online gradient descent algorithm to update the parameter ω_i locally in (15). In our simulations, we get the global model trained in a centralized manner if all the features were collected centrally by stochastic-gradient descent (SGD) algorithm. Next, we compare our algorithm against SGD algorithm proposed in [38] and keep track of the loss for different datasets and parameters settings. Figures 3 and 4 present the evolution of the cost during the training procedure for a9a and bank datasets, respectively. In addition, to make a fair comparison, we analyse the convergence curve based on the count of gradient calculated. Table 2 shows the testing error for different datasets and the error of parameters for DFOG and SGD. The results show that DFOG can converge to the centralized solution of SGD, while keeping local feature sets to the corresponding agent. That is, DFOG can deal with the online supervised learning problem caused by distributed features over networks.

We next show how the performance depends on different M. Note that when M is larger, we need to do more communication on the consistency step (14). Figure 5 shows the evolution of the cost with different M. It can be found that the larger the M we set, the faster the DFOG approaches to the centralized SGD algorithm.

6. Conclusions

In this paper, we considered an online supervised learning problem where the features are split across agents in online settings. We proposed an online supervised learning algorithm with distributed features over multiagent system. We first formulated the centralized cost as a "cost of sum" form. By dynamic consensus algorithm, each agent could effectively estimate the sum term, which is calculated based on the entire features at each round time. Then, with the help of online gradient descent algorithm, each agent locally updates the parameters. The algorithm designed does not require the data of the total number times and does not communicate the raw data between neighbors. We proved that local solution converges to the centralize minimizer, which is the best decision trained based on the entire dataset, and the proposed algorithm achieves an $\mathcal{O}(\sqrt{2T})$ regret bound. Simulations with real dataset verify the conclusion.

Distributed machining learning algorithms are worth of further studies due to their promising future, including distributed online boosting, distributed decision tree [39], the use of Big data-aided learning [40], and distributed learning over time-varying communication topology in networks.

Appendix

Proof of Theorem 1: let Assumption 2 holds, for each time t, then we have

$$Q\left(\omega_{t,i}\right) - Q(u) \leq < \omega_{t,i} - u, \quad \nabla_{\omega_{t,i}} Q >, \tag{A.1}$$

where $\langle \omega_i - u, \nabla_{\omega_{t,i}} Q\rangle$ is the inner product between vectors $\omega_i - u$ and $\nabla_{\omega_{t,i}} Q$. Moreover, we denote $\|\omega\| = \sqrt{\langle \omega, \omega\rangle}$.

Using Lemma 2, we obtain

$$\mathrm{Reg}_i^T \leq r(u) - r(\omega_1) + \sum_{t=1}^{T}\left(Q\left(\omega_{t,i}\right) - Q\left(\omega_{t+1,i}\right)\right) \leq \frac{1}{2\mu}\|u\|_2^2 + \sum_{t=1}^{T}\langle \omega_{t,i} - \omega_{t+1,i}, \nabla_{\omega_{t,i}} Q\rangle. \tag{A.2}$$

From equation (15), we have

$$\sum_{t=1}^{T}\langle \omega_{t,i} - \omega_{t+1,i}, \nabla_{\omega_{t,i}} Q\rangle = \sum_{t=1}^{T}\left(\nabla_{\widehat{z}_{t,i}} Q\left(\omega_{t,i}\right) x_{t,i} + \nabla_{\omega_{t,i}} r\left(\omega_{t,i}\right)\right)\nabla_{\omega_{t,i}} Q, \tag{A.3}$$

where $\widehat{z}_{t,i} = \sum_{i=1}^{N}[P^M]_{ij}\omega_{t,i}x_{t,i}$.

We derive the gradient of cost (12) as follows:

$$\nabla_{\omega_{t,i}} Q = \nabla_{z_t} Q\left(\omega_{t,i}\right) x_{t,i}. \tag{A.4}$$

Substituting (A.4) into (A.3),

$$\begin{aligned}\sum_{t=1}^{T}\langle \omega_{t,i} - \omega_{t+1,i}, \nabla_{\omega_{t,i}} Q\rangle &= \sum_{t=1}^{T}\left(\nabla_{\widehat{z}_{t,i}} Q \cdot x_{t,i} + \nabla_{\omega_{t,i}} r\right)\nabla_{\omega_{t,i}} Q \\ &= \sum_{t=1}^{T}\left\|\nabla_{\omega_{t,i}} Q\right\|_2^2 + \sum_{t=1}^{T}\left(\nabla_{\widehat{z}_{t,i}} Q - \nabla_{\overline{z}_{t,i}} Q\right) x_{t,i}\nabla_{\omega_{t,i}} Q + \sum_{t=1}^{T}\nabla_{\omega_{t,i}} r\nabla_{\omega_{t,i}} Q \\ &\leq TC^2 + \sum_{t=1}^{T} C\left\|\left(\nabla_{\widehat{z}_{t,i}} Q\left(\omega_{t,i}\right) - \nabla_{\overline{z}_{t,i}} Q\left(\omega_{t,i}\right)\right)\right\|\left\|x_{t,i}\right\| + 2\mathrm{TCU}.\end{aligned} \tag{A.5}$$

Let Assumption 2 holds such that $\|(\nabla_{\widehat{z}_{t,i}} Q(\omega_{t,i}) - \nabla_{\overline{z}_{t,i}} Q(\omega_{t,i}))\| \leq C\|\widehat{z}_{t,i} - \overline{z}_{t,i}\|$. Using Lemma 1, we have

$$\left\|\left(\nabla_{\widehat{z}_{t,i}} Q\left(\omega_{t,i}\right) - \nabla_{\overline{z}_{t,i}} Q\left(\omega_{t,i}\right)\right)\right\| \cdot \left\|x_{t,i}\right\| \leq C\left\|\widehat{z}_{t,i} - \overline{z}_{t,i}\right\| \leq C\left(1 - \frac{\zeta}{4N^2}\right)^{(M/B)-2}\left\|\sum_{i=1}^{N}\omega_{t,i}x_{t,i}\right\|. \tag{A.6}$$

Denoting $\|x\|_* = \max\|x_{t,i}\|$ and $\|z_t\|_* = \max\|\sum_{i=1}^{N}\omega_{t,i}x_{t,i}\|$ for $t = 1, \ldots, T$, we derive

$$\mathrm{Reg}_i^T \leq \frac{1}{2\mu}\|u\|_2^2 + \mu\mathrm{TC}^2 + \mu\mathrm{TC}\left(1 - \frac{\zeta}{4N^2}\right)^{(M/B)-2}\left\|z_t\right\|_*\|x\|_* + 2\mu\mathrm{CTU}. \tag{A.7}$$

If $\|u\| \leq U$ and $\mu_t = U/C\sqrt{2T}$, then

$$\mathrm{Reg}_i^T \leq \mathrm{CU}\sqrt{2T} + C_2U\sqrt{2T} + U^2\sqrt{2T}, \tag{A.8}$$

where $C_2 = (1/2)(1 - (\zeta/4N^2))^{(M/B)-2}\|z_t\|_*\|x\|_*$. Theorem 1 has been proved.

Data Availability

a9a dataset has been derived from https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html. Bank dataset has been derived from https://archive.ics.uci.edu/ml/datasets/Bank+Marketing.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Hu, D. Niu, J. Yang, and S. Zhou, "FDML: a collaborative machine learning framework for distributed features," in *Proceedings of the Knowledge Discovery and Data Mining*, pp. 2232–2240, Anchorage, AK, USA, August 2019.

[2] A. T. Vu, G. D. F. Morales, J. Gama et al., "Distributed adaptive model rules for mining big data streams," in *Proceedings of the 2014 IEEE International Conference on Big Data (Big Data)*, IEEE, Washington DC, USA, October 2014.

[3] N. Cai, C. Diao, and M. J. Khan, "A novel clustering method based on quasi-consensus motions of dynamical multiagent systems," *Complexity*, vol. 2017, Article ID 4978613, , 2017.

[4] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 482–497, 2015.

[5] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, p. 4585, 2020.

[6] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 7, pp. 2442–2454, 2020.

[7] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[8] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, p. 1, 2020.

[9] S. Liu, Z. Ji, and H. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, Article ID 7685460, 2020.

[10] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[11] B. Ying and A. H. Sayed, "Diffusion gradient boosting for networked learning," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2512–2516, New Orleans, LA, USA, March 2017.

[12] B. Ying, K. Yuan, and A. H. Sayed, "Supervised learning under distributed features," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 977–992, 2019.

[13] N. Cai, M. He, Q. Wu, and M. J. Khan, "On almost controllability of dynamical complex networks with noises," *Journal of Systems Science and Complexity*, vol. 32, no. 4, pp. 1125–1139, 2019.

[14] D. Ciuonzo, P. S. Rossi, and P. Willett, "Generalized rao test for decentralized detection of an uncooperative target," *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 678–682, 2017.

[15] J. B. Predd, S. B. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, 2006.

[16] P. Sharma, P. Khanduri, and L. Shen, "On distributed online convex optimization with sublinear dynamic regret and fit," 2020, https://arxiv.org/abs/2001.03166.

[17] A. Murdopo, "Distributed decision tree learning for mining big data streams," Master of Science Thesis, European Master in Distributed Computing, Dresden, Germany, 2013.

[18] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[19] D. Yuan, D. W. C. Ho, and G. P. Jiang, "An adaptive primal-dual subgradient algorithm for online distributed constrained optimization," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3045–3055, 2017.

[20] J. Chen, Z. J. Towfic, and A. H. Sayed, "Dictionary learning over distributed models," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1001–1016, 2015.

[21] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization with coupling constraints," *Automatica*, vol. 84, pp. 149–158, 2017.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[23] Y. Hu, P. Liu, L. Kong, and D. Niu, "Learning privately over distributed features: an ADMM sharing approach," 2019, https://arxiv.org/abs/1907.07735.

[24] G. C. Rota, "Algebraic graph theory," *Graduate Texts in Mathematics*, vol. 207, no. 32, p. 298, 1994.

[25] A. Nedic, A. Olshevsky, and A. Ozdaglar, "Distributed subgradient methods and quantization effects," in *Proceedings of the Conference on Decision and Control*, pp. 4177–4184, Cancun, Mexico, December 2008.

[26] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the admm in decentralized consensus optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1750–1761, 2014.

[27] T.-H. Chang, A. Nedic, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1524–1538, 2014.

[28] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: an exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

[29] K. Yuan, B. Ying, X. Zhao, and A. H. Sayed, "Exact diffusion for distributed optimization and learning–part I: algorithm development," 2017, https://arxiv.org/abs/1702.05122.

[30] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

[31] A. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[32] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[33] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental

limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, 2011.

[34] R. A. Freeman, P. Yang, and K. M. Lynch, "Stability and convergence properties of dynamic average consensus estimators," in *Proceedings of the Conference on Decision and Control*, pp. 338–343, San Diego, CA, USA, December 2006.

[35] T. Yang, L. Zhang, R. Jin, and J. Yi, "Tracking slowly moving clairvoyant: optimal dynamic regret of online learning with true and noisy gradient," in *Proceedings of The 33rd International Conference on Machine Learning, ser. Proceedings of Machine Learning Research*, pp. 449–457, PMLR, New York, NY, USA, June 2016.

[36] H. Kasai, "SGDLibrary: a MATLAB library for stochastic optimization algorithms," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 7942–7946, 2017.

[37] P. Nazari, T. Davoud Ataee, and M. George, "Dadam: a consensus-based distributed adaptive gradient method for online optimization," 2019, https://arxiv.org/abs/1901.09109v6.

[38] D. Saad, "Online algorithms and stochastic approximations," *Online Learning*, vol. 5, pp. 6–13, 1998.

[39] J. Ye, "Stochastic gradient boosted distributed decision trees," in *Proceedings of the Conference on Information and Knowledge Management*, pp. 2061–2064, Hong Kong, China, November 2009.

[40] G. Aceto, "Know your big data trade-offs when classifying encrypted mobile traffic with deep learning," in *Proceedings of the Traffic Monitoring and Analysis Conference*, pp. 121–128, Paris, France, June 2019.

Hindawi
Complexity
Volume 2020, Article ID 8816578, 13 pages
https://doi.org/10.1155/2020/8816578

Research Article

Guaranteed Cost Formation Tracking Control for Swarm Systems with Intermittent Communications

Purui Zhang,[1] Xiaoqian Chen,[1] and Xiaogang Yang[2]

[1]*College of Aerospace and Engineering, National University of Defense Technology, Changsha 410003, China*
[2]*High-Tech Institute of Xi'an, Xi'an 710025, China*

Correspondence should be addressed to Xiaogang Yang; doctoryxg@163.com

Received 25 August 2020; Revised 7 October 2020; Accepted 17 October 2020; Published 16 November 2020

Academic Editor: Ning Cai

The current paper studies guaranteed cost time-varying formation tracking design and analysis problems of high-order swarm systems subject to intermittent communications. Different from the existing work of the time-varying formation control, the time-varying formation tracking can be achieved while certain performance can be guaranteed, and the impacts of the intermittent communications and switching topologies are considered. First, a new intermittent time-varying formation tracking control protocol with a global performance index is proposed, where not only the formation regulation performances but also the control energy expenditures are involved. The codesign of the gain matrix with the performance index is achieved to compromise the formation regulation performances against control energy expenditures, and the guaranteed cost is determined to restrain the upper bound of the performance index. Then, guaranteed cost time-varying formation tracking design and analysis criteria are given, where the matrix variable of the linear matrix inequality conditions is used to design the gain matrix and to determine the guaranteed cost. Finally, a simulation example is provided to illustrate the effectiveness of the theoretical results.

1. Introduction

As one of the most important topics of the distributed cooperative control of swarm systems, formation control has aroused many attentions from researchers in recent years [1–7]. Distributed formation control means to design formation control protocol using only local information such that a team of autonomous agents forms and maintains the expected geometrical shape. Recently, due to the rapid development of the consensus theory, many scholars investigated the formation control problem via consensus-based approaches [8–14]. The core idea of the consensus-based formation control is to drive the agents to the desired states such that they can keep the specified difference from the virtual agreement states, which can be determined by the consensus control. Formation can be categorized into leaderless formation and formation tracking according to the different communication topology structure. For the leaderless one, each agent plays equal roles to determine the formation shape cooperatively. However, for the formation tracking, the followers should form the expected formation and track the leader, which determines the swarm property of the whole swarm systems.

A basic problem of the consensus-based formation control is the time-invariant formation control, where the expected formation shape is fixed. In this case, the relative position between any two agents will not change when the formation is formed. Time-invariant formation control/formation tracking can be regarded as the directed extension of the consensus/consensus tracking, and it was widely investigated in recent years [15–19]. However, due to the complicated task requirements and task updates, time-varying formation tracking is often needed in many applications, such as cooperative attack task, obstacle avoidance, and resource exploration. Compared with the time-invariant formation, time-varying formation tracking is more challenging since it should consider the impacts of the derivate of the formation function and the formation changes in time. For second-order swarm systems with switching topologies, the time-varying formation tracking

control was studied [20]. Wang et al. [21] investigated the robust time-varying formation design problems for second-order swarm systems with external disturbances where a new distributed extended state observer was constructed to compensate the influence of the disturbances. For general high-order swarm systems, time-varying formation tracking control and adaptive time-varying formation control were addressed [22, 23], respectively.

In many practical applications, the communication among agents may not always be continuous due to some communication faults including congestion of communication channels, packet losses, and sensing device failures. These communication faults can be modeled as intermittent communication where each agent can exchange its information to its neighbors over connected communication time units, but the interaction among agents will disappear in disconnected communication time units. Consensus of second-order swarm systems with intermittent communications was studied in [24]. Considering the impact of time delays and intermittent communications, Fattahi and Afshar [25] investigated the adaptive consensus control problem for high-order swarm systems. Sun and Wang [26] addressed the consensus problem for high-order swarm systems with Lipschitz nonlinear dynamics and intermittent communications where an interesting sampling time unit approach was proposed to transfer the intermittent consensus control problems to asymptotical stability problems of the swarm systems with input delays. In [24–26], the communication topology was fixed over connected communication time units. However, many swarm systems suffer switching topologies due to the changes of communication channels among agents as shown in [27–29] where the intermittent communications were not considered. It should be noted that it is difficult to deal with the intermittent communications and the switching topologies over connected communication time units simultaneously, and they were not considered in time-varying formation tracking problems, which form one of the motivations of the current paper.

Note that the abovementioned works only studied the formation achievement strategy without considering the performance constraints. However, it makes much sense to investigate the time-varying formation tracking control method with optimal/suboptimal performance indexes since there exist many resource limitations, and the control protocol should be optimized in real-world applications. In this sense, it is challenging to design a proper formation control protocol such that the time-varying formation tracking can be achieved, while the associated performance is guaranteed. For consensus control of swarm systems, there are some interesting works that address the optimal/suboptimal control problems. In [30], the optimal consensus algorithm was proposed with global performance indexes for first-order swarm systems where it was shown that the complete graph is needed to realize the global optimization of consensus. To relax the topology from the complete graph to the connected undirected graph, guaranteed cost consensus was achieved with different conditions in [31–33]. However, there are rare works that consider the time-varying formation tracking with guaranteed cost performance analysis, and it is interesting to investigate how the formation variance affects the guaranteed cost of swarm systems, which also motivates the study of the current paper.

Guaranteed cost time-varying formation tracking problems for high-order swarm systems with intermittent communications and switching topologies are addressed in the current paper. First, an intermittent time-varying formation tracking control protocol containing the switching topologies is constructed with the corresponding performance index. Then, the dynamics of the whole closed-loop system is decomposed into two parts by a nonsingular transformation and an orthonormal transformation successively, which can convert the formation tracking problem of the swarm system to the asymptotical stability problem of the reduced-order system. The stability of the reduced-order system is analyzed, respectively, in the connected communication time units and the disconnected communication time units to obtain the exponential condition of the asymptotical stability. Sufficient conditions of guaranteed cost time-varying formation tracking design and analysis are derived in the form of linear matrix inequality, and the guaranteed cost is determined to restrain the upper bound of the performance index.

Compared with the relative works about the formation control, the contributions of the current paper are twofold. First, the guaranteed cost time-varying formation tracking control problem is addressed, which can ensure that swarm systems can not only achieve the time-varying formation tracking but also satisfy the guaranteed cost constraints; that is, the compromise design between the formation regulation performance and the control energy expenditure should be realized with respect to the proposed performance index, and the guaranteed cost is determined to describe the upper bound of the performance index. However, the results on the time-varying formation control in [20–23] did not consider the impact of the guaranteed cost constraints. Second, the communication constraints of both intermittent communications and switching topologies are introduced into the design process of the guaranteed cost time-varying formation tracking, which contains two challenging problems. The first one is that the swarm stability of the whole swarm systems should be analyzed in the connected communication time units and the disconnected communication time units, respectively. In this case, the stability analysis methods in [20–23] are invalid, and the divergent property of the whole system is tackled in the disconnected communication time units by proposing the new method. The second one is that the impact of the intermittent communications should be considered for the guaranteed cost performance analysis. In this sense, the performance index becomes a piecewise continuous integral function and is difficult to be addressed when deriving the main results of the current paper.

The rest of the current paper is shown in the following sections orderly. Section 2 formulates the problem model where the basic concepts of the communication topology are introduced and the formation tracking control protocol is proposed. In Section 3, guaranteed cost time-varying formation tracking design and analysis criteria are given, respectively, and the guaranteed cost is determined. Section 4

presents a numerical simulation to illustrate the correctness of the proposed theorems. In Section 5, the main results of the current paper are summarized. Throughout the whole paper, $\mathbb{N}$ and $\mathbb{N}^+$ are used to stand for the natural numbers and positive natural numbers, respectively. $\mathbb{R}$ represents the real matrices with proper dimensions. $\otimes$ is the Kronecker product, and $*$ is the symmetric terms in the matrix. The positive definite and symmetric matrix is denoted by $P^T = P > 0$.

2. Problem Description

2.1. Model the Switching Communication Topology. Each communication topology in the switching topology set, $G_a = \{\mathcal{G}_a^1, \mathcal{G}_a^2, \ldots, \mathcal{G}_a^j\}$ $(j \geq 2)$ is modeled as the directed graph $\mathcal{G}_a$, where the node set is represented by $\mathcal{G} = \{n_1, n_2, \ldots, n_N\}$, and the edge set is denoted by $\mathcal{E} = \{(n_m, n_k)\colon n_m, n_k \in \mathcal{G}\}$. Let $\kappa(t)\colon [0, +\infty) \longrightarrow \{1, 2, \ldots, j\}$ be the switching signal; then, the edge weighting $a_{mk}^{\kappa(t)}$ is positive if the edge $(n_k, n_m), n_m, n_k \in \mathcal{G}$, exists from n_k to n_m, and $a_{mk}^{\kappa(t)}$ is zero otherwise. Define the neighboring set and the indegree of the node n_m as $N_m^{\kappa(t)} = \{n_k \in \mathcal{G}\colon (n_k, n_m) \in \mathcal{E}\}$ and $d_{mm}^{\kappa(t)} = \deg(n_m) = \sum_{k \in N_m^{\kappa(t)}} a_{mk}^{\kappa(t)}$, respectively. The Laplacian matrix is defined as $L^{\kappa(t)} = [l_{mk}^{\kappa(t)}]_{N \times N}$ with $l_{mk}^{\kappa(t)} = d_{mm}^{\kappa(t)} - a_{mk}^{\kappa(t)}$. Note that the switching instants t^{s_i} $(s_i \in \mathbb{N})$ of the switching topology set should satisfy $t^{s_i+1} - t^{s_i} > T_d > 0$ with T_d the dwell time. It is assumed that there is no self-loop for all nodes. A directed path from node n_m to node n_k is a sequence of edges in the form of $\{(n_m, n_i), (n_i, n_j), \ldots, (n_s, n_k)\}$. The directed graph is said to have a spanning tree if a directed path exists from the root node to any other nodes. To address the formation tracking problems, it is supposed that each communication topology in the switching topology set has a spanning tree with the leader locating at the root node. More details for the basic concept of graph theory can be found in [34].

2.2. Design the Intermittent Formation Tracking Protocol. Consider a group of N agents, where agent 1 is the leader, and the other $N-1$ agents are the homogenous followers. The dynamics model of the leader-follower swarm system is described as follows:

$$\begin{cases} \dot{x}_1(t) = Ax_1(t), \\ \dot{x}_m(t) = Ax_m(t) + Bu_m(t), \end{cases} \tag{1}$$

where $m \in \{2, 3, \ldots, N\}$, $A \in \mathbb{R}^{p \times p}$, $B \in \mathbb{R}^{p \times q}$, $x_m(t)$ is the state, and $u_m(t)$ is the control signal of agent m. Notice that the leader lies on the root node of the spanning tree and receives no information from followers. It is supposed that the leader's control input is zero, and the communication between neighboring followers is undirected.

Since the swarm system (1) is subjected to the intermittent communications and the switching topologies, their effects should be analyzed before moving on. It is assumed that a sequence of uniformly bounded time units $[t_{2i}, t_{2i+2}) = [t_{2i}, t_{2i+1}) \cup [t_{2i+1}, t_{2i+2})$ $(i \in \mathbb{N})$ exists such that $t_{2i} = t_{2i}^1 < t_{2i}^2 < \cdots < t_{2i}^{s_i} = t_{2i+1} < t_{2i+1} + \sigma_i = t_{2i+2}$, where s_i and σ_i are integers. Let $t_0 = 0$ and $0 < \theta_{\min} \leq \theta_i = t_{2i+2} - t_{2i} \leq \theta_{\max}$. Define the communication failure rate as $\varepsilon_i = (t_{2i+2} - t_{2i+1})/(t_{2i+2} - t_{2i})$ with $0 < \varepsilon_i \leq \varepsilon_{\max} < 1$. In general, one can find that the communication channel between neighboring agents is smooth, and the communication topology may be switched in time units $[t_{2i}, t_{2i+1})$, but all the communication channels will be disappeared in time units $[t_{2i+1}, t_{2i+2})$. Hence, the communication is intermittent for the swarm system (1). Moreover, it should be pointed out that $0 < T_d \leq t_{2i}^{s_i} - t_{2i}^{s_{i-1}} \leq \overline{T}_d$, where $\overline{T}_d$ is bounded and T_d is called the minimum dwell time, and the communication topology switches at time instant $t_{2i}^{s_i}$.

Let a piecewise continuous differentiable function $h_m(t)$ $(m = 2, 3, \ldots, N)$ represent the expected time-varying formation structure formed by the followers; then, a new guaranteed cost formation tracking control protocol with the associated performance index is proposed as

$$\begin{cases} u_m(t) = \begin{cases} K_l \sum\limits_{k \in N_m^{\kappa(t)}, k \neq 1} a_{mk}^{\kappa(t)} \left(x_k(t) - h_k(t) - x_m(t) + h_m(t)\right) & t \in [t_{2i}, t_{2i+1}), \\ + K_l a_{m1}^{\kappa(t)} \left(x_1(t) - x_m(t) + h_m(t)\right), & \\ 0, & t \in [t_{2i+1}, t_{2i+2}), \end{cases} \\ J_c = \sum\limits_{i=0}^{+\infty} \left(\int_{t_{2i}}^{t_{2i+1}} \left(J_h(t) + J_u(t)\right) \mathrm{d}t + \int_{t_{2i+1}}^{t_{2i+2}} J_h(t) \mathrm{d}t \right), \end{cases} \tag{2}$$

where $m \in \{2, 3, \ldots, N\}$,

$$\begin{aligned} J_h(t) &= \sum_{m=1}^{N} \sum_{k \in N_m^{\kappa(t)}} a_{mk}^{\delta(t)} \left(x_k(t) - h_k(t) - x_m(t) + h_m(t)\right)^T Q \left(x_k(t) - h_k(t) - x_m(t) + h_m(t)\right), \\ J_u(t) &= \sum_{m=2}^{N} u_m^T(t) R u_m(t). \end{aligned} \tag{3}$$

$i \in \mathbb{N}$, $Q = Q^T > 0$, $R = R^T > 0$, and K_l is denoting the control gain matrix. J_c is the performance index representing the total of the formation regulation performance and the control energy expenditure.

For swarm systems subjected to intermittent communications and switching topologies, the definition of the guaranteed cost formation tracking control is given as follows.

Definition 1. For any given bounded initial states $x_m(0) - h_m(0)$ $(m = 2, 3, \cdots, N)$, the swarm system (1) is said to be guaranteed cost formation tracking achievable if there exists a gain matrix K_l such that $\lim_{t \longrightarrow +\infty} (x_m(t) - h_m(t) - x_1(t)) = 0$ $(m = 2, 3, \cdots, N)$ and $J_c \leq C_{\text{ost}}$, where C_{ost} is called the guaranteed cost.

The current paper mainly focuses on the guaranteed cost formation tracking design problems, in which the gain matrix is designed, and the guaranteed cost is determined. Moreover, for the given gain matrix, the guaranteed cost formation tracking analysis criterion is derived.

Remark 1. Protocol (2) consists of two parts. The first one is the intermittent control input, which is constructed by the state and formation errors between neighboring followers and those between the leader and the followers over the time units $[t_{2i}, t_{2i+1})$ and is set as zero in the time units $[t_{2i+1}, t_{2i+2})$, $i \in \mathbb{N}$. With the intermittent control input and the switching neighboring sets and edge weightings, protocol (2) is piecewise continuous, which will lead to the piecewise continuous right hand sides of the closed-loop system in the system stability analysis and is challenging to be dealt with. The second one is the performance index, which describes the total cost of the guaranteed cost formation tracking design. The weighting matrices Q and R represent the proportion of the formation regulation performance and the control energy expenditure in the performance index, respectively, which will be taken into consideration in the gain matrix design. Note that the performance index J_c is a piecewise continuous integral function due to the intermittent control input. Moreover, different from the guaranteed cost consensus, the guaranteed cost formation tracking can drive the swarm system to form an expected formation structure, while the guaranteed cost can be satisfied. Note that the expected formation structure can be time-varying and can be designed as much as required if it can satisfy the formation feasibility condition as shown in Theorem 1 in the following content.

3. Main Results

In this section, first, the formation tracking problem of the swarm system (1) is converted to the asymptotical stability problem of a reduced-order subsystem via nonsingular transformation. Then, guaranteed cost formation tracking design and analysis criteria are derived, and the guaranteed cost is determined to show the upper bound of the performance index.

Denote $\varphi_m(t) = x_m(t) - h_m(t)$, and one can obtain from (1) and (2) that

$$\dot{\varphi}_m(t) = \begin{cases} A(\varphi_m(t) + h_m(t)) + BK_l \sum\limits_{k \in N_m^{\kappa(t)}} a_{mk}^{\kappa(t)} (\varphi_k(t) - \varphi_m(t)) & t \in [t_{2i}, t_{2i+1}), \\ +BK_l a_{m1}^{\kappa(t)} (x_1(t) - \varphi_m(t)) - \dot{h}_m(t), & \\ A(\varphi_m(t) + h_m(t)) - \dot{h}_m(t), & t \in [t_{2i+1}, t_{2i+2}). \end{cases} \tag{4}$$

Since no formation is required to be formed by the leader, one can define the auxiliary variable $h_1(t) \equiv 0$ and set $\varphi_1(t) = x_1(t) - h_1(t)$. Let $\varphi(t) = [\varphi_1^T(t), \varphi_2^T(t), \cdots, \varphi_N^T(t)]^T$, $x(t) = [x_1^T(t), x_2^T(t), \ldots, x_N^T(t)]^T$, and $h(t) = [h_1^T(t), h_2^T(t), \ldots, h_N^T(t)]^T$, then equation (4) can be represented by the compact form as

$$\dot{\varphi}(t) = \begin{cases} (I_N \otimes A)(\varphi(t) + h(t)) - \left(L^{\kappa(t)} \otimes BK_l\right)\varphi(t) - \left(I_N \otimes I_p\right)\dot{h}(t), & t \in [t_{2i}, t_{2i+1}), \\ (I_N \otimes A)(\varphi(t) + h(t)) - \left(I_N \otimes I_p\right)\dot{h}(t), & t \in [t_{2i+1}, t_{2i+2}), \end{cases} \tag{5}$$

where the structure of $L^{\kappa(t)}$ is shown as follows:

$$L^{\kappa(t)} = \begin{bmatrix} 0 & 0 \\ L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)} & -\Gamma_l^{\kappa(t)} \end{bmatrix},$$
$$\Lambda_l^{\kappa(t)} = \mathrm{diag}\left\{a_{21}^{\kappa(t)}, a_{31}^{\kappa(t)}, \ldots, a_{N1}^{\kappa(t)}\right\}, \tag{6}$$
$$\Gamma_l^{\kappa(t)} = \left[a_{21}^{\kappa(t)}, a_{31}^{\kappa(t)}, \ldots, a_{N1}^{\kappa(t)}\right]^T,$$

and $L_f^{\kappa(t)}$ represents the Laplacian matrix of followers.

In the sequel, by nonsingular transformation, the closed-loop system (5) will be decomposed into two subsystems. First, define the following nonsingular matrix:

$$U^{\kappa(t)} = \begin{bmatrix} 1 & 0 \\ \mathbf{1}_{N-1} & I_{N-1} \end{bmatrix}, \tag{7}$$

such that

$$\left(\left(U^{\kappa(t)}\right)^{-1} \otimes I_p\right)\varphi(t) = \left[x_1^T(t), \tilde{\varphi}_2^T(t), \ldots, \tilde{\varphi}_N^T(t)\right]^T, \tag{8}$$

with $\tilde{\varphi}_m(t) = \varphi_m(t) - x_1(t)$, $m = 2, 3, \ldots, N$, and

$$\left(U^{\kappa(t)}\right)^{-1} L^{\kappa(t)} U^{\kappa(t)} = \begin{bmatrix} 0 & 0 \\ 0 & L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)} \end{bmatrix}, \tag{9}$$

where the fact $\Lambda_l^{\kappa(t)}\mathbf{1}_{N-1} = \Gamma_l^{\kappa(t)}$ is utilized.

Then, the block $L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)}$ is diagonalized. For each communication topology in the switching topology set, since there at least exists a spanning tree with the leader locating at the root node and the communication channels among followers are undirected and connected, the eigenvalue 0 of $L_f^{\kappa(t)}$ is simple and the block $L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)}$ is positive definite and symmetric. In this sense, there exists an orthonormal matrix $W^{\kappa(t)}$ such that

$$\left(W^{\kappa(t)}\right)^T \left(L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)}\right) W^{\kappa(t)} = D_f^{\kappa(t)} = \mathrm{diag}\left\{\lambda_2^{\kappa(t)}, \lambda_3^{\kappa(t)}, \ldots, \lambda_N^{\kappa(t)}\right\}, \tag{10}$$

where $\lambda_m^{\kappa(t)}$ $(m = 2, 3, \cdots, N)$ are the eigenvalues of $L_f^{\kappa(t)}$ with the order $0 < \lambda_2^{\kappa(t)} \le \lambda_3^{\kappa(t)} \le \cdots \le \lambda_N^{\kappa(t)}$. Denote $\tilde{\varphi}(t) = [\tilde{\varphi}_2^T(t), \tilde{\varphi}_3^T(t), \ldots, \tilde{\varphi}_N^T(t)]^T$ and $((W^{\kappa(t)})^T \otimes I_p)\tilde{\varphi}(t) = \phi(t) = [\phi_2^T(t), \phi_3^T(t), \ldots, \phi_N^T(t)]^T$; then, equation (5)fd5 is converted to the following two subdynamics:

$$\dot{x}_1(t) = Ax_1(t), \tag{11}$$

$$\dot{\phi}(t) = \begin{cases} (I_{N-1} \otimes A)\left(\phi(t) + \left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes I_p\right)h(t)\right) \\ -\left(D_f^{\kappa(t)} \otimes BK_l\right)\phi(t) - \left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes I_p\right)\dot{h}(t), & t \in [t_{2i}, t_{2i+1}), \\ (I_{N-1} \otimes A)\left(\phi(t) + \left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes I_p\right)h(t)\right) \\ -\left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes I_p\right)\dot{h}(t), & t \in [t_{2i+1}, t_{2i+2}). \end{cases} \tag{12}$$

Because $U^{\kappa(t)}$ is nonsingular and $W^{\kappa(t)}$ is orthonormal, one can derive that the swarm system (1) with control protocol (2) achieves the time-varying formation tracking if subdynamics (12) is asymptotical stable; that is, $\lim_{t \longrightarrow +\infty} \phi(t) = 0$.

Let $\lambda_{\min} = \min\{\lambda_2^s \colon \forall s \in \{1, 2, \cdots, j\}\}$ and $\lambda_{\max} = \max\{\lambda_N^s \colon \forall s \in \{1, 2, \ldots, j\}\}$; then, the following theorem gives the guaranteed cost formation tracking criterion for the swarm system (1) with protocol (2).

Theorem 1. *Swarm system (1) is guaranteed cost formation tracking achievable by protocol (2) with $K_l = \lambda_{min}^{-1}\gamma B^T P^{-1}/2$, if $\mu(1 - \varepsilon_{max}) > \varpi\varepsilon_{max} e^{\varpi\varepsilon_{max}\theta_{max}}$, $\dot{h}_m(t) = Ah_m(t)$, $m = 2, 3, \ldots, N$, and there exist $\gamma > 0$ and $P = P^T > 0$ such that*

$$\Theta_\varpi = \begin{bmatrix} AP + PA^T - \varpi P & 2\lambda_{\max}PQ \\ * & -2\lambda_{\max}Q \end{bmatrix} < 0,$$
$$\Theta_\mu = \begin{bmatrix} AP + PA^T + \mu P - \gamma BB^T & 2\lambda_{\max}PQ & \lambda_{\max}\lambda_{\min}^{-1}\gamma BR/2 \\ * & -2\lambda_{\max}Q & 0 \\ * & * & -R \end{bmatrix} < 0. \tag{13}$$

In this case, the guaranteed cost is

$$C_{\mathrm{ost}} = (x(0) - h(0))^T \left(\begin{bmatrix} N-1 & -\mathbf{1}_{N-1}^T \\ -\mathbf{1}_{N-1} & I_{N-1} \end{bmatrix} \otimes P^{-1}\right)(x(0) - h(0)). \tag{14}$$

Proof. Construct Lyapunov functional candidate as follows:

$$V(t) = \phi^T(t)\left(I_{N-1} \otimes P^{-1}\right)\phi(t). \tag{15}$$

For $t \in [t_{2i}, t_{2i+1})$, $i \in \mathbb{N}$, taking the time derivate of $V(t)$ with respect to the trajectories of equation (12) gives

$$\begin{aligned} V(t) = {} & \phi^T(t)\left(I_{N-1} \otimes \left(P^{-1}A + A^T P^{-1}\right) - D_f^{\kappa(t)} \otimes \left(P^{-1}BK_u + K_u^T B^T P^{-1}\right)\right)\phi(t) \\ & + 2\phi^T(t)\left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes P^{-1}A\right)h(t) \\ & - 2\phi^T(t)\left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes P^{-1}\right)\dot{h}(t). \end{aligned} \tag{16}$$

Define an auxiliary variable

$$\Xi(t) = \left[\left(Ah_2(t) - \dot{h}_2(t)\right)^T, \left(Ah_3(t) - \dot{h}_3(t)\right)^T, \cdots, \left(Ah_N(t) - \dot{h}_N(t)\right)^T\right]^T. \tag{17}$$

Then, one can find that $\Xi(t) = 0$, if $\dot{h}_m(t) = Ah_m(t)$, $m = 2, 3, \cdots, N$. Based on the above fact, it can be obtained that

$$V(t) = \phi^T(t)\left(I_{N-1} \otimes \left(P^{-1}A + A^T P^{-1}\right) - D_f^{\kappa(t)} \otimes \left(P^{-1}BK_u + K_u^T B^T P^{-1}\right)\right)\phi(t). \tag{18}$$

Let $K_u = \lambda_{\min}^{-1}\gamma B^T P^{-1}/2$; then, one can show that

$$\dot{V}(t) + \mu V(t) = \sum_{m=2}^{N} \phi_i^T(t)\left(P^{-1}A + A^T P^{-1} + \mu P^{-1} - \lambda_m^{\kappa(t)}\lambda_{\min}^{-1}\gamma P^{-1}BB^T P^{-1}\right)\phi_i(t). \tag{19}$$

It can be derived by pre- and postmultiplying $AP + PA^T + \mu P - \gamma BB^T < 0$ with P^{-1} that

$$P^{-1}A + A^T P^{-1} + \mu P^{-1} - \gamma P^{-1}BB^T P^{-1} < 0. \tag{20}$$

Since $\lambda_m^{\kappa(t)}\lambda_{\min}^{-1} \geq 1$, $m = 2, 3, \cdots, N$, one can deduce that

$$\dot{V}(t) < -\mu V(t). \tag{21}$$

For $t \in [t_{2i+1}, t_{2i+2})$, $i \in \mathbb{N}$, taking the time derivate of $V(t)$ along equation (12) yields

$$\begin{aligned} V(t) = {} & \phi^T(t)\left(I_{N-1} \otimes \left(P^{-1}A + A^T P^{-1}\right)\right)\phi(t) \\ & + 2\phi^T(t)\left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes P^{-1}A\right)h(t) \\ & - 2\phi^T(t)\left(\left(W^{\kappa(t)}\right)^T [0, I_{N-1}]\left(U^{\kappa(t)}\right)^{-1} \otimes P^{-1}\right)\dot{h}(t). \end{aligned} \tag{22}$$

Due to $\dot{h}_m(t) = Ah_m(t)$, $m = 2, 3, \cdots, N$, one can derive by similar analysis that

$$\dot{V}(t) - \varpi V(t) = \sum_{m=2}^{N} \phi_i^T(t)\left(P^{-1}A + A^T P^{-1} - \varpi P^{-1}\right)\phi_i(t). \tag{23}$$

Note that $P^{-1}A + A^T P^{-1} - \varpi P^{-1} < 0$ can be obtained by pre- and postmultiplying $AP + PA^T - \varpi P < 0$ with P^{-1}; then, it holds that

$$\dot{V}(t) < \varpi V(t). \tag{24}$$

For $t \in [t_0, t_2)$, i.e., $i = 0$, one has

$$V(t_2) < e^{\varpi(t_2 - t_1)}V(t_1) < e^{\varpi(t_2 - t_1)}e^{-\mu(t_1 - t_0)}V(t_0) = e^{-(\mu - (\mu + \varpi)\varepsilon_0)\theta_0}V(0). \tag{25}$$

In virtue of $\mu(1 - \varepsilon_{\max}) > \varpi\varepsilon_{\max}e^{\varpi\varepsilon_{\max}\theta_{\max}}$ and the fact that $e^{\varpi\varepsilon_{\max}\theta_{\max}} > 1$, it can be deduced that $-(\mu - (\mu + \varpi)\varepsilon_0)\theta_0 < 0$. Then, one can obtain for $\forall i \in \mathbb{N}$ that

$$V(t_{r+1}) < V(0)e^{-\sum_{d=0}^{r}(\mu - (\mu + \varpi)\varepsilon_d)\theta_d}. \tag{26}$$

Hence, one can show that for $\forall t > 0$, there exists a $\nu \in \mathbb{N}^+$ such that $t_{2k} < t \le t_{2k+2}$. In this case, one has

$$\begin{aligned} V(t) &\le e^{\varpi\theta_{\max}} V(t_\nu) \le e^{\varpi\theta_{\max}} V(0) e^{-\sum_{s=0}^{\nu-1}\varphi_s} \le e^{\varpi\theta_{\max}} V(0) e^{-\nu\left(\mu-(\mu+\varpi)\varepsilon_{\max}\right)\theta_{\min}} \\ &\le e^{\varpi\theta_{\max}} V(0) e^{-\left(\left(\left(\mu-(\mu+\varpi)\varepsilon_{\max}\right)\theta_{\min}\right)/\theta_{\max}\right)t}. \end{aligned} \tag{27}$$

From (27), it can be concluded that subdynamics (12) is asymptotical stable; that is, $\lim_{t\longrightarrow+\infty}\phi(t) = 0$. Therefore, one can find that the formation tracking can be achieved for the swarm system (1) with protocol (2).

In the next content, the guaranteed cost formation tracking achievability is discussed with the performance index J_c. From (2), one can deduce that

$$\begin{aligned} J_h(t) &= 2\widetilde{\varphi}^T(t)\left(\left(L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)}\right)\otimes Q\right)\widetilde{\varphi}(t), \\ J_u(t) &= \widetilde{\varphi}^T(t)\left(\left(L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)}\right)^2 \otimes K_l^T R K_l\right)\widetilde{\varphi}(t). \end{aligned} \tag{28}$$

Due to $\phi(t) = ((W^{\kappa(t)})^T \otimes I_p)\widetilde{\varphi}(t)$ and $K_u = \lambda_{\min}^{-1}\gamma B^T P^{-1}/2$, one can obtain that

$$J_c \le \sum_{i=0}^{+\infty}\sum_{m=2}^{N}\int_{t_{2i}}^{t_{2i+1}} \frac{1}{4}\lambda_{\max}^2\lambda_{\min}^{-2}\gamma^2\phi_m^T(t)\left(P^{-1}BRB^TP^{-1}\right)\phi_m(t)\mathrm{d}t + \sum_{i=0}^{+\infty}\sum_{m=2}^{N}\int_{t_{2i}}^{t_{2i+2}} 2\overline{\lambda}_{\max}\phi_m^T(t)Q\phi_m(t)\mathrm{d}t. \tag{29}$$

According to (19), (23), and (29), it can be derived that

$$\begin{aligned} J_c \le& \sum_{i=0}^{+\infty}\sum_{m=2}^{N}\int_{t_{2i}}^{t_{2i+1}} \frac{1}{4}\lambda_{\max}^2\lambda_{\min}^{-2}\gamma^2\phi_m^T(t)\left(P^{-1}BRB^TP^{-1}\right)\phi_m(t)\mathrm{d}t \\ &+ \sum_{i=0}^{+\infty}\sum_{m=2}^{N}\int_{t_{2i}}^{t_{2i+1}} \phi_m^T(t)\left(P^{-1}A + A^TP^{-1} + \mu P^{-1} - \lambda_m^{\kappa(t)}\lambda_{\min}^{-1}\gamma P^{-1}BB^TP^{-1}\right)\phi_m(t)\mathrm{d}t \\ &+ \sum_{i=0}^{+\infty}\sum_{m=2}^{N}\int_{t_{2i+1}}^{t_{2i+2}} \phi_m^T(t)\left(P^{-1}A + A^TP^{-1} - \varpi P^{-1}\right)\phi_m(t) \\ &+ \sum_{i=0}^{+\infty}\sum_{m=2}^{N}\int_{t_{2i}}^{t_{2i+2}} 2\lambda_{\max}\phi_m^T(t)Q\phi_m(t)\mathrm{d}t \\ &- \sum_{i=0}^{+\infty}\left(\int_{t_{2i}}^{t_{2i+1}}(\dot{V}(t) + \mu V(t))\mathrm{d}t + \int_{t_{2i+1}}^{t_{2i+2}}(\dot{V}(t) - \varpi V(t))\mathrm{d}t\right). \end{aligned} \tag{30}$$

By $\lambda_m^{\kappa(t)}\lambda_{\min}^{-1} \ge 1$, $\Theta_\varpi < 0$, $\Theta_\mu < 0$, and Schur complement, it holds as $i \longrightarrow +\infty$ that

$$J_c \le V(0) - \sum_{i=0}^{+\infty}\left(\int_{t_{2i}}^{t_{2i+1}} \mu V(t)\mathrm{d}t - \int_{t_{2i+1}}^{t_{2i+2}} \varpi V(t)\mathrm{d}t\right). \tag{31}$$

Utilizing the mean value theorem of integrals gives

$$\begin{aligned} J_c &\le V(0) - \sum_{i=0}^{+\infty}\left(\mu V(t_{2i+1})(t_{2i+1} - t_{2i}) - \varpi V(t_{2i+2})(t_{2i+2} - t_{2i+1})\right) \\ &= V(0) - \sum_{i=0}^{+\infty}\left(\mu(1-\varepsilon_i)V(t_{2i+1}) - \varpi\varepsilon_i V(t_{2i+2})\right)\theta_i \\ &\le V(0) - \sum_{i=0}^{+\infty}\left(\mu(1-\varepsilon_i) - \varpi\varepsilon_i e^{\varpi\varepsilon_i\theta_i}\right)V(t_{2i+1})\theta_i. \end{aligned} \tag{32}$$

By $\mu(1-\varepsilon_{\max}) > \varpi\varepsilon_{\max}e^{\varpi\varepsilon_{\max}\theta_{\max}}$, one can deduce that

$$J_c \le V(0) = \phi^T(0)\left(I_{N-1}\otimes P^{-1}\right)\phi(0). \tag{33}$$

Since $\phi(0) = ((W^{\kappa(t)})^T \otimes I_p)\widetilde{\varphi}(0)$, one can find that

$$\phi^T(0)\left(I_{N-1}\otimes P^{-1}\right)\phi(0) = \widetilde{\varphi}^T(0)\left(W^{\kappa(t)}\left(W^{\kappa(t)}\right)^T \otimes P^{-1}\right)\widetilde{\varphi}(0). \tag{34}$$

Then, due to $\widetilde{\varphi}(0) = ([0, I_{N-1}](U^{\kappa(t)})^{-1} \otimes I_p)\varphi(0)$, it follows from (33) that

$$C_{\text{ost}} = V(0) = \varphi^T(0)\left(\begin{bmatrix} N-1 & -\mathbf{1}_{N-1}^T \\ -\mathbf{1}_{N-1} & I_{N-1} \end{bmatrix}\otimes P^{-1}\right)\varphi(0). \tag{35}$$

This completes the proof of Theorem 1. □

Remark 2. Note that the condition $\dot{h}_m(t) = Ah_m(t)$ in Theorem 1 is called the formation feasibility condition, which indicates whether an expected formation is feasible or not to be achieved by swarm systems. It should be pointed out that not all formation can be achieved due to the

dynamic constraint of the agent. For time-varying formation, the formation function derivate $\dot{h}_m(t)$ affects the feasibility of the formation, whose constraint is shown in the condition $\dot{h}_m(t) = Ah_m(t)$. It can be found that the condition is associated with the dynamic matrix A of each agent. However, if $\dot{h}_m(t) \equiv 0$, which means that the formation is time-invariant, then the formation feasibility becomes $Ah_m = 0$, which can be found in [19].

Remark 3. Due to the jointed effect of the intermittent communication and the switching topology, the right hand side of the closed-loop system becomes piecewise continuous. Besides, from the proof of Theorem 1, it can be found that the system stability analysis is divided into two parts due to the jointed effect of the intermittent communication and the switching topology. On the one hand, for time units $[t_{2i}, t_{2i+1})$, $i \in \mathbb{N}$, it can be concluded that the Lyapunov functional candidate is decreased exponentially by a rate faster than μ. On the other hand, the value of the Lyapunov functional candidate may be increased along a rate less than ϖ in time units $[t_{2i+1}, t_{2i+2})$. By combining these two aspects of the stability analysis, it can be shown that the Lyapunov functional candidate converges with the rate $(\mu - (\mu + \varpi)\varepsilon_{\max})\theta_{\min}/\theta_{\max}$ exponentially according to the condition $\mu(1 - \varepsilon_{\max}) > \varpi\varepsilon_{\max} e^{\varpi\varepsilon_{\max}\theta_{\max}}$. Note that if the guaranteed cost performance is not considered, then the condition $\mu(1 - \varepsilon_{\max}) > \varpi\varepsilon_{\max}$ can guarantee the stability of subdynamics (12). The condition $\mu(1 - \varepsilon_{\max}) > \varpi\varepsilon_{\max} e^{\varpi\varepsilon_{\max}\theta_{\max}}$ ensures that the performance index J_c can be upper bounded by the guaranteed cost C_{ost}. Generally speaking, the condition $\mu(1 - \varepsilon_{\max}) > \varpi\varepsilon_{\max} e^{\varpi\varepsilon_{\max}\theta_{\max}}$ can always guarantee $\mu(1 - \varepsilon_{\max}) > \varpi\varepsilon_{\max}$ since $e^{\varpi\varepsilon_{\max}\theta_{\max}} > 1$ for positive ϖ, $\varepsilon_{\max}$, and $\theta_{\max}$.

Theorem 1 provides the criterion of the guaranteed cost formation tracking design where the gain matrix K_l is determined. However, if K_l is given, then it is interesting to analyze whether K_l is feasible to solve the guaranteed cost formation tracking problems. Set $\overline{P} = P^{-1}$ and use the convex property of linear matrix inequalities, then the following theorem gives the sufficient conditions of the guaranteed cost formation tracking analysis for given K_l.

Theorem 2. *For any given K_l, the swarm system (1) with protocol (2) achieves guaranteed cost formation tracking if $K_u = \lambda_{\min}^{-1}\gamma B^T P^{-1}/2$ $\dot{h}_m(t) = Ah_m(t)$, $m = 2, 3, \cdots, N$, and there exists a matrix $\overline{P} = \overline{P}^T > 0$ such that*

$$\begin{gathered}
\overline{P}A + A^T\overline{P} - \varpi\overline{P} + 2\lambda_{\max}Q < 0, \\
\begin{bmatrix} \overline{P}A + A^T\overline{P} + \mu\overline{P} - \lambda_{\min}\left(\overline{P}BK_l + K_l^T B^T\overline{P}\right) & 2\lambda_{\min}Q & \lambda_{\min}K_l^T R \\ * & -2\lambda_{\min}Q & 0 \\ * & * & -R \end{bmatrix} < 0, \\
\begin{bmatrix} \overline{P}A + A^T\overline{P} + \mu\overline{P} - \lambda_{\max}\left(\overline{P}BK_l + K_l^T B^T\overline{P}\right) & 2\lambda_{\max}Q & \lambda_{\max}K_l^T R \\ * & -2\lambda_{\max}Q & 0 \\ * & * & -R \end{bmatrix} < 0.
\end{gathered} \tag{36}$$

In this case, the guaranteed cost is

$$C_{\text{ost}} = (x(0) - h(0))^T \left(\begin{bmatrix} N-1 & -\mathbf{1}_{N-1}^T \\ -\mathbf{1}_{N-1} & I_{N-1} \end{bmatrix} \otimes P^{-1} \right) (x(0) - h(0)). \tag{37}$$

Remark 4. The formation design in [20–23] only took care about how to design a proper gain matrix such that the expected formation can be achieved, but they did not consider the guaranteed cost performance when designing the formation control protocol. In contrast, the current paper constructs a performance index to describe the total cost, where the weighting matrices between the formation regulation performance and the control energy expenditure are denoted by Q and R. In this case, weighting matrices Q and R are introduced into the design procedure of the gain matrix, which can assure that not only the formation tracking can be achieved but also the performance index can be constrained by the guaranteed cost. By adjusting the relative value of Q and R, the compromise design between the formation regulation performance and the control energy expenditure can be achieved. Moreover, the guaranteed costs obtained in Theorems 1 and 2 are associated with the initial states and formations and the interaction matrix. Note that the initial states and formations are often available in applications and the interaction matrix is related to a time-invariant star graph, which can be obtained when the number of agents is determined. Furthermore, in the gain matrix design, the eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$ are needed, which is difficult to be calculated. Fortunately, $\lambda_{\min}$ can be obtained via the method in [35], and $\lambda_{\max}$ can be estimated by Gersgorin's disc theorem in [36].

Remark 5. Swarm system with the leaderless structure describes the dynamics of each agent, where each agent plays the equal role of the collaborative behavior. However, the swarm system with the leader-following structure describes the dynamics of the leader with no control input and that of the follower. Different from the formation design of

leaderless swarm systems, the guaranteed cost formation tracking problem of leader-follower swarm systems owns two interesting features. First, although the communication topology among followers is undirected and connected, the Laplacian matrix of the whole system is asymmetric due to the existence of the leader. In this sense, a nonsingular transformation and an orthonormal transformation are adopted successively to diagonalize the block $L_f^{\kappa(t)} + \Lambda_l^{\kappa(t)}$ of the Laplacian matrix such that the dynamics of the closed-loop system can be linearly decoupled to solve the guaranteed cost formation tracking problem. Second, the guaranteed cost is associated with the Laplacian matrix $\begin{bmatrix} N-1 & -\mathbf{1}_{N-1}^T \\ -\mathbf{1}_{N-1} & I_{N-1} \end{bmatrix}$ of a star graph with the leader locating at the center, which indicates that the global interaction mechanism of the whole swarm system is determined by the leader for the guaranteed cost formation tracking problem. Besides, the formation tracking movement is fully determined by the state response leader.

4. Numerical Simulation

In this section, a simulation is given to illustrate the effectiveness of the proposed guaranteed cost time-varying formation tracking design method in the above sections.

The third-order swarm system considered in the simulation is composed with one leader labeled by 1 and five followers labeled from 2 to 6 whose dynamics is modeled as follows:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & -1 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}. \tag{38}$$

The switching topology set of the swarm system is shown in Figure 1, where the topology is switched among topologies $\mathcal{G}_a^1$, $\mathcal{G}_a^2$, $\mathcal{G}_a^3$, and $\mathcal{G}_a^4$ with the dwell time $T_d = 0.3$ s in the connected communication time units $t \in [0.6i, 0.6i + 0.51)$ s, $i \in \mathbb{N}$, and the communication among all agents is interrupted in the disconnected communication time units $t \in [0.6i + 0.51, 0.6(i+1))$ s. In this case, the maximum communication failure rate is $\varepsilon_{\max} = 0.15$. The initial states of the whole swarm system are given as follows:

$$\begin{aligned} x_1(0) &= [3.5, -2.7, 1.5]^T, \\ x_2(0) &= [3.5, 5.2, -1.3]^T, \\ x_3(0) &= [4.2, -2.5, 2.3]^T, \\ x_4(0) &= [-3.1, -2.5, 4.6]^T, \\ x_5(0) &= [-2.1, 4.8, -1.5]^T, \\ x_6(0) &= [-6.2, 2.4, -3.5]^T. \end{aligned} \tag{39}$$

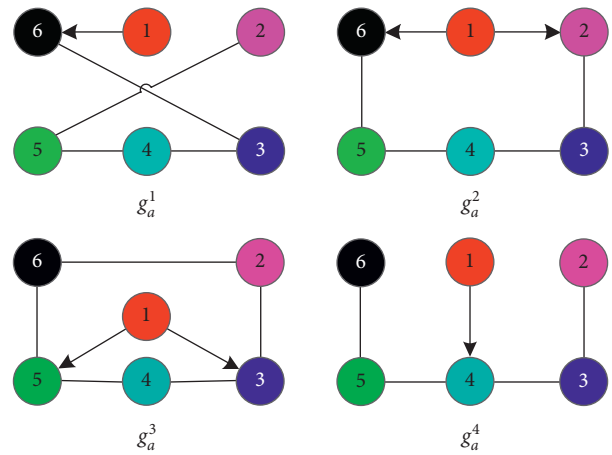


Figure 1: Switching topology set G_a.

The expected time-varying formation function is chosen as follows:

$$h_m(t) = \begin{bmatrix} \sin\left(t + \dfrac{2(m-1)\pi}{5}\right) \\ \cos\left(t + \dfrac{2(m-1)\pi}{5}\right) \\ -\sin\left(t + \dfrac{2(m-1)\pi}{5}\right) \end{bmatrix}, \quad m = 2, 3, \ldots, 6. \tag{40}$$

According to the above form of $h_m(t)$, it can be found that five followers should shape into a regular pentagon and keep rotating around its center. Meanwhile, the conditions $\dot{h}_m(t) = Ah_m(t)$, $(m = 2, 3, \cdots, 6)$ are satisfied. Set $\mu = 0.9$, $\varpi = 5$, $R = 0.1$, and $Q = \mathrm{diag}\{0.3, 0.1, 0.2\}$. By Theorem 1, it can be calculated by the FEASP solver in MATLAB that $\gamma = 0.0072$ and

$$P = \begin{bmatrix} 0.1002 & -0.0646 & 0.0442 \\ -0.0646 & 0.0508 & -0.0335 \\ 0.0442 & -0.0335 & 0.0570 \end{bmatrix}. \tag{41}$$

In this case, the guaranteed cost is determined as $C_{\mathrm{ost}} = 11171.4191$, and the gain matrix is design as

$$K_l = (7.7145, 13.5458, 4.3332). \tag{42}$$

Figure 2 depicts the error trajectory between the state and formation of each follower and the leader within 15 s, where the trajectories of followers are full curves with different colors and that of the leader is a red imaginary line. One can see from Figure 2 that $\varphi_m(t)$ $(m = 2, 3, \cdots, 6)$ of five followers converge to the same value which equals to $\varphi_1(t)$ of the leader, which means that the error state $\varphi_m(t)$ of five followers achieve consensus and track to that of the leader.

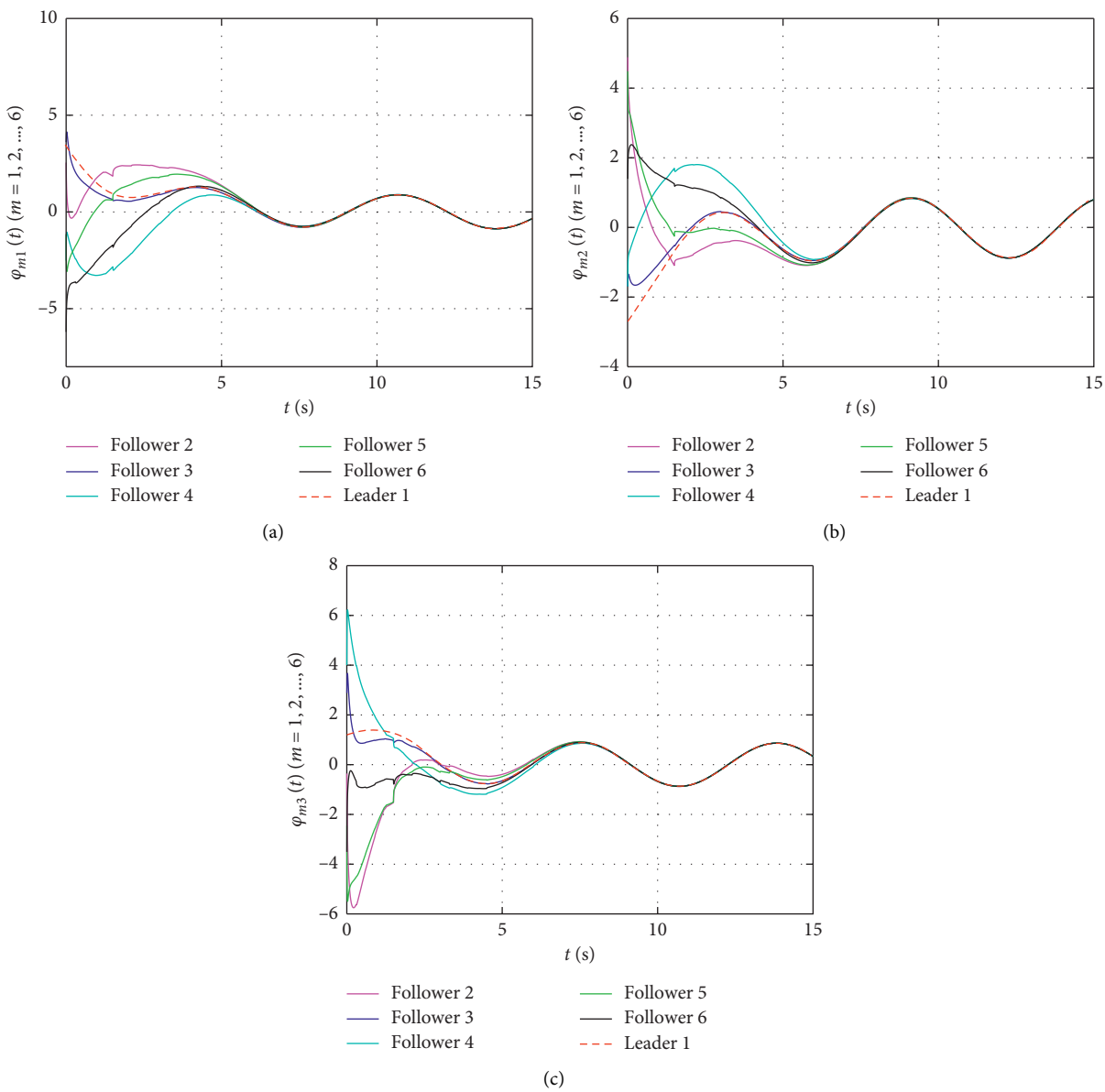


Figure 2: Curves of $\varphi_m(t)$ $(m = 1, 2, \ldots, 6)$. (a) $\varphi_{m1}(t)$. (b) $\varphi_{m2}(t)$. (c) $\varphi_{m3}(t)$.

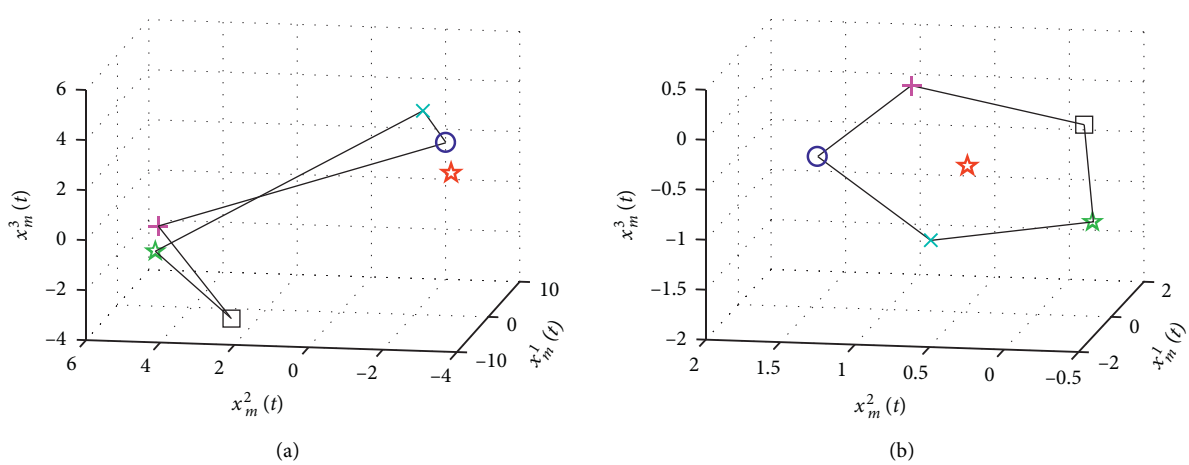


Figure 3: Continued.

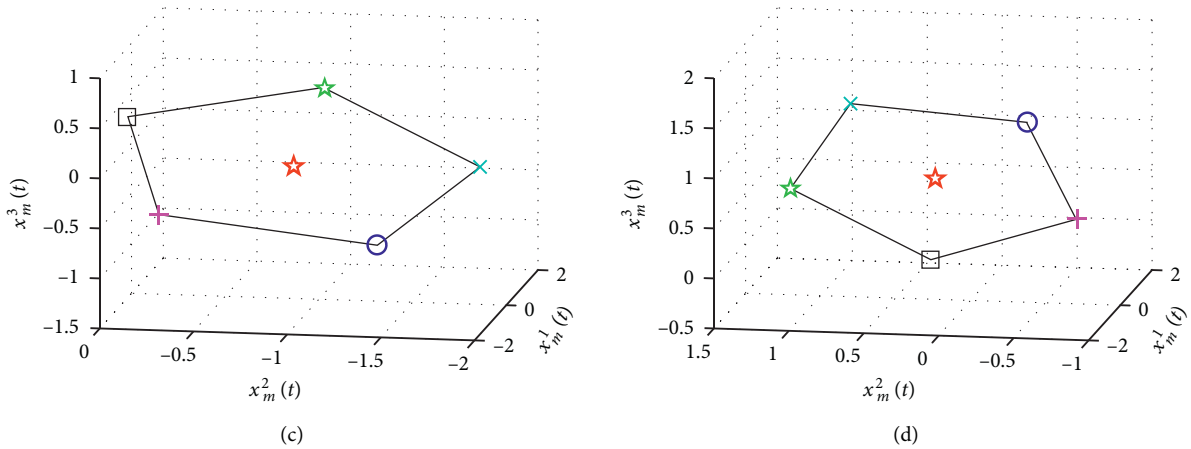


Figure 3: State of five followers and the leader at different moments. (a) $t = 0$ s. (b) $t = 10$ s. (c) $t = 12$ s. (d) $t = 14$ s.

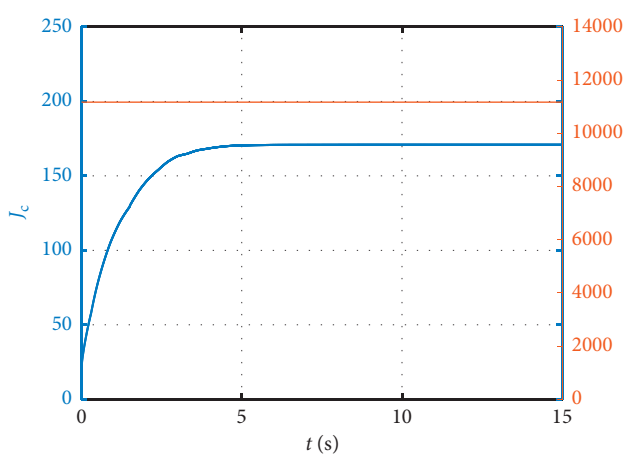


Figure 4: Curves of the performance index and the guaranteed cost.

The state snapshots of five followers and the leader are shown in Figure 3, where the state of the leader is described as the red pentacle and those of the five followers are depicted as pink pluses, blue circles, bluish x-marks, green pentacles, and black squares, orderly. From Figures 3(a)-3(b), it can be found that the formation of five followers is achieved with the geometrical shape of the regular pentagon, and the state of the leader locates at the center of the regular pentagon. From Figures 3(b)–3(d), one can see that the formation of five followers keeps rotating around the leader; that is, the time-varying formation tracking is achieved.

Figure 4 describes the curves of the performance index and the guaranteed cost, respectively. It can be shown that the value of the performance index increases to a finite value that is less than the guaranteed cost, i.e., $J_c \leq C_{\text{ost}}$.

From the simulation results in Figures 2–4, it can be concluded that the swarm system (1) with intermittent communications and switching topologies is guaranteed cost time-varying formation tracking achievable by protocol (2).

5. Conclusions

Guaranteed cost time-varying formation tracking design and analysis problems were studied for the swarm system with intermittent communications and switching topologies. An intermittent guaranteed cost formation tracking control protocol was constructed, which consisted of an intermittent control input and a performance index. It was shown that by designing the gain matrix of the control protocol, the time-varying formation tracking was achieved, while the certain performance was satisfied, where the upper bound of the performance index was restrained by determining the guaranteed cost. By adjusting the weighting matrices of the performance index, the compromised design between the control energy expenditure and the formation regulation performance was achieved. Sufficient conditions of the guaranteed cost time-varying formation design and analysis were given, and the guaranteed cost was determined. It was proven that if the formation and the communication failure rate satisfy the corresponding conditions in Theorem 1, then the high-order swarm system with intermittent communications and switching topologies can achieve the guaranteed cost time-varying formation tracking by designing the gain matrix of the formation control protocol. The further works will extend the main results of this paper from the switching connected topologies to the jointly switching topologies, and the communication among followers can be directed.

Data Availability

The data used to support this study are included within this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Purui Zhang and Xiaoqian Chen were involved in conceptualization; Purui Zhang and Xiaogang Yang were involved in methodology; Purui Zhang was involved in

validation, formal analysis, investigation, and writing original draft preparation; Xiaoqian Chen and Xiaogang Yang were involved in writing the review and editing and funding acquisition; Xiaoqian Chen was involved in supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61867005, 61763040, and 61703411).

References

[1] H. Weimerskirch, J. Martin, Y. Clerquin, P. Alexandre, and S. Jiraskova, "Energy saving in flight formation," *Nature*, vol. 413, no. 18, pp. 697-698, 2001.

[2] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.

[3] L. Consolini, F. Morbidi, D. Prattichizzo, and M. Tosques, "Leader-follower formation control of nonholonomic mobile robots with input constraints," *Automatica*, vol. 44, no. 5, pp. 1343–1349, 2008.

[4] J. Qu, Z. Ji, C. Lin, and H. Yu, "Fast consensus seeking on networkswith antagonistic interactions," *Complexity*, vol. 2018, Article ID 7831317, 15 pages, 2018.

[5] N. Cai, M. He, Q. Wu, and M. J. Khan, "On almost controllability of dynamical complex networks with noises," *Journal of Systems Science and Complexity*, vol. 32, no. 4, pp. 1125–1139, 2019.

[6] F. Li, Y. Ding, M. Zhou, K. Hao, and L. Chen, "An affection-based dynamic leader selection model for formation control in multirobot systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1217–1228, 2017.

[7] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 7, pp. 2442–2454, 2020.

[8] W. Ren, "Consensus strategies for cooperative control of vehicle formations," *IET Control Theory & Applications*, vol. 1, no. 2, pp. 505–512, 2007.

[9] R. Rahimi, F. Abdollahi, and K. Naqshi, "Time-varying formation control of a collaborative heterogeneous multi agent system," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1799–1805, 2014.

[10] J. Xi, M. He, H. Liu, and J. Zheng, "Admissible output consensualization control for singular multi-agent systems with time delays," *Journal of the Franklin Institute*, vol. 353, no. 16, pp. 4074–4090, 2016.

[11] B. Cheng and Z. Li, "Fully distributed event-triggered protocols for linear multiagent networks," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1655–1662, 2019.

[12] Y. Zhang, J. Sun, H. Liang, and H. Li, "Event-triggered adaptive tracking control for multi-agent systems with unknown disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 3, pp. 890–901, 2020.

[13] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4585–4598, 2020.

[14] M. Razaq, M. Rehan, C. Ahn, A. Khan, and M. Tufail, "Consensus of one-sided Lipschitz multiagents under switching topologies," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 2019, to be published.

[15] A. Abdessameud and A. Tayebi, "Formation control of VTOL unmanned aerial vehicles with communication delays," *Automatica*, vol. 47, no. 11, pp. 2383–2394, 2011.

[16] W. Qin, Z. Liu, and Z. Chen, "A novel observer-based formation for nonlinear multi-agent systems with time delay and intermittent communication," *Nonlinear Dynamics*, vol. 79, no. 3, pp. 1651–1664, 2015.

[17] M. Jafarian, E. Vos, C. De Persis, J. Scherpen, and A. van der Schaft, "Disturbance rejection in formation keeping control of nonholonomic wheeled robots," *International Journal of Robust and Nonlinear Control*, vol. 26, no. 15, pp. 3344–3362, 2016.

[18] H. Du, G. Wen, Y. Cheng, Y. He, and R. Jia, "Distributed finite-time cooperative control of multiple high-order nonholonomic mobile robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2998–3006, 2017.

[19] H. Liu, T. Ma, F. L. Lewis, and Y. Wan, "Robust formation control for multiple quadrotors with nonlinearities and disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1362–1371, 2020.

[20] X. Dong, Y. Zhou, Z. Ren, and Y. Zhong, "Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5014–5024, 2017.

[21] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[22] X. Dong and G. Hu, "Time-varying formation tracking for linear multiagent systems with multiple leaders," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3658–3664, 2017.

[23] W. Jiang, G. Wen, Z. Peng, T. Huang, and A. Rahmani, "Fully distributed formation-containment control of heterogeneous linear multiagent systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 9, pp. 3889–3896, 2019.

[24] Z. Yu, H. Jiang, C. Hu, and X. Fan, "Consensus of second-order multi-agent systems with delayed nonlinear dynamics and aperiodically intermittent communications," *International Journal of Control*, vol. 90, no. 5, pp. 909–922, 2016.

[25] M. Fattahi and A. Afshar, "Distributed consensus of multi-agent systems with fault in transmission of control input and time-varying delays," *Neurocomputing*, vol. 189, pp. 11–24, 2016.

[26] J. Sun and Z. Wang, "Consensus of multi-agent systems with intermittent communications via sampling time unit approach," *Neurocomputing*, vol. 397, pp. 149–159, 2020.

[27] W. Ni and D. Cheng, "Leader-following consensus of multi-agent systems under fixed and switching topologies," *Systems & Control Letters*, vol. 59, no. 3-4, pp. 209–217, 2010.

[28] J. Shao, W. X. Zheng, T.-Z. Huang, and A. N. Bishop, "On leader-follower consensus with switching topologies: an analysis inspired by pigeon hierarchies," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3588–3593, 2018.

[29] R. Wang, "Adaptive output-feedback time-varying formation tracking control for multi-agent systems with switching directed networks," *Journal of the Franklin Institute*, vol. 357, no. 1, pp. 551–568, 2020.

[30] Y. Cao and W. Ren, "Optimal linear-consensus algorithms: an LQR perspective," *IEEE Transactions on Systems, Man, and Cybernetics-PartB Cybernetics*, vol. 40, no. 3, pp. 819–829, 2010.

[31] J. Xi, Z. Fan, H. Liu, and T. Zheng, "Guaranteed-cost consensus for multiagent networks with Lipschitz nonlinear dynamics and switching topologies," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 7, pp. 2841–2852, 2018.

[32] J. B. Rejeb, I.-C. Morărescu, and J. Daafouz, "Control design with guaranteed cost for synchronization in networks of linear singularly perturbed systems," *Automatica*, vol. 91, pp. 89–97, 2018.

[33] J. Xi, C. Wang, H. Liu, and L. Wang, "Completely distributed guaranteed-performance consensualization for high-order multiagent systems with switching topologies," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1338–1348, 2019.

[34] C. Godsil and G. Royal, *Algebraic Graph Theory*, Springer-Verlag, Berlin, Germany, 2001.

[35] A. Berman and X. Zhang, "Lower bounds for the eigenvalues of Laplacian matrices," *Linear Algebra and Its Applications*, vol. 316, no. 1–3, pp. 13–20, 2000.

[36] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1990.

Hindawi
Complexity
Volume 2020, Article ID 8867820, 15 pages
https://doi.org/10.1155/2020/8867820

Research Article

Energy-Limited Time-Varying Formation Control for Second-Order Multiagent Systems

Wanzhen Quan,[1] Yulong Zhao,[2] Le Wang,[1] and Xiaogang Yang[1]

[1]*High-Tech Institute of Xi'an, Xi'an 710025, China*
[2]*College of Information and Communication, University of National Defense Technology, Wuhan 430010, China*

Correspondence should be addressed to Xiaogang Yang; doctoryxg@163.com

Received 20 August 2020; Revised 12 September 2020; Accepted 8 October 2020; Published 12 November 2020

Academic Editor: Ning Cai

The energy-limited time-varying formation (ETVF) control problem of second-order multiagent systems (MAS) is addressed for both leaderless and leader-following communication topologies in this paper. Different from the previous results, the joint consideration of energy limitation and formation design is more challenging and practical. First, an ETVF control protocol is presented, and the total energy supply is pregiven and limited, which is more common in practical applications. Then, by an orthogonal transformation, the formation control problem is converted into the consensus stabilization problem for second-order leaderless MAS, where sufficient conditions for the ETVF are derived by joint design of control gains and the total energy. At the same time, the explicit formula that forms the formation center function is obtained to depict the macroscopic movement of the multiagent system as a whole. Moreover, the proposed method is also extended to the leader-following communication structure. Finally, two examples are given to verify the effectiveness of our theoretical results.

1. Introduction

Recently, the topic of cooperative control of MAS has been paid increasing attention due to its wide application in many fields, such as sensor network synchronization [1–3], formation control [4–7], flocking [8, 9], and so on. Cooperative control of MAS originated from bioscience, where many researchers aroused great interest in birds, bees, and fishes in nature. Many scholars found that the above cooperative behavior among agents is based on the swarm consensus [10–17], and all agents could perform tasks that cannot be achieved by a single one as a formation structure. The formation control, as a fundamental and key issue of cooperative control, refers to a team of interacting agents to perform complex tasks together in a certain geometric structure, which is applied to many fields, such as networking system of satellite clusters, movements of mobile robotics [18, 19], cooperative attack of multiple missiles [20], and unmanned aerial vehicle formations [21, 22].

It is well known that the formation can be divided into the time-invariant and time-varying formation according to variation of formation structures. Fabris et al. [23] investigated the time-invariant formation tracking problem for second-order MAS, and a special potential function was designed to obtain a specified geometric structure, where the position and velocity of second-order MAS followed their expected values. Based on the virtual structure and consensus method, a formation control strategy was proposed in [24] to analyze nonholonomic intelligent vehicles. However, these methods are only suitable for the fixed formation structure. In practice, the application scenario is dynamic, and the formation of MAS needs to vary over time, such as obstacle avoidance. Accordingly, it is necessary to investigate the problem of time-varying formation control. In [25], an observer-based protocol was proposed to obtain the distributed time-varying group formation under the directed communication topology for MAS. Dong et al. [26] focused on the time-varying formation tracking problem with switching communication topologies for second-order

MAS, and the effectiveness of the proposed theoretical results were proved by the multiquadrotor unmanned aerial vehicle system.

Besides, there are two types of formation, called leaderless formation and leader-following formation, depending on whether or not the communication topology among agents has a specified leader. A finite-time cooperative controller was proposed for multiple leaderless nonholonomic mobile robots via the finite-time control technique and the model features [27]. Jia et al. [28] investigated the three-dimensional leaderless flocking problem and proposed a simplified distributed control algorithm, which was verified by the LaSalle–Krasovskii invariance principle. Furthermore, the leader-following formation aroused much attention in [29–33], where followers tracked the leader and maintained a specified formation with it. Considering the condition of uncertain factors and time delay, the leader-follower collaborative formation was achieved for the multi-unmanned underwater vehicle formation in [32]. The time-varying protocol of second-order leader-following MAS was obtained in [33] through solving an algebraic Riccati equation.

Despite the numerous results mentioned above, there is still much uncompleted interesting work when considering the energy limitation problem in practical engineering applications. For example, when unmanned aerial vehicles execute some flight tasks of agricultural seeding, forest fire prevention, and cooperative reconnaissance, their fuel or battery power is limited, and it is difficult to further recharge energy during the flight. Therefore, it is necessary not only to achieve formation control but also to consider limited energy. Under the circumstance, the energy-limited formation control problem is converted to the optimal/suboptimal problem to guarantee some performance index. However, the time-varying formation design for second-order MAS was addressed in [34, 35], where the problem of energy limitation was not considered. Note that the optimizing consensus was investigated in [36, 37]. An optimal consensus control for multiagent linear systems was investigated in the cases of the all-to-all and general communication [38], and the result illustrated that the proposed controller made the state of the closed-loop system converge exponentially. In [36–38], optimizing control was just the optimization of consensus, but the formation control problem of energy limitation was not taken into consideration. In practice, the total energy of MAS is usually constrained, and it is significant to study the formation control problem of energy limitation for MAS, which is an important basis for our investigation of this paper.

Motivated by the aforementioned facts and challenges, the problems of time-varying formation for MAS with limited energy in leaderless and leader-following cases are studied in this paper, respectively. First, an ETVF control protocol for second-order MAS is proposed for the leaderless communication topology. In addition, through orthogonal transformation, the dynamics of the multiagent system can be resolved into two independent parts: the consistency component and the inconsistent component. The former determines the whole macroscopic motion, and the microscopic motion is obtained by the latter. Second, sufficient conditions with analytic solutions of control gains and a formation center function are obtained in the leaderless case, in which the formation center function refers to macroscopic movements of the whole multiagents. Third, the research on the leaderless case is extended to the leader-following one. It is worth mentioning that a two-step method to transform the leader-follower into the leaderless framework is proposed.

The main contributions of this paper are threefold. Firstly, the problem of limited energy supply is taken into consideration based on practical engineering, where a joint design strategy combining time-varying formation control and energy constraints is investigated. In this case, the desired formation structure not only must be formed but also the practical energy consumption of the time-varying formation is less than the given energy, where the practical energy is a performance index function composed of the quadratic form of the control protocol. However, the formation control problem in [32–34] did not consider energy limitation. Secondly, in the process of solving the control gain, the numerical solution solved by Matlab's LMI Toolbox is replaced by a feasible analytical solution, so it is not needed to verify the feasibility of the solution. In [32–34], the control gains were obtained by the LMI Toolbox instead of the analytical solution. Thirdly, the critical contribution is that a two-step transformation method is presented so as to transform the whole Laplacian matrix into the dimension-lowering diagonal matrix; in this case, the leaderless and leader-following communication topologies are unified into the identical multiagent system framework.

This paper is organized as follows. The communication topology modeling and the problem description are introduced in Section 2. In Section 3, the stability of ETVF for second-order MAS is analyzed in the leaderless case, and the formation center function is determined. Section 4 extends the ETVF for the leaderless case to the leader-following one. Section 5 verifies the theoretical results by two numerical simulation examples. A brief conclusion is drawn in Section 6.

Notations. Let the symbol $\mathbb{R}$ denote a real constant. The matrix I_2 is an identity matrix of dimension 2. 1_N stands for the N-dimensional column vector with all elements being 1. 0 represents the zero matrix or zero vector with all elements equal to 0. Let the symbol $\otimes$ denote the Kronecker product.

2. Problem Description

2.1. Preliminaries: Modeling Communication Topology. For the second-order MAS with N homogeneous agents, the communication topology among them is described by an undirected weighted graph $G = (V(G), E(G))$, where $V(G) = \{v_1, v_2, \ldots, v_N\}$ denotes the set of nodes, $E(G) \subseteq V(G) \times V(G)$ is the set of edges with $E(G) = \{e_{ij} = (v_i, v_j)\}$ being a communication link from the i-th agent to the jth agent, and edge (v_i, v_i) is called the self-loop of vertex v_i. The self-loops of topology in this paper are not considered, and we assume that the topology has no self-

loops. The index of the neighbor set of agent i is denoted as $N_i = \{v_j\colon (v_j, v_i) \in E(G)\}$, and the weight adjacency matrix is defined as $W(G) = [w_{ij}] \in \mathbb{R}^{N\times N}$, where the weight $w_{ij} > 0$ if $(v_j, v_i) \in E(G)$, and $w_{ij} = 0$ otherwise. Let the degree matrix be $D(G) = \mathrm{diag}\{d_1, d_2, \ldots, d_N\}$, where the in-degree matrix of agent i is defined by $d_i = \sum_{j=1}^{N} w_{ij}$. The Laplacian matrix L is defined as $L = D(G) - W(G)$. More details about the graph theory can be found in [39].

2.2. Designing Energy-Limited Formation Control Protocol. Consider a group of second-order homogeneous agents. The dynamics of each agent is modeled by

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = u_i(t), \end{cases} \tag{1}$$

where $i = 1, 2, \ldots, N$, $x_i(t) \in \mathbb{R}$ and $v_i(t) \in \mathbb{R}$ are the position and the velocity terms of the i-th agent, respectively, and $u_i(t) \in \mathbb{R}$ is the control input. The desired formation is designed by the time-varying formation function $f(t) = [f_1^T(t), f_2^T(t), \ldots, f_N^T(t)]^T$, where $f_i(t) = [f_{ix}(t), f_{iv}(t)]^T$ is piecewise continuous differentiable with $f_{ix}(t)$ representing the position term and $f_{iv}(t)$ denotes the velocity term.

For multiagent system (1), an energy-limited formation control protocol is put forward as follows:

$$\begin{cases} u_i(t) = \sum\limits_{j\in N_i} w_{ij}\Big(k_1\big(x_j(t) - f_{jx}(t) - x_i(t) + f_{ix}(t)\big) + k_2\big(v_j(t) - f_{jv}(t) - v_i(t) + f_{iv}(t)\big)\Big), \\ J_u = \sum\limits_{i=1}^{N} \int_0^{+\infty} \kappa u_i^2(t)\mathrm{d}t, \end{cases} \tag{2}$$

where $i = 1, 2, \ldots, N$ and κ is a positive constant (it is a weight coefficient for performance index function of system (1), and we can effectively inhibit the amplitude growth of the control quantity $u(t)$ by increasing κ). k_1 and k_2 are control gains with $k_1, k_2 > 0$, and J_u is an energy index function, which refers to the control energy consumption. Let J_u^* be the total energy supply of second-order MAS; then, the definition of the energy-limited formation for second-order MAS is given below.

Definition 1. For any given $J_u^* > 0$ and nonidentical initial states $x_i(0)$ and $v_i(0)$ $(i = 1, 2, \ldots, N)$, multiagent system (1) with formation control protocol (2) is said to be energy-limited formation achievable if there exist control gains k_1 and k_2 such that $\lim_{t\longrightarrow+\infty}(x_i(t) - f_{ix}(t) - c_x(t)) = 0$, $\lim_{t\longrightarrow+\infty}(v_i(t) - f_{iv}(t) - c_v(t)) = 0$, and $J_u \le J_u^*$, where $c_x(t)$ and $c_v(t)$ correspond to the position and velocity terms of the formation center function $c(t) = [c_x(t), c_v(t)]^T$, respectively.

The fundamental purpose of this paper lies in the following two major aspects: (i) how to design control gains k_1 and k_2 such that multiagent system (1) with formation control protocol (2) can achieve the desired formation in which the total energy supply is limited and (ii) how to obtain the explicit formula of the formation center function.

Remark 1. It is important to emphasize that controller (2) looks so complicated, which shows the essence of the state feedback control. Firstly, controller (2) about $u(t)$ is the distributed consensus control protocol, which mainly consists of two parts, one is $x_j(t) - f_{jx}(t) - x_i(t) + f_{ix}(t)$ denoting the tracking error of position, and another is the tracking error of velocity $v_j(t) - f_{jv}(t) - v_i(t) + f_{iv}(t)$. Secondly, controller (2) about J_u refers to the cost function that is the integral of $u(t)$ with respect to time t, which reflects the practical energy consumption for the whole system. The goal of consensus protocol (2) is to guarantee that $x_j(t) - f_{jx}(t) - x_i(t) + f_{ix}(t) \longrightarrow 0$ and $v_j(t) - f_{jv}(t) - v_i(t) + f_{iv}(t) \longrightarrow 0$ as $t \longrightarrow \infty$; at the same time, the practical energy consumption is less than the total energy supply budget. And it is important to emphasize that controller (2) looks so complicated, but the states are knowable and easy to implement.

Remark 2. The formation control protocol (2) has two important characteristics. On the one hand, it contains an energy index function which is significant for practical engineering applications due to limited resources, such as fuel supply and battery power in practical engineering applications. Under this condition, it is crucial to propose a new idea that joins design between control gains k_1, k_2 and energy constraints. In other words, the control gains k_1 and k_2 are designed to maintain the desired formation and ensure that energy consumption of second-order MAS is within its total energy supply budget. On the other hand, the problem of the time-varying formation control is given, and a time-varying formation function is required to introduce into the control protocol (2), which makes the ETVF control problem for second-order MAS more challenging and practical than time-invariant ones, such as avoiding dynamic obstacles in time. And one can find that the time-varying formation not only forms a certain structure but also moves in a desired geometric formation according to the requirement of the realistic environments.

3. ETVF Control for the Leaderless Case

This section presents sufficient conditions for multiagent system (1) with formation control protocol (2) to achieve the ETVF. Then, an explicit formula of the formation center function is presented.

Let $\delta_{ix}(t) = x_i(t) - f_{ix}(t)$ and $\delta_{iv}(t) = v_i(t) - f_{iv}(t)$; then, the dynamics of multiagent system (1) with formation control protocol (2) can be depicted as

$$\begin{cases} \dot{\delta}_{ix}(t) = \delta_{iv}(t) + f_{iv}(t) - \dot{f}_{ix}(t), \\ \dot{\delta}_{iv}(t) = k_1 \sum_{j\in N_i} w_{ij}\left(\delta_{jx}(t) - \delta_{ix}(t)\right) + k_2 \sum_{j\in N_i} w_{ij}\left(\delta_{jv}(t) - \delta_{iv}(t)\right) - \dot{f}_{iv}(t). \end{cases} \tag{3}$$

Let $\delta_i(t) = [\delta_{ix}(t), \delta_{iv}(t)]^T$ and $\delta(t) = [\delta_1^T(t), \delta_2^T(t), \ldots, \delta_N^T(t)]^T$; then, multiagent system (3) can be rewritten in the Kronecker form as follows:

$$\begin{aligned} \dot{\delta}(t) = &\left(L \otimes \begin{bmatrix} 0 & 0 \\ -k_1 & -k_2 \end{bmatrix} + I_N \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)\delta(t) \\ &+ \left(I_N \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) f(t) - \dot{f}(t). \end{aligned} \tag{4}$$

Since the communication topology is undirected and connected, the corresponding Laplacian matrix is symmetric and has one zero eigenvalue. Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of the Laplacian matrix L with $0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_N$; then, there exists an orthonormal matrix $Q = [1_N/\sqrt{N}, \bar{Q}]$ such that $Q^T L Q = \Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$. Let $\tilde{\delta}(t) = (Q^T \otimes I_2)\delta(t) = [\tilde{\delta}_1^T(t), \tilde{\delta}_2^T(t), \ldots, \tilde{\delta}_N^T(t)]^T$, and then multiagent system (4) is transformed into

$$\dot{\tilde{\delta}}_1(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\tilde{\delta}_1(t) + \left(\frac{1_N^T}{\sqrt{N}} \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) f(t) - \left(\frac{1_N^T}{\sqrt{N}} \otimes I_2\right)\dot{f}(t), \tag{5}$$

$$\dot{\tilde{\delta}}_i(t) = \begin{bmatrix} 0 & 1 \\ -\lambda_i k_1 & -\lambda_i k_2 \end{bmatrix}\tilde{\delta}_i(t) + \left(e_i^T Q^T \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) f(t) - \left(e_i^T Q^T \otimes I_2\right)\dot{f}(t), \tag{6}$$

where $(i \in \{2, 3, \ldots, N)$ and e_i is an N-dimensional column with the i-th element being 1 and the others being 0.

In the following theorem, a design approach of control gains k_1 and k_2 is given to achieve the ETVF.

Theorem 1. *For any given $J_u^* > 0$, multiagent system (1) with formation control protocol (2) achieves the ETVF if $f_{iv}(t) - \dot{f}_{ix}(t) = 0$, $\dot{f}_{iv}(t) = 0\,(i = 1, 2, \ldots, N)$ and control gains k_1 and k_2 satisfy*

$$\begin{cases} 0 < k_1 < \min\left(-\lambda_{\max} k_2^2 + 2k_2\kappa^{-1}, \dfrac{J_u^* - k_2^2\lambda_{\max}\delta_x^T(0)\delta_x(0) - 2k_2\lambda_{\max}\left|\delta_x^T(0)\delta_v(0)\right| - 2\delta_v^T(0)\delta_v(0)}{2\lambda_{\max}\delta_x^T(0)\delta_x(0)}\right), \\ 0 < k_2 < 2\lambda_{\max}^{-1}\kappa^{-1}. \end{cases} \tag{7}$$

Proof. One can set that

$$\tilde{\delta}_c(t) = (Q \otimes I_2)\left[\tilde{\delta}_1^T(t), 0\right]^T = \frac{1_N}{\sqrt{N}} \otimes \tilde{\delta}_1(t), \tag{8}$$

$$\tilde{\delta}_{\bar{c}}(t) = (Q \otimes I_2)\left[0, \tilde{\delta}_2^T(t), \ldots, \tilde{\delta}_N^T(t)\right]^T = \sum_{i=2}^{N} Q e_i \otimes \tilde{\delta}_i(t). \tag{9}$$

Because Q is nonsingular, $\tilde{\delta}_c(t)$ and $\tilde{\delta}_{\bar{c}}(t)$ are linearly independent by (8) and (9), and then we can obtain

$$\tilde{\delta}(t) = \tilde{\delta}_c(t) + \tilde{\delta}_{\bar{c}}(t). \tag{10}$$

According to the structure of $\tilde{\delta}_c(t)$ given in (8), $\delta_{ix}(t) = x_i(t) - f_{ix}(t)$ and $\delta_{iv}(t) = v_i(t) - f_{iv}(t)$ $(i = 1, 2, \ldots, N)$, $1/\sqrt{N}\tilde{\delta}_1(t)$ can be used as a candidate of the formation center function $c(t)$ with $c(t) = [c_x(t), c_v(t)]^T$, and multiagent system (1) with control protocol (2) achieves the desired formation structure if and only if $\lim_{t\longrightarrow+\infty}\tilde{\delta}_i(t) = 0\,(i = 2, 3, \ldots, N)$.

In order to design control gains k_1 and k_2 to satisfy $\lim_{t\longrightarrow+\infty}\tilde{\delta}_i(t) = 0\,(i = 2, 3, \ldots, N)$, the Lyapunov function is chosen as follows:

$$V_i(t) = \tilde{\delta}_i^T(t)\Omega_i\tilde{\delta}_i(t), \tag{11}$$

where the matrix $\Omega_i = \begin{bmatrix} \lambda_i^2k_2^2 + 2\lambda_ik_1 & \lambda_ik_2 \\ \lambda_ik_2 & 2 \end{bmatrix}$. Because $\lambda_i > 0$ and control gains $k_1, k_2 > 0$, it can be obtained that $\lambda_i^2k_2^2 + 2\lambda_ik_1 > 0$ and $\det(\Omega_i) = \lambda_i^2k_2^2 + 4\lambda_ik_1 > 0$. In this case, the matrix Ω_i is positive definite, and the Lyapunov function $V_i(t) > 0$ if $\tilde{\delta}_i(t) \neq 0$. The time derivative of $V_i(t)$ along with (6) is

$$\begin{aligned}\dot{V}_i(t) = &-2\lambda_i^2k_1k_2\tilde{\delta}_{ix}^2(t) - 2\lambda_ik_2\tilde{\delta}_{iv}^2(t) + 2\tilde{\delta}_i^T(t)\Omega_i \\ &\times\left(\left(e_i^TQ^T \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)f(t) - \left(e_i^TQ^T \otimes I_2\right)\dot{f}(t)\right),\end{aligned} \tag{12}$$

where one can find that $-2\lambda_i^2k_1k_2 < 0$ and $-2\lambda_ik_2 < 0$. Since

$$\left(e_i^TQ^T \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)f(t) = \left(\left(e_i^TQ^T \otimes I_2\right)\left(I_N \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)\right)f(t), \tag{13}$$

we have

$$\begin{aligned}&\left(e_i^TQ^T \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)f(t) - \left(e_i^TQ^T \otimes I_2\right)\dot{f}(t) \\ &\quad = \left(e_i^TQ^T \otimes I_2\right)\left(\left(I_N \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)f(t) - \dot{f}(t)\right).\end{aligned} \tag{14}$$

It can be found that $Q^T \otimes I_2$ is nonsingular, and if $f_{iv}(t) - \dot{f}_{ix}(t) = 0$, $\dot{f}_{iv}(t) = 0\,(i = 1, 2, \ldots, N)$, then one can obtain that

$$\left(e_i^TQ^T \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)f(t) - \left(e_i^TQ^T \otimes I_2\right)\dot{f}(t) = 0. \tag{15}$$

According to (12) and (15), it is verified that $\dot{V}_i(t) < 0$ if $\tilde{\delta}_i(t) \neq 0$; then, one can obtain that $\lim_{t\longrightarrow+\infty}\tilde{\delta}_i(t) = 0\,(i = 2, 3, \ldots, N)$. Hence, multiagent system (1) with formation control protocol (2) can achieve the desired time-varying formation.

The next step is to focus on the problem of the energy-limited formation. It can be seen that

$$\sum_{i=1}^N u_i^2(t) = \delta^T(t)\left(L^2 \otimes \begin{bmatrix} k_1^2 & k_1k_2 \\ k_1k_2 & k_2^2 \end{bmatrix}\right)\delta(t). \tag{16}$$

Since $\delta(t) = (Q \otimes I_2)\tilde{\delta}(t)$ and $\lambda_1 = 0$, (16) can be converted into

$$\tilde{\delta}^T(t)\left(\Lambda^2 \otimes \begin{bmatrix} k_1^2 & k_1k_2 \\ k_1k_2 & k_2^2 \end{bmatrix}\right)\tilde{\delta}(t) = \sum_{i=2}^N \lambda_i^2\tilde{\delta}_i^T(t)\begin{bmatrix} k_1^2 & k_1k_2 \\ k_1k_2 & k_2^2 \end{bmatrix}\delta_i(t). \tag{17}$$

From (16) and (17), it can be seen that

$$\begin{aligned}J_\Delta &= \int_0^\Delta \sum_{i=1}^N \kappa u_i^2(t)\mathrm{d}t \\ &= \int_0^\Delta \kappa\left(\sum_{i=2}^N \lambda_i^2\left(k_1\tilde{\delta}_{ix}(t) + k_2\tilde{\delta}_{iv}(t)\right)^2\right)\mathrm{d}t,\end{aligned} \tag{18}$$

where $\Delta \geq 0$. Because $\int_0^\Delta \dot{V}_i(t)dt = V_i(\Delta) - V_i(0)$, one can get

$$\begin{aligned}J_\Delta &= \int_0^\Delta \kappa\left(\sum_{i=2}^N \lambda_i^2\left(k_1\tilde{\delta}_{ix}(t) + k_2\tilde{\delta}_{iv}(t)\right)^2\right)\mathrm{d}t + \sum_{i=2}^N\left(\int_0^\Delta \dot{V}_i(t)\mathrm{d}t - V_i(\Delta) + V_i(0)\right) \\ &= \sum_{i=2}^N\left(\int_0^\Delta\left(\left(\kappa\lambda_i^2k_1^2 - 2\lambda_i^2k_1k_2\right)\tilde{\delta}_{ix}^2(t) + 2\kappa\lambda_i^2k_1k_2\tilde{\delta}_{ix}(t)\tilde{\delta}_{iv}(t) + \left(\kappa\lambda_i^2k_2^2 - 2\lambda_ik_2\right)\tilde{\delta}_{iv}^2(t)\right)\mathrm{d}t - V_i(\Delta)\right) + \sum_{i=2}^N V_i(0).\end{aligned} \tag{19}$$

Let

$$\Xi_i = \begin{bmatrix} \kappa\lambda_i^2k_1^2 - 2\lambda_i^2k_1k_2 & \kappa\lambda_i^2k_1k_2 \\ \kappa\lambda_i^2k_1k_2 & \kappa\lambda_i^2k_2^2 - 2\lambda_ik_2 \end{bmatrix}. \tag{20}$$

If $\Xi_i < 0$, then $\kappa\lambda_i^2k_1^2 - 2\lambda_i^2k_1k_2 < 0$ and $(\kappa\lambda_i^2k_1^2 - 2\lambda_i^2k_1k_2)(\kappa\lambda_i^2k_2^2 - 2\lambda_ik_2) - \kappa^2\lambda_i^4k_1^2k_2^2 > 0$. Hence, it can be concluded that the value range of control gains k_1 and k_2: $0 < k_1 < -\lambda_{\max}k_2^2 + 2k_2\kappa^{-1}, 0 < k_2 < 2\lambda_{\max}^{-1}\kappa^{-1}$. In this case, the matrix Ξ_i is negative definite, and from (12) to (19), one can find that

$$J_u \leq \sum_{i=2}^N V_i(0). \tag{21}$$

Because $\tilde{\delta}(t) = (Q^T \otimes I_2)\delta(t)$, one can obtain that $\tilde{\delta}_x(t) = Q^T\delta_x(t)$ and $\tilde{\delta}_v(t) = Q^T\delta_v(t)$, where $\tilde{\delta}_x(t) = [\tilde{\delta}_{1x}(t), \tilde{\delta}_{2x}(t), \ldots, \tilde{\delta}_{Nx}(t)]^T$, $\tilde{\delta}_v(t) = [\tilde{\delta}_{1v}(t), \tilde{\delta}_{2v}(t), \ldots, \tilde{\delta}_{Nv}(t)]^T$, $\delta_x(t) = [\delta_{1x}(t), \delta_{2x}(t), \ldots, \delta_{Nx}(t)]^T$, and $\delta_v(t) = [\delta_{1v}(t), \delta_{2v}(t), \ldots, \delta_{Nv}(t)]^T$. Moreover, (21) can be rewritten as

$$\begin{aligned}\sum_{i=2}^N V_i(0) &= \sum_{i=2}^N\left(\lambda_i^2k_2^2 + 2\lambda_ik_1\right)\tilde{\delta}_{ix}^2(0) + 2\lambda_ik_2\tilde{\delta}_{ix}(0)\tilde{\delta}_{iv}(t) + 2\tilde{\delta}_{iv}^2(0) \\ &= k_2^2\tilde{\delta}_x^T(0)Q^TL^2Q\tilde{\delta}_x(0) + 2k_1\tilde{\delta}_x^T(0)Q^TLQ\tilde{\delta}_x(0) \\ &\quad + 2k_2\tilde{\delta}_x^T(0)Q^TLQ\tilde{\delta}_v(0) + \sum_{i=2}^N 2\tilde{\delta}_{iv}^2(0) \\ &\leq k_2^2\lambda_{\max}^2\delta_x^T(0)\delta_x(0) + 2k_1\lambda_{\max}\delta_x^T(0)\delta_x(0) + 2k_2\lambda_{\max} \\ &\quad \cdot\left|\delta_x^T(0)\delta_v(0)\right| \\ &\quad + 2\delta_v^T(0)\delta_v(0).\end{aligned} \tag{22}$$

From (21) to (22), one can obtain that

$$J_u \le k_2^2\lambda_{\max}^2\delta_x^T(0)\delta_x(0) + 2k_1\lambda_{\max}\delta_x^T(0)\delta_x(0) + 2k_2\lambda_{\max} \cdot \left|\delta_x^T(0)\delta_v(0)\right| 2\delta_v^T(0)\delta_v(0). \tag{23}$$

In this case, there exists

$$k_1 \le \frac{J_u^* - k_2^2\lambda_{\max}^2\delta_x^T(0)\delta_x(0) - 2k_2\lambda_{\max}\left|\delta_x^T(0)\delta_v(0)\right| - 2\delta_v^T(0)\delta_v(0)}{2\lambda_{\max}\delta_x^T(0)\delta_x(0)}, \tag{24}$$

such that $J_u \le J_u^*$. Hence, the conclusion of Theorem 1 is drawn. It is worth noting that the conclusion is related to the maximum eigenvalue of the topology. Based on the theoretical analysis, the stability of the system is closely related to the Laplacian matrix, and then it can be converted into the conclusion about the eigenvalues of the Laplacian matrix. □

Remark 3. Theoretically, any desired geometric structure formation can be realized by selecting the corresponding formation function. However, from the perspective of mechanism modeling, the feasibility condition of the formation is restricted by the structure of dynamic systems in practice. Accordingly, it is significant to choose a proper time-varying formation function that meets the formation feasibility constraint condition $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} f_i(t) - \dot{f}_i(t) = 0$ $(i = 1, 2, \ldots, N)$, that is, $f_{iv}(t) - \dot{f}_{ix}(t) = 0$, $\dot{f}_{iv}(t) = 0$ $(i = 1, 2, \ldots, N)$, which is echoed by the dynamics of each agent (1). In this case, the derivative of the position component of the formation function is equal to its velocity component, and the center of the formation moves in a uniform straight line. Note that if $\dot{f}_i(t) \equiv 0\,(i = 1, 2, \ldots, N)$, then the formation feasibility constraint condition is $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} f_i(t) = 0\,(i = 1, 2, \ldots, N)$, where the formation is time-invariant. In this case, the time-invariant formation only forms a certain static geometric formation structure, but the time-varying one not only forms a certain structure but also moves in a desired geometric formation according to the requirement of the realistic environments. On the whole, it can be illustrated that the feasible condition of the time-varying formation has wide scope, while for the time-invariant formation, it is just a special case.

Remark 4. In most works, the control gain is obtained by the FEASP solver of Matlab's LMI Toolbox, which is a numerical solution algorithm. However, it is difficult and unable for the FEASP solver to find feasible solutions in some situations, which makes solutions severely limited. Therefore, it is necessary to find the analytical solutions of k_1 and k_2 that do not need to verify their feasibility. In this case, the nonlinear relationship of the control gains k_1 and k_2 is separated. It can be seen that k_2 is determined independently, and k_1 relies on k_2, and the value ranges of control gains k_1 and k_2 can be simultaneously determined by using the function monotony. Therefore, one can choose any value in the range of control gains k_1 and k_2, instead of trying to find a certain existing value that must satisfy the feasibility.

From Theorem 1, one can find that the candidate of the formation center function is given as $c(t) = 1/\sqrt{N}\tilde{\delta}_1(t)$, and a method to determine the formation center function is presented in the following theorem.

Theorem 2. *If multiagent system (1) under formation control protocol (2) achieves the ETVF determined by $f_i(t)\,(i = 1, 2, \ldots, N)$, then the formation center function $c(t)$ satisfies that*

$$\lim_{t\longrightarrow\infty}\left(c(t) - \frac{1}{N}\left(e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t}\sum_{i=1}^{N}\left[x_i(0), v_i(0)\right]^T + \sum_{i=1}^{N} f_i(t)\right)\right) = 0. \tag{25}$$

Proof. Since $Q = [1_N/\sqrt{N}, \tilde{Q}]$ and $\tilde{\delta}(t) = (Q^T \otimes I_2)\delta(t)$, one can find that

$$\tilde{\delta}_1(0) = \left(\frac{1_N^T}{\sqrt{N}} \otimes I_2\right)\delta(0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_i(0). \tag{26}$$

According to $\delta_i(t) = [\delta_{ix}(t), \delta_{iv}(t)]^T$ with $\delta_{ix}(t) = x_i(t) - f_{ix}(t)$ and $\delta_{iv}(t) = v_i(t) - f_{iv}(t)\,(i = 1, 2, \ldots, N)$, (26) is converted into

$$\tilde{\delta}_1(0) = \frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}\left[x_i(0), v_i(0)\right]^T - \sum_{i=1}^{N} f_i(0)\right). \tag{27}$$

In the following, the dynamic response of subsystem (5) is obtained by $\tilde{\delta}_1(0)$, $f(t)$, and $\dot{f}(t)$, respectively.

$$c_0(t) = \frac{1}{\sqrt{N}} e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t} \left(\sum_{i=1}^{N} [x_i(0), v_i(0)]^T - \sum_{i=1}^{N} f_i(0) \right), \quad (28)$$

$$c_f(t) = \int_0^t e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}(t-\tau)} \left(\frac{1_N^T}{\sqrt{N}} \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right) f(\tau) \mathrm{d}\tau, \quad (29)$$

$$c_{\dot{f}}(t) = -\int_0^t e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}(t-\tau)} \left(\frac{1_N^T}{\sqrt{N}} \otimes I_2 \right) \dot{f}(\tau) \mathrm{d}\tau. \quad (30)$$

Because

$$\begin{aligned} \left(1_N^T \otimes I_2\right) f(t) &= \sum_{i=1}^{N} f_i(t), \\ &\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}(t-\tau)} \left(1_N^T \otimes I_2\right) f(\tau) \\ &= e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}(t-\tau)} \left(1_N^T \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right) f(\tau), \end{aligned} \quad (31)$$

by (28), (29), and (30), one can obtain that the dynamic response of subsystem (5) can be written as follows:

$$\tilde{\delta}_1(t) = \frac{1}{\sqrt{N}} e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t} \sum_{i=1}^{N} [x_i(0), v_i(0)]^T - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f_i(t). \quad (32)$$

Furthermore, the formation center function is

$$c(t) = \frac{1}{\sqrt{N}} \tilde{\delta}_1(t) = \frac{1}{N} e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t} \sum_{i=1}^{N} [x_i(0), v_i(0)]^T - \frac{1}{N} \sum_{i=1}^{N} f_i(t). \quad (33)$$

From (26) to (33), the conclusion of Theorem 2 is drawn. □

Remark 5. The most critical problem is to develop the formula of formation center function $c(t)$ in Theorem 2. The formula is composed of the initial state and the formation function, in which the former is connected with the average value of initial states $x_i(0)$ and $v_i(0)$ and the dynamics of each agent and the latter is associated with the average value of the formation function. Under this circumstance, the formation center function $c(t)$ is located at the geometric center of the whole formation structure, which depicts the whole macroscopic motion, and the formation function $f_i(t)(i = 1, 2, \ldots, N)$ denotes expected formation state, which describes relative microscopic motion among agents. Besides, formation center function $c(t)$ has nothing to do with the derivative of the formation function, and energy limitation does not affect the formation center function.

4. ETVF Control for the Leader-Following Case

The energy-limited multiagent system consisting of a leader and $N-1$ followers is focused on in this section. The dynamics of the leader is shown as

$$\begin{cases} \dot{x}_1(t) = v_1(t), \\ \dot{v}_1(t) = 0, \end{cases} \quad (34)$$

where $x_1(t)$ and $v_1(t)$ are the position and velocity terms of the leader, and the leader does not obtain any state information from all followers. The dynamics of the followers is given by

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = u_i(t), \end{cases} \quad (35)$$

where $i = 2, 3, \ldots, N$. For each follower, there is at least one path to get the state information from the leader, and the leader as a reference object for all followers is uncontrolled; it is only a one-way transmission. Let $f_1(t) \equiv 0$; then, the formation control protocol for the energy-limited formation is given as follows:

$$\begin{cases} u_i(t) = w_{i1} k_1 \left(x_1(t) - x_i(t) + f_{ix}(t)\right) + w_{i1} k_2 \left(v_1(t) - v_i(t) + f_{iv}(t)\right) \\ + \sum_{j \in N_i, j \neq 1} w_{ij} \left(k_1 \left(x_j(t) - f_{jx}(t) - x_i(t) + f_{ix}(t)\right) + k_2 \left(v_j(t) - f_{jv}(t) - v_i(t) + f_{iv}(t)\right)\right), \\ J_u = \sum_{i=2}^{N} \int_0^{+\infty} \kappa u_i^2(t) \mathrm{d}t. \end{cases} \quad (36)$$

The states of followers are required to maintain a time-varying formation determined by the formation function $f_i(t) = [f_{ix}(t), f_{iv}(t)]^T$ and to track the states of the leader. Suppose that the formation center function $\bar{c}(t) = [x_1(t), v_1(t)]^T$ for the leader-following case, where $x_1(t)$ and $v_1(t)$ correspond to the position and velocity terms of the leader, respectively.

Let $\psi_{ix}(t) = x_i(t) - f_{ix}(t)$ and $\psi_{iv}(t) = v_i(t) - f_{iv}(t)(i = 2, 3, \ldots, N)$; then, the dynamics of multiagent system (35) under formation control protocol (36) can be shown as

$$\begin{cases} \dot{\psi}_{ix}(t) = \psi_{iv}(t) + f_{iv}(t) - \dot{f}_{ix}(t), \\ \dot{\psi}_{iv}(t) = k_1 w_{i1} x_1(t) + k_2 w_{i1} v_1(t) + k_1 \sum_{j\in N_i, j\neq 1} w_{\sigma(t),ij}\left(\psi_{jx}(t) - \psi_{ix}(t)\right) \\ +k_2 \sum_{j\in N_i, j\neq 1} w_{\sigma(t),ij}\left(\psi_{jv}(t) - \psi_{iv}(t)\right) - \dot{f}_{iv}(t). \end{cases} \tag{37}$$

Let $\psi_i(t) = [\psi_{ix}(t), \psi_{iv}(t)]^T (i = 2, 3, \ldots, N)$ and $\psi(t) = [\bar{c}^T(t), \psi_2^T(t), \ldots, \psi_N^T(t)]^T$; then, multiagent system (35) can be rewritten in the following Kronecker form:

$$\begin{aligned} \dot{\psi}(t) = &\left(L \otimes \begin{bmatrix} 0 & 0 \\ -k_1 & -k_2 \end{bmatrix} + I_N \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)\psi(t) \\ &+\left(I_N \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) f(t) - \dot{f}(t), \end{aligned} \tag{38}$$

where the Laplacian matrix $L = \begin{bmatrix} 0 & 0 \\ -l_{\text{fl}} & L_{\text{ff}} + \Lambda_{\text{fl}} \end{bmatrix}$ with $l_{\text{fl}} = [w_{21}, w_{31}, \ldots, w_{N1}]^T$, $\Lambda_{\text{fl}} = \text{diag}\{w_{21}, w_{31}, \ldots, w_{N1}\}$, and L_{ff} represents the Laplacian matrix among the followers.

Let the matrix $\widehat{T} = \begin{bmatrix} 1 & 0 \\ 1_{N-1} & I_{N-1} \end{bmatrix}$; then, one can obtain that

$$\widehat{T}^{-1} L \widehat{T} = \begin{bmatrix} 0 & \\ & L_{\text{ff}} + \Lambda_{\text{fl}} \end{bmatrix}. \tag{39}$$

Build the state differences of the leader and each follower $\overline{\psi}_i(t) = \psi_i(t) - \bar{c}(t) (i = 2, 3, \ldots, N)$; one can find that $(T \otimes I_2)\psi(t) = [\bar{c}^T(t), \overline{\psi}^T(t)]^T$ with $\overline{\psi}(t) = [\overline{\psi}_2^T(t), \overline{\psi}_3^T(t), \ldots, \overline{\psi}_N^T(t)]^T$. Since the local communication topology among followers is undirected, the Laplacian matrix $L_{\text{ff}} + \Lambda_{\text{fl}}$ is symmetric. Without loss of generality, suppose that $\widehat{\lambda}_2, \widehat{\lambda}_3, \ldots, \widehat{\lambda}_N > 0 (i = 2, 3, \ldots, N)$ are eigenvalues of Laplacian matrix $L_{\text{ff}} + \Lambda_{\text{fl}}$. Under the circumstances, there exists an orthonormal matrix $\widetilde{Q}$ such that $\widetilde{Q}^T (L_{\text{ff}} + \Lambda_{\text{fl}})\widetilde{Q} = \widetilde{\Lambda}$, where $\widetilde{\Lambda} = \text{diag}\{\widehat{\lambda}_2, \widehat{\lambda}_3, \ldots, \widehat{\lambda}_N\}$. Let $(\widetilde{Q}^T \otimes I_2)\overline{\psi}(t) = \widetilde{\psi}(t) = [\widetilde{\psi}_2^T(t), \ldots, \widetilde{\psi}_N^T(t)]^T$; then, (38) can be converted into

$$\begin{aligned} \dot{\bar{c}}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \bar{c}(t), \\ \dot{\widetilde{\psi}}(t) &= \begin{bmatrix} 0 & 1 \\ -\widehat{\lambda}_i k_1 & -\widehat{\lambda}_i k_2 \end{bmatrix} \widetilde{\psi}_i(t) \\ &\quad + \left(e_{i-1}^T \widetilde{Q}^T [0, I_{N-1}] \widehat{T}^{-1} \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) f(t) \\ &\quad - \left(e_{i-1}^T \widetilde{Q}^T [0, I_{N-1}] \widehat{T}^{-1} \otimes I_2\right) \dot{f}(t), \end{aligned} \tag{40}$$

where $i = 2, 3, \ldots, N$, $\bar{c}(t)$ is the formation center function, and multiagent (1) under formation control protocol (2) achieves the desired formation.

Next, it can be found that

$$\sum_{i=2}^{N} u_i^2(t) = \overline{\psi}^T(t)\left(\left(L_{ff} + \Lambda_{fl}\right)^2 \otimes \begin{bmatrix} k_1^2 & k_1 k_2 \\ k_1 k_2 & k_2^2 \end{bmatrix}\right)\overline{\psi}(t). \tag{41}$$

Since $(\widetilde{Q}^T \otimes I_2)\overline{\psi}(t) = \widetilde{\psi}(t)$, (41) can be transformed into

$$\widetilde{\psi}^T(t)\left(\widetilde{\Lambda}^2 \otimes \begin{bmatrix} k_1^2 & k_1 k_2 \\ k_1 k_2 & k_2^2 \end{bmatrix}\right)\widetilde{\psi}(t) = \sum_{i=2}^{N} \widehat{\lambda}_i^2 \left(k_1 \widetilde{\psi}_{ix}(t) + k_2 \widetilde{\psi}_{iv}(t)\right)^2. \tag{42}$$

Let $\overline{\psi}_x(t) = [\overline{\psi}_{2x}(t), \overline{\psi}_{3x}(t), \ldots, \overline{\psi}_{Nx}(t)]^T$ and $\overline{\psi}_v(t) = [\overline{\psi}_{2v}(t), \overline{\psi}_{3v}(t), \ldots, \overline{\psi}_{Nv}(t)]^T$; one can obtain that

$$\begin{aligned} \sum_{i=2}^{N} u_i^2(t) &= k_2^2 \widetilde{\psi}_x^T(t) \widetilde{Q}^T \left(L_{ff} + \Lambda_{fl}\right)^2 \widetilde{Q} \widetilde{\psi}_x(t) + 2k_1 \widetilde{\psi}_x^T(t) \widetilde{Q}^T \left(L_{ff} + \Lambda_{fl}\right) \widetilde{Q} \widetilde{\psi}_x(t) \\ &\quad + 2k_2 \widetilde{\psi}_x^T(t) \widetilde{Q}^T \left(L_{ff} + \Lambda_{fl}\right) \widetilde{Q} \widetilde{\psi}_v(t) + 2\widetilde{\psi}_v^T(t) \widetilde{Q}^T \widetilde{Q} \widetilde{\psi}_v(t) \\ &\le k_2^2 \widehat{\lambda}_{\max}^2 \overline{\psi}_x^T(0) \overline{\psi}_x(0) + 2k_1 \widehat{\lambda}_{\max} \overline{\psi}_x^T(0) \overline{\psi}_x(0) + 2k_2 \widehat{\lambda}_{\max} \left|\overline{\psi}_x^T(0) \overline{\psi}_v(0)\right| \\ &\quad + 2\widehat{\lambda}_{\max} \overline{\psi}_v^T(0) \overline{\psi}_v(0). \end{aligned} \tag{43}$$

The proof procedure from (41) to (43) is similar to the Theorem 1. Hence, the conclusions about energy-limited formation can be drawn in the following.

Theorem 3. *For any given $J_u^* > 0$, multiagent system (1) with formation control protocol (2) achieves the ETVF if $f_{iv}(t) - \dot{f}_{ix}(t) = 0$, $\dot{f}_{iv}(t) = 0\,(i = 2, 3, \ldots, N)$ and there exist control gains k_1 and k_2 satisfying*

$$\begin{cases} 0 < k_1 < \min\left(-\widehat{\lambda}_{\max} k_2^2 + 2k_2\kappa^{-1}, \dfrac{J_u^* - k_2^2\widehat{\lambda}_{\max}^2 \overline{\psi}_x^T(0)\overline{\psi}_x(0) - 2k_2\widehat{\lambda}_{\max}\left|\overline{\psi}_x^T(0)\overline{\psi}_v(0)\right| - 2\overline{\psi}_v^T(0)\overline{\psi}_v(0)}{2\widehat{\lambda}_{\max}\overline{\psi}_x^T(0)\overline{\psi}_x(0)}\right), \\ 0 < k_2 < 2\widehat{\lambda}_{\max}^{-1}\kappa^{-1}. \end{cases} \tag{44}$$

Remark 6. Different from the leaderless case in Theorem 1, the leader-following one has two characteristics. Firstly, the leader-following communication topology is asymmetric, which means the leader does not get any information from followers, while at least one follower can get it from the leader. Besides, the local communication topology among the followers is undirected. In this case, a two-step transformation method is proposed, wherein the most critical step is to construct state differences between followers and the leader using a special transformation, so that leaderless and leader-following cases can be unified into an identical framework and the derivation process of leader-following cases is simplified. From a macro perspective, the whole movement is determined by all agents for leaderless multiagent systems, but it only relies on the leader for leader-following ones. Secondly, the range of control gains k_1, k_2 for leaderless case is about the initial states of each agent, but it is related to the state differences of the initial values between the leader and each follower for leader-following case, which is the state difference method proposed to convert the asymmetric matrix into a symmetric one.

5. Numerical Simulation

In this section, two numerical examples of leaderless and leader-following MAS are used to verify correctness and effectiveness of the theoretical results, respectively.

Example 1. (the leaderless case) Consider the leaderless MAS with five agents moving on a 2-dimensional plane (XY plane). The dynamics of each agent is expressed by (1) with $x_i(t) = [x_{iX}(t), x_{iY}(t)]^T$, $v_i(t) = [v_{iX}(t), v_{iY}(t)]^T$, and $u_i(t) = [u_{iX}(t), u_{iY}(t)]^T\,(i \in \{1, 2, \ldots, 5\})$, in which $x_{iX}(t), v_{iX}(t), u_{iX}(t)$ and $x_{iY}(t), v_{iY}(t), u_{iY}(t)$ denoting position, velocity, and control input are along the X-axis and Y-axis in the XY plane, respectively. The communication among agents is subject to the undirected topology G shown in Figure 1, where the weight of the adjacent matrix is 0-1.

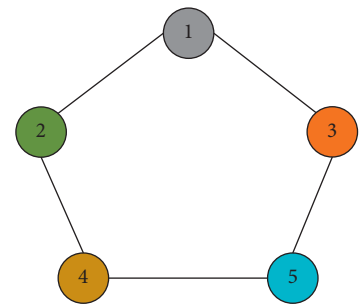


Figure 1: Communication topology G for the leaderless case.

The initial states $[x_i(0), v_i(0)]^T = [x_{iX}(0), x_{iY}(0), v_{iX}(0), v_{iY}(0)]^T\,(i \in \{1, 2, \ldots, 5\})$ are given as follows:

$$\begin{aligned} [x_1(0), v_1(0)]^T &= [1.1, 1.3, 0.4, -1.2]^T, \\ [x_2(0), v_2(0)]^T &= [0.2, -0.8, -1.3, 0.5]^T, \\ [x_3(0), v_3(0)]^T &= [-0.3, 1.1, 0.5, 0.6]^T, \\ [x_4(0), v_4(0)]^T &= [-1.3, -0.7, -1.2, 1.1]^T, \\ [x_5(0), v_5(0)]^T &= [0.9, 0.8, 1.1, 0.2]^T. \end{aligned} \tag{45}$$

The formation function for leaderless MAS is given as

$$f_i(t) = \left[t\sin\left(\frac{2(i-1)\pi}{5}\right), t\cos\left(\frac{2(i-1)\pi}{5}\right), \sin\left(\frac{2(i-1)\pi}{5}\right), \cos\left(\frac{2(i-1)\pi}{5}\right)\right]^T, \quad i \in \{1, 2, \ldots, 5\}. \tag{46}$$

Let $\kappa = 0.4$ and the total energy $J_u^* = 120$; then, one can find by Theorem 1 that $0 < k_1 < 0.3455$ and $0 < k_2 < 0.6910$. Here, one chooses $k_1 = 0.2$ and $k_2 = 0.5$.

In Figures 2 and 3, state trajectories of $\delta_{ix}(t)$ and $\delta_{iv}(t)\,(i = 1, 2, \ldots, 5)$ for leaderless MAS along the X-axis and Y-axis are depicted, respectively, where the curve marked by blue circles refers to the trajectories of the formation center function. Figure 4 depicts snapshots of the state of five agents and a formation center at different times, where five agents are marked with a red square, a green diamond, a magenta circle, a sky blue triangle, and a black pentagon, and the blue hexagon stands for the formation center. From Figures 4(a)–4(d), one can see that the pentagon formation is formed, and the pentagon formation structure is changing over time at the same multiple. Figure 5 shows the state trajectories of each agent in the whole time, and the trajectory of the formation center is given in Figure 6. Figure 7 shows that the practical energy consumption J_u converges to a finite value with $J_u < J_u^*$. It is clear that the desired pentagonal time-varying formation under

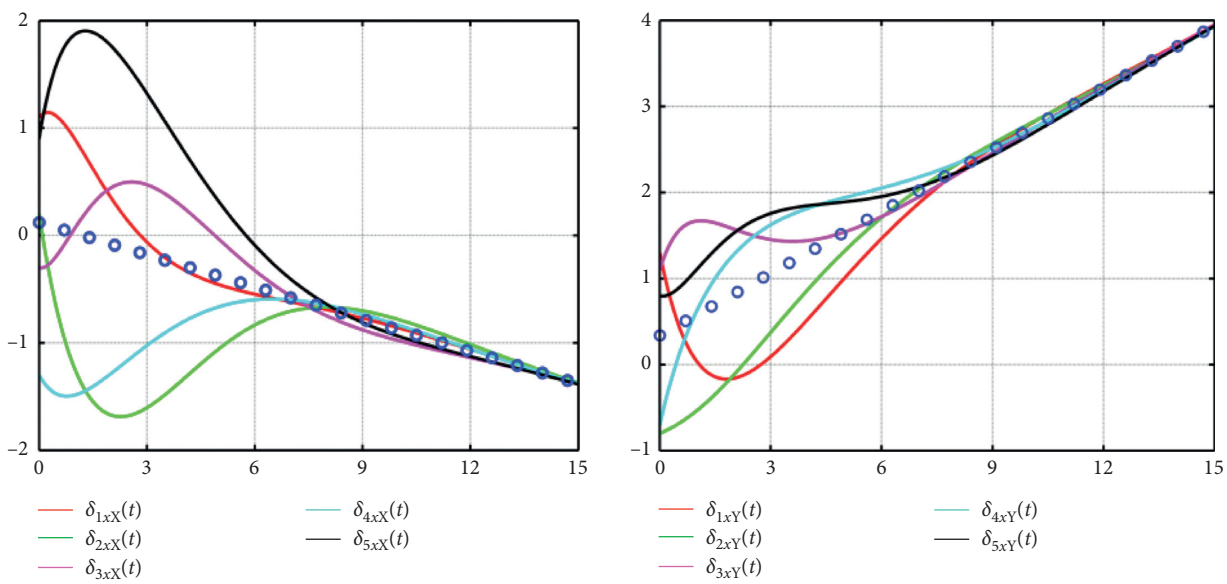


Figure 2: Trajectories of $\delta_{ix}(t)\,(i = 1, 2, \ldots, 5)$ in X and Y directions.

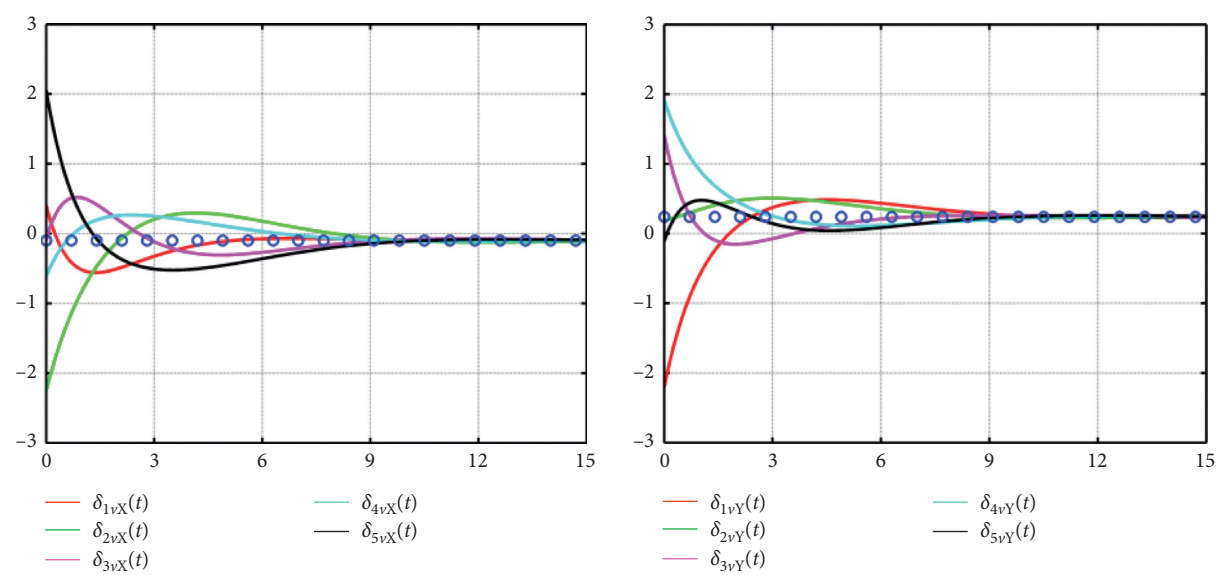


Figure 3: Trajectories of $\delta_{iv}(t)\,(i = 1, 2, \ldots, 5)$ in X and Y directions.

energy-limited formation control protocol (2) for leaderless MAS is achieved.

Example 2. (the leader-following case) In this example, the energy-limited multiagent system is composed of a leader and six followers, whose communication topology G is shown in Figure 8. The dynamics of the followers are modeled by (35) with $x_i(t) = [x_{iX}(t), x_{iY}(t), x_{iZ}(t)]^T$, $v_i(t) = [v_{iX}(t), v_{iY}(t), v_{iZ}(t)]^T$ and $u_i(t) = [u_{iX}(t), u_{iY}(t), u_{iZ}(t)]^T$ $(i \in \{2, 3, \ldots, 7\})$. The followers need to keep a time-varying formation structure with the formation following function:

$$f_i(t) = \left[t\sin\left(\frac{(i-2)\pi}{3} + \frac{\pi}{2}\right), \frac{\sqrt{2}}{2}t\sin\left(\frac{(i-2)\pi}{3}\right), \frac{\sqrt{2}}{2}t\sin\left(\frac{(i-2)\pi}{3}\right),\right.$$
$$\left.\sin\left(\frac{(i-2)\pi}{3} + \frac{\pi}{2}\right), \frac{\sqrt{2}}{2}\sin\left(\frac{(i-2)\pi}{3}\right), \frac{\sqrt{2}}{2}\sin\left(\frac{(i-2)\pi}{3}\right)\right]^T, \quad (i \in \{2, 3, \ldots, 7\}). \tag{47}$$

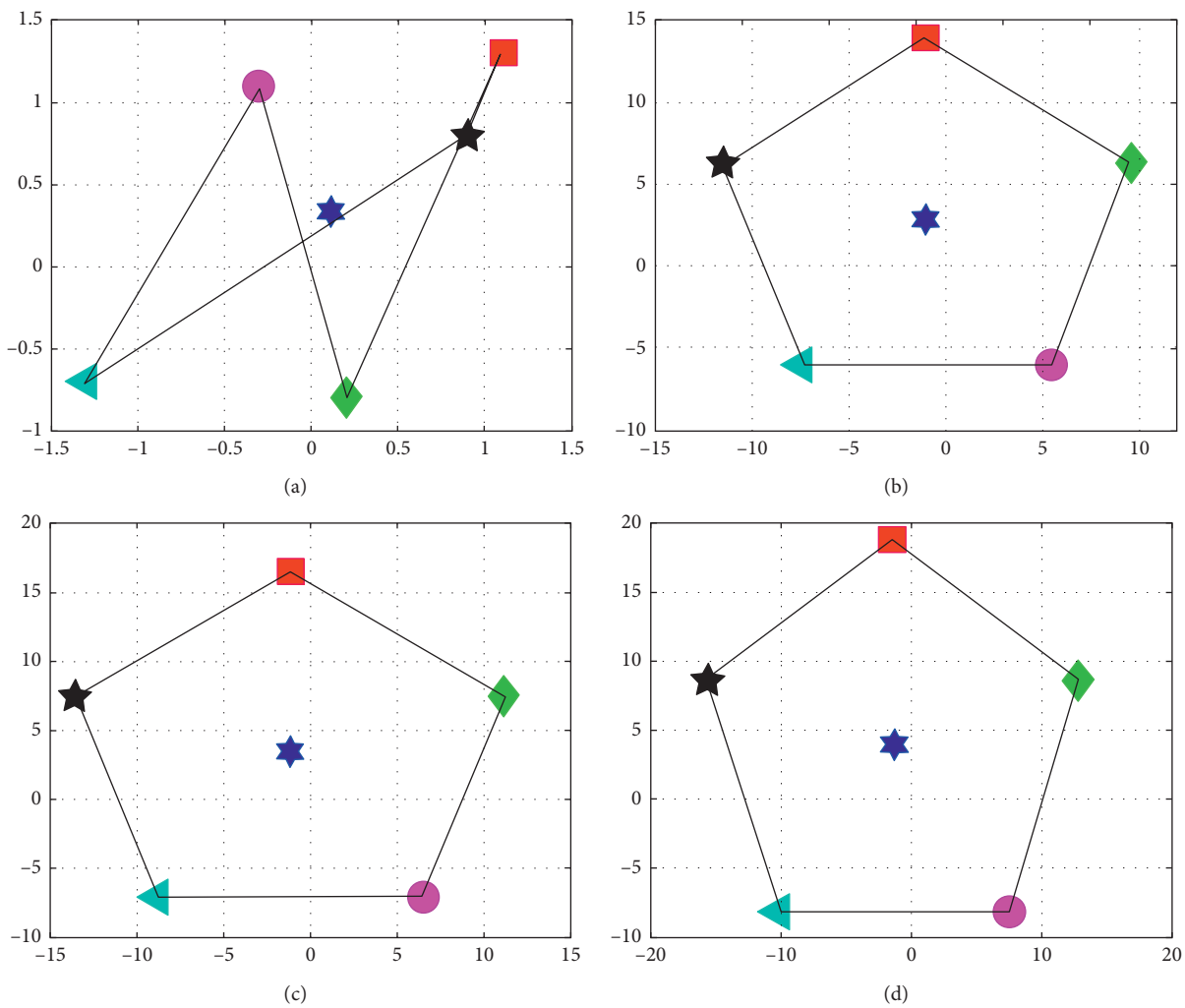


Figure 4: State snapshots of five agents at different times for the leaderless case. (a) $t = 0$s. (b) $t = 12$s. (c) $t = 14$s. (d) $t = 15$s.

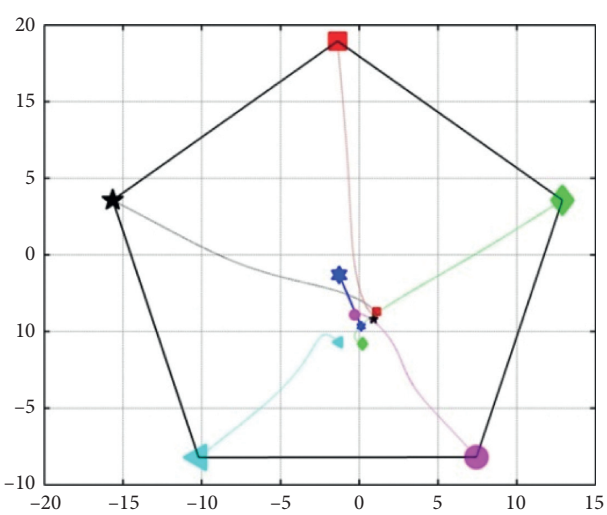


Figure 5: State trajectories of each agent for the leaderless case.

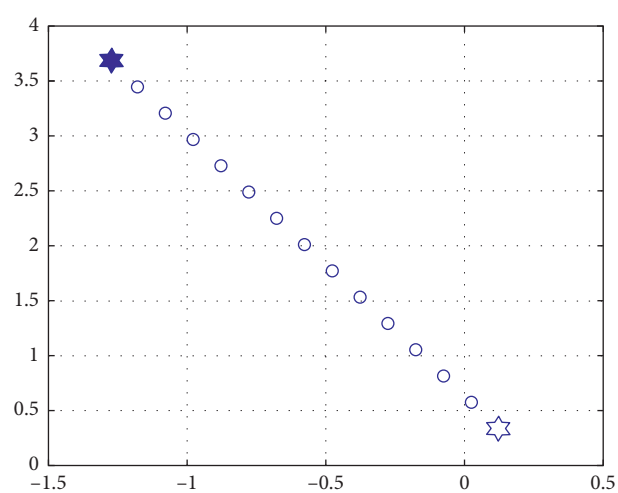


Figure 6: Trajectory of the formation center for the leaderless case.

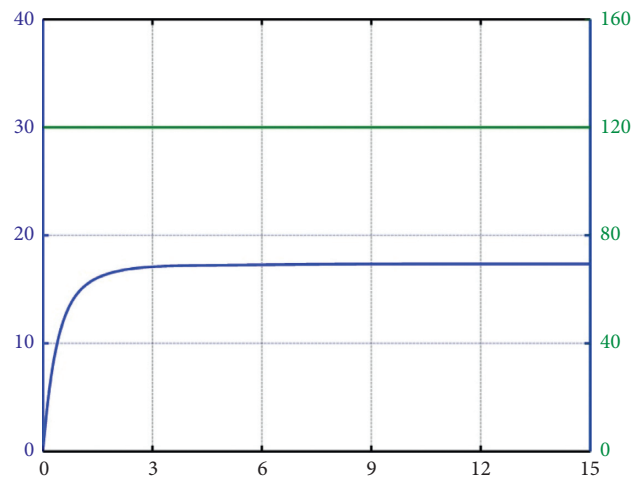


Figure 7: Trajectories of practical energy consumption and total energy supply for the leaderless case.

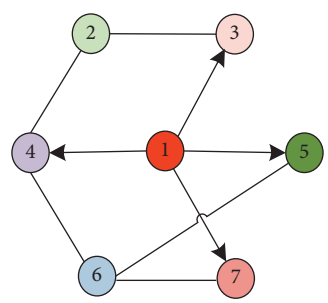


Figure 8: Communication topology G for the leader-following case.

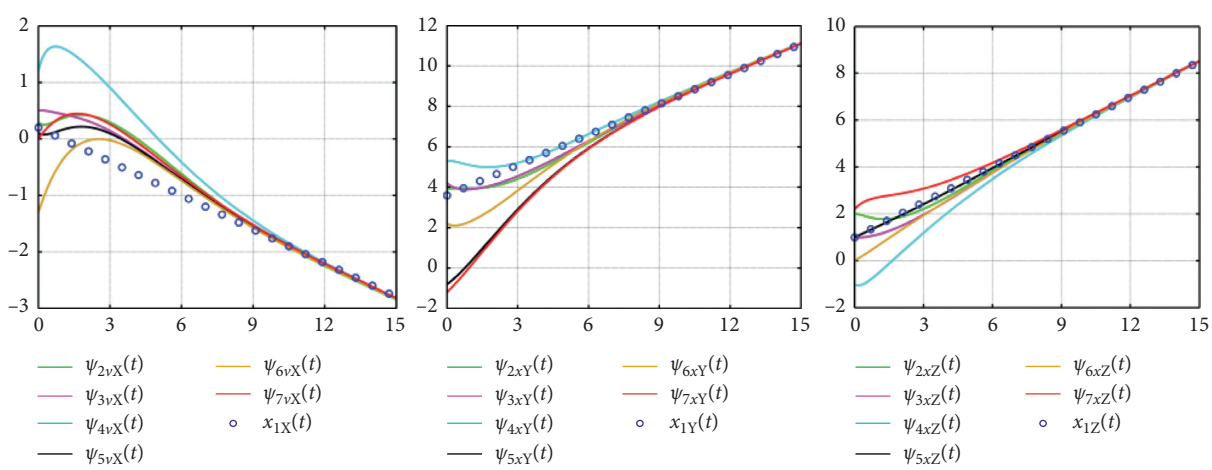


Figure 9: Trajectories of $\psi_{ix}(t)\,(i = 2, 3, \ldots, 7)$ in X, Y, and Z directions.

The initial states $[x_i(0), v_i(0)]^T = [x_{iX}(0), x_{iY}(0), x_{iZ}(0), v_{iX}(0), v_{iY}(0), v_{iZ}(0)]^T\,(i \in \{1, 2, \ldots, 7\})$ are as follows:

$$\begin{aligned}
[x_1(0), v_1(0)]^T &= [0.2, 3.6, 1.0, -0.2, 0.5, 0.5]^T,\\
[x_2(0), v_2(0)]^T &= [0.3, 3.9, 2.0, 0.5, 0.2, 0.0]^T,\\
[x_3(0), v_3(0)]^T &= [0.5, 4.2, 1.0, 0.6, -0.4, 0.5]^T,\\
[x_4(0), v_4(0)]^T &= [1.2, 5.3, -1.0, 1.3, 0.8, 0.0]^T,\\
[x_5(0), v_5(0)]^T &= [0.1, -0.8, 1.0, -1.2, 0.9, 0.5]^T,\\
[x_6(0), v_6(0)]^T &= [-1.3, 2.2, 0.0, 0.7, -1.3, 0.0]^T,\\
[x_7(0), v_7(0)]^T &= [0.0, -1.2, 2.2, 1.1, 0.6, 0.5]^T.
\end{aligned} \tag{48}$$

Choose $\kappa = 0.12$ and the total energy $J_u^* = 500$. According to Theorem 3, one can obtain that $0 < k_1 < 0.3250$ and $0 < k_2 < 1.8045$. In this case, we can choose $k_1 = 0.3$ and $k_2 = 1.0$.

Figures 9 and 10 show that state trajectories of $\psi_{ix}(t)$ and $\psi_{iv}(t)\,(i = 2, 3, \ldots, 7)$ for all followers along the X-axis, Y-axis, and Z-axis are asymptotically converge to the leader, respectively. In Figure 11, each follower is scattered in the three-dimensional plane by triangle, circle, parallelogram, hexagon, square, and pentagon at $t = 0$ s, $t = 12$ s, $t = 14$ s, and $t = 15$ s, where star denotes the leader. It can be seen

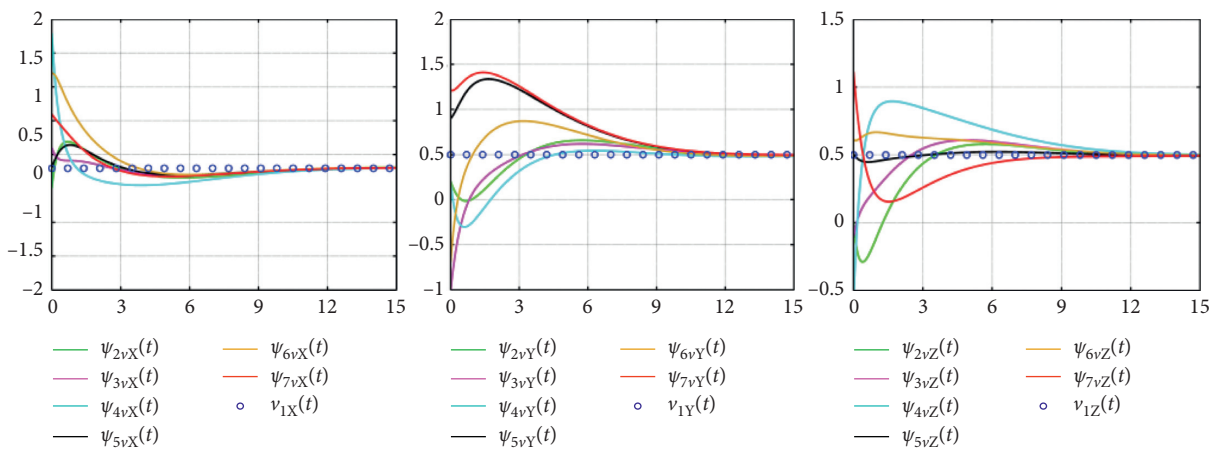


Figure 10: Trajectories of $\psi_{iv}(t)\,(i = 2, 3, \ldots, 7)$ in X, Y, and Z directions.

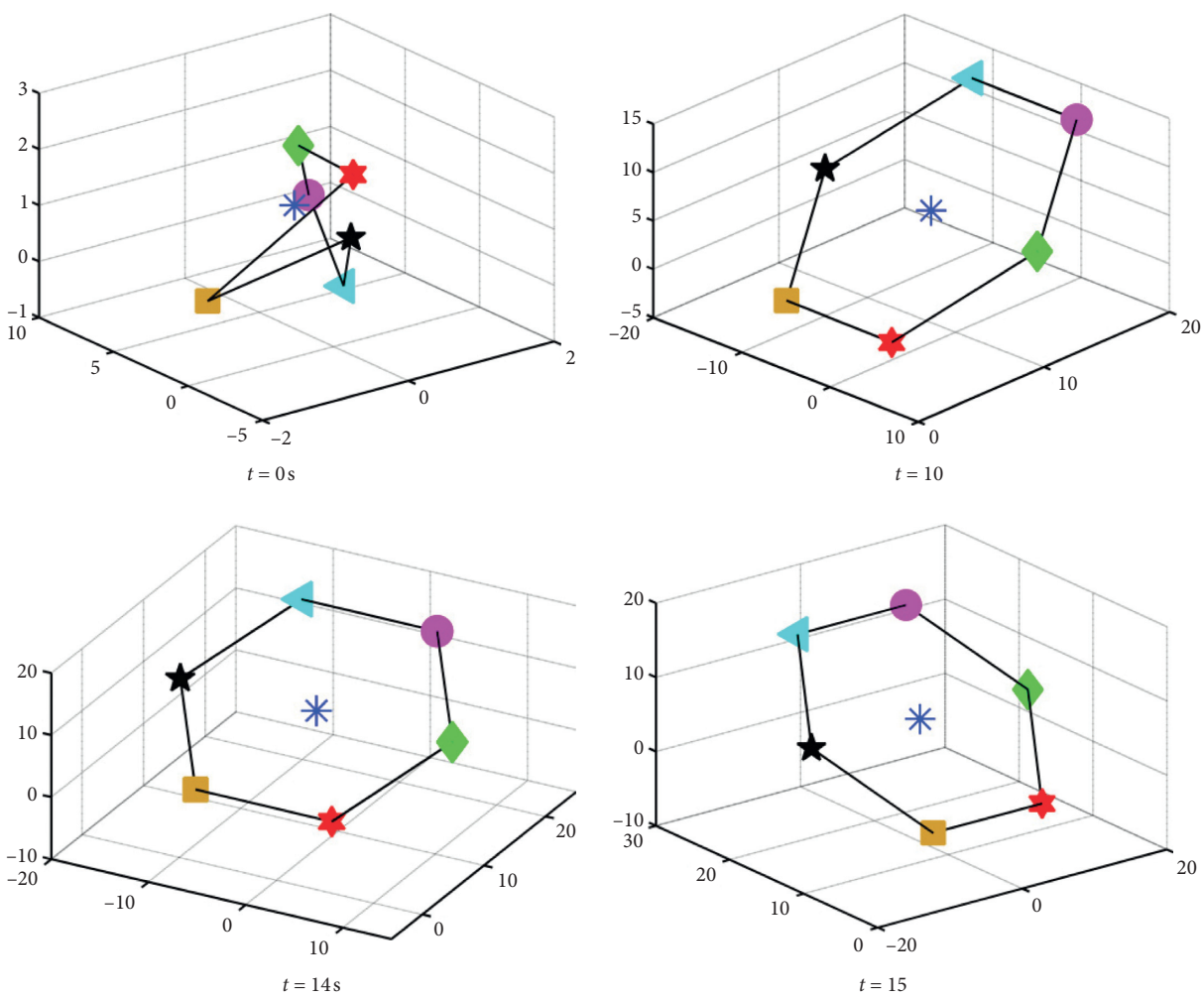


Figure 11: State snapshots of all agents at different times for the leader-following case. (a) $t = 0\,s$. (b) $t = 12\,s$. (c) $t = 14\,s$. (d) $t = 15\,s$.

that the followers finally form and move in the desired hexagonal formation structure around the leader. The state trajectory of the leader is shown in Figure 12. Compared with the leaderless simulation case, the biggest difference of the formation center function is the actual leader, but it is a virtual point for the leaderless case. From Figure 13, it can be noticed that the practical energy consumption J_u converges to a finite value with $J_u < J_u^*$. From those simulation trajectories, it reveals that all followers can form a time-varying hexagonal structure formation to track the leader.

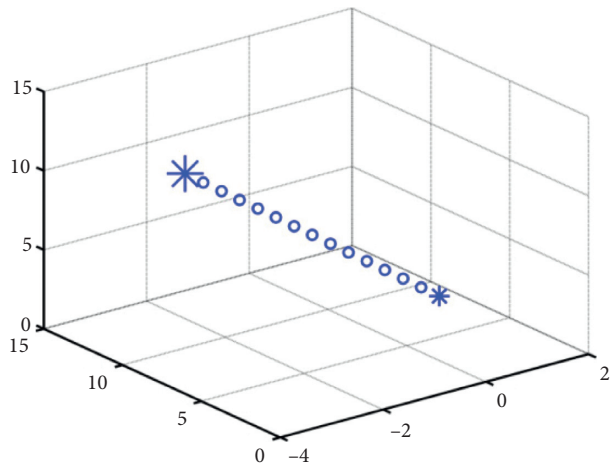


Figure 12: State trajectory of the leader for the leader-following case.

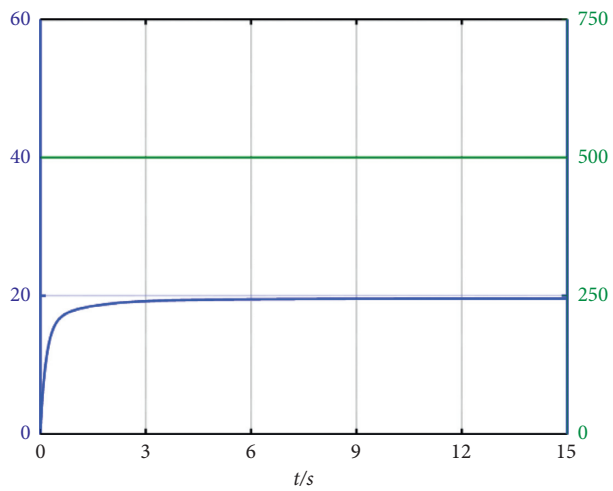


Figure 13: Trajectories of practical energy consumption and total energy supply for the leader-following case.

6. Conclusions

This paper investigates the ETVF control problem of second-order MAS for leaderless and leader-following cases, respectively. In comparison with the existing results, the energy-limited time-varying protocol is firstly presented. A key idea of our method is the employment of energy constraint. Then, the stability is analyzed according to Lyapunov theory and graph theory, and the sufficient condition for second-order MAS to achieve the time-varying formation is derived by using the joint design method of formation control gains and formation total energy supply. Meanwhile, the explicit formula of formation center function is used to describe the macroscopic movement of the whole multiagents. Moreover, we propose a two-step method to solve the problem of the asymmetric structure of the Laplacian matrix for the leader-following topology, which transforms the leader-following one into the leaderless framework so that our work is simplified. In the future, research is aimed at extending the results to nonlinear MAS.

Data Availability

All data, such as the initial values and system dynamics, are provided in the simulation part of this paper and are verified by Matlab experiments.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yulong Zhao was responsible for validation and project administration. Wanzhen Quan contributed to conceptualization, methodology, and original draft preparation. Wanzhen Quan and Le Wang were responsible for investigation. Xiaogang Yang and Le Wang supervised the study. Xiaogang Yang was responsible for funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This study was supported by the National Natural Science Foundation of China under grant nos. 61867005, 61806209, 61763040, and 61703411 and in part by the Open Foundation of Shaanxi Key Laboratory of Integrated and Intelligent Navigation under grant no. SKLIIN-20180103.

References

[1] F. Derakhshan and S. Yousefi, "A review on the applications of multiagent systems in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 15, no. 5, 2019.

[2] J. Qin, W. Fu, H. Gao, and W. X. Zheng, "Distributed k-means algorithm and fuzzy c-means algorithm for sensor networks based on multiagent consensus theory," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 772–783, 2017.

[3] J. Qu, Z. Ji, C. Lin, and H. Yu, "Fast consensus seeking on networks with antagonistic interactions," *Complexity*, vol. 2018, Article ID 7831317, 15 pages, 2018.

[4] H. Liu, T. Ma, F. L. Lewis, and Y. Wan, "Robust formation control for multiple quadrotors with nonlinearities and disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1362–1371, 2020.

[5] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 7, pp. 2442–2454, 2020.

[6] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4585–4598, 2020.

[7] K.-K. Oh and H.-S. Ahn, "Formation control of mobile agents based on distributed position estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 737–742, 2013.

[8] S. Martin, "Multi-agent flocking under topological interactions," *Systems & Control Letters*, vol. 69, pp. 53–61, 20149.

[9] G. Wen, Z. Duan, H. Su, G. Chen, and W. Yu, "A Connectivity-preserving flocking algorithm for multi-agent

dynamical systems with bounded potential function," *IET Control Theory & Applications*, vol. 6, no. 6, pp. 813–812, 2012.

[10] D. Yu, C. L. P. Chen, C.-E. Ren, and S. Sui, "Swarm control for self-organized system with fixed and switching topology," *IEEE Transactions on Cybernetics*, vol. 50, no. 10, p. 4481, 2020.

[11] M. Komareji, Y. Shang, and R. Bouffanais, "Consensus in topologically interacting swarms under communication constraints and time-delays," *Nonlinear Dynamics*, vol. 93, no. 3, pp. 1287–1300, 2018.

[12] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, 2020, In press.

[13] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[14] S. Liu, Z. Ji, and H. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, Article ID 7685460, 18 pages, 2020.

[15] X. Jia, L. Hu, F. Feng, and J. Xu, "Robust H∞ consensus control for linear discrete-time swarm systems with parameter uncertainties and time-varying delays," *International Journal of Aerospace Engineering*, vol. 2019, Article ID 7278531, 16 pages, 2019.

[16] H. Zhao and J. H. Park, "Dynamic output feedback consensus of continuous-time networked multiagent systems," *Complexity*, vol. 20, no. 5, pp. 35–42, 2014.

[17] J. Xi, M. He, H. Liu, and J. Zheng, "Admissible output consensualization control for singular multi-agent systems with time delays," *Journal of the Franklin Institute*, vol. 353, no. 16, pp. 4074–4090, 2016.

[18] M. Ou, H. Du, and S. Li, "Finite-time formation control of multiple nonholonomic mobile robots," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 1, pp. 140–165, 2014.

[19] T. Murayama, "Distributed model predictive consensus control for robotic swarm system," *Artificial Life and Robotics*, vol. 23, no. 4, pp. 628–635, 2018.

[20] J. Zhou and J. Yang, "Distributed guidance law design for cooperative simultaneous attacks with multiple missiles," *Journal of Guidance, Control and Dynamics*, vol. 39, no. 10, pp. 1–9, 2016.

[21] P. Dasgupta, "A multiagent swarming system for distributed automatic target recognition using unmanned aerial vehicles," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 38, no. 3, pp. 549–563, 2008.

[22] Z. Lin and H. Liu, "Topology-based distributed optimization for multi-UAV cooperative wildfire monitoring," *Optimal Control Applications and Methods*, vol. 38, no. 3, pp. 1530–1548, 2018.

[23] M. Fabris, A. Cenedese, and J. Hauser, "Optimal time-invariant formation tracking for a second-order multi-agent system," in *Proceedings of the 2019 European Control Conference*, pp. 1556–1561, Naples, Italy, June 2019.

[24] L. Dong, Y. Chen, X. Qu, and X. Qu, "Formation control strategy for nonholonomic intelligent vehicles based on virtual structure and consensus approach," *Procedia Engineering*, vol. 137, pp. 415–424, 2016.

[25] M. Li, Q. Ma, C. Zhou, J. Qin, and Y. Kang, "Distributed time-varying group formation control for generic linear systems with observer-based protocols," *Neurocomputing*, vol. 397, pp. 244–252, 2020.

[26] X. Dong, Y. Zhou, Z. Ren, and Y. Zhong, "Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5014–5024, 2017.

[27] H. Du, G. Wen, Y. Cheng, Y. He, and R. Jia, "Distributed finite-time cooperative control of multiple high-order nonholonomic mobile robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2998–3006, 2016.

[28] Y. Jia, J. Du, W. Zhang, and L. Wang, "Three-Dimensional leaderless flocking control of large-scale small unmanned aerial vehicles," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6208–6213, 2017.

[29] Y. Li, C. Hua, and X. Guan, "Distributed output feedback leader-following control for high-order nonlinear multiagent system using dynamic gain method," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 640–649, 2020.

[30] L. Gao, Y. Cui, W. Chen, and W. Chen, "Leader-following consensus for discrete-time descriptor multi-agent systems with observer-based protocols," *Transactions of the Institute of Measurement and Control*, vol. 38, no. 11, pp. 1353–1364, 2016.

[31] Z. Zhang, L. Zhang, F. Hao, and L. Wang, "Leader-following consensus for linear and Lipschitz nonlinear multiagent systems with quantized communication," *IEEE Transactions on Cybernetics*, vol. 47, no. 8, pp. 1970–1982, 2017.

[32] Z. Yan, X. Pan, Z. Yang, and L. Yue, "Formation control of leader-following multi-UUVs with uncertain factors and time-varying delays," *IEEE Access*, vol. 7, pp. 118792–118805, 2019.

[33] X. Dong, J. Xiang, L. Han, Q. Li, and Z. Ren, "Distributed time-varying formation tracking analysis and design for second-order multi-agent systems," *Journal of Intelligent & Robotic Systems*, vol. 86, no. 2, pp. 277–289, 2016.

[34] X. Dong, Q. Li, Q. Zhao, and Z. Ren, "Time-varying group formation analysis and design for second-order multi-agent systems with directed topologies," *Neurocomputing*, vol. 205, pp. 367–374, 2016.

[35] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 7, no. 30, pp. 2796–2808, 2020.

[36] Z.-H. Guan, B. Hu, M. Chi, D.-X. He, and X.-M. Cheng, "Guaranteed performance consensus in second-order multi-agent systems with hybrid impulsive control," *Automatica*, vol. 50, no. 9, pp. 2415–2418, 2014.

[37] D. M. Zhang, L. Meng, X. G. Wang, and L. L. Ou, "Linear quadratic regulator control of multi-agent systems," *Optimal Control Applications and Methods*, vol. 36, no. 1, pp. 45–59, 2015.

[38] W. Dong, "Distributed optimal control of multiple systems," *International Journal of Control*, vol. 83, no. 10, pp. 2067–2079, 2010.

[39] C. Godsil and G. Royal, *Algebraic Graph Theory*, Springer-Verlag, New York, NY, USA, 2001.

Hindawi
Complexity
Volume 2020, Article ID 6678190, 14 pages
https://doi.org/10.1155/2020/6678190

Research Article

Formation Tracking for High-Order Time-Invariant Swarm Systems with Limited Energy and Fixed Topologies

Jianye Yang,[1] **Cheng Wang**,[2] **Hongtao Dang**,[1] **and Xinzhong Han**[3]

[1]*School of Science, Xijing University, Xi'an 710123, China*
[2]*High-Tech Institute of Xi'an, Xi'an 710025, China*
[3]*Beijing BlueVision Technology Limited Company, Beijing 100070, China*

Correspondence should be addressed to Cheng Wang; m15212783833@163.com

Received 3 October 2020; Accepted 13 October 2020; Published 28 October 2020

Academic Editor: Ning Cai

The current paper investigates design and analysis problems of formation tracking for high-order linear time-invariant swarm systems, where communication topologies among agents have leader-following structures and the whole energy supply is limited. Firstly, the communication topology of a swarm system is depicted by a directed graph with a spanning tree, where the communication channels from the leader to followers are directional and the communication channels among followers are bidirectional, and a new formation tracking protocol with an energy integral term and a given upper bound is proposed to achieve formation tracking with the limited energy. Then, sufficient conditions for time-varying formation tracking design and analysis with the limited energy are presented, respectively, which include two/three linear matrix inequality constraints associated with the maximum and minimum nonzero eigenvalues of the Laplacian matrix of a communication topology. Especially, time-invariant formation tracking criteria are further deduced. Finally, two numerical examples are revealed to verify main theoretical conclusions.

1. Introduction

In the last decade, many scientists paid their attentions to design different cooperative control strategies of swarm systems with different application backgrounds, such as explanations for animal flocking phenomena [1, 2], information consensus analysis and design [3–9], analysis for different source data [10], and formation structure design and maintenance [10–13]. Distributed formation is inspired by the biological formation flying without the superintend vertex, where the formation structure maintenance is realized via distributed information transmissions and local interactions. It should be pointed out that distributed formation control has extensive applications in the cooperative operation of satellite clusters, the coordinated assembly of multiple robots, and the collaborative attack of multiple unmanned aerial ships, and Ren [14] summarized the existing formation control approaches including the virtual-structure one, the behavior-based one, and the consensus-based one and pointed out that the consensus-based approach can achieve distributed formation control and can develop the precise mathematical analysis. The consensus-based formation of linear swarm systems as well as their controllability was discussed in [15–17].

Based on the time-varying characteristics of formation structures, distributed formation can be divided into time-varying distributed formation and time-invariant distributed formation. For time-varying distributed formation control, the formation geometric structure is time-varying even if the formation structure is formed. For time-invariant distributed formation control, the formation geometric structure is time-invariant once the formation structure is finished. By using the Nyquist stability criterion, time-invariant formation criteria for swarm systems were proposed in [18], where the communication topology was modeled by graph theory and several important features of the communication topology were shown. By a nonlinear formation protocol, sufficient conditions for finite-time time-invariant formation were revealed in [19],

where theoretical results were applied into multiple nonholonomic robot systems. Qin et al. [20] studied the impacts of intermittent communications on time-invariant formation control for swarm systems with time delays. In [21–23], sufficient and/or necessary conditions of time-varying formation design and analysis were presented, where it was shown that time-varying formation control is more challenging than time-invariant formation control and formation feasible conditions of the time-varying formation are more complex than formation feasible conditions of the time-invariant formation.

According to communication topology structure properties among agents, distributed formation can be divided into two types: the leaderless distributed formation and the leader-following distributed formation. In the leaderless distributed formation, all agents in a swarm system have the same decision weight as shown in [18–23]. The leader-following distributed formation is also called formation tracking, where the leader plays a role similar to the superintend vertex and all following vertexes track the leader in a specific formation structure. For first-order swarm systems, Xiao et al. [24] proposed some novel sufficient conditions for formation tracking, where each agent was described as a first-order integrator, which means that the state of one agent is not time-invariant if it does not receive any collaborative information from its neighboring agents. Dong et al. [25] presented several sufficient conditions for formation tracking of second-order swarm systems, where each agent was described as a second-order integrator. In this case, the velocity-similar term converges to a constant and the position-similar term linearly diverges away if the convergence constant of the velocity-similar term is nonzero. In [26], formation tracking criteria for high-order linear time-invariant swarm systems were given, where the formation tracking problem is more challenging since the dynamics of each agent is of high order.

For practical swarm systems, the whole energy supply is usually limited and the above literature studies on formation control did not considered the impacts of the limited energy on formation design and analysis problems. Similar to the guaranteed-cost control of isolated systems, the whole guaranteed-cost function was constructed on the basis of the integral principle in [27, 28], where sufficient conditions for the guaranteed-cost formation were proposed via the linear matrix inequality tool and the nonzero eigenvalues of the Laplacian matrix of the communication topology. Yu et al. [29] presented the guaranteed-cost time-varying formation criteria for swarm systems with external disturbances and time-varying delays, but they did not give an approach to determine the upper bound of the performance function. Especially, it should be pointed out that the whole energy supply of a swarm system was assumed to be infinite when distributed formation control was carried out in [27–29]. Actually, the whole energy supply is usually limited for practical swarm systems and its impacts on time-varying formation control are critically important. To the best of our knowledge, the design and analysis problems of time-varying formation tracking for high-order linear time-variant swarm systems with the limited energy and fixed topologies are still open and are not comprehensively discussed.

The current paper investigated time-varying formation tracking problems for high-order linear time-variant swarm systems with the limited energy and fixed topologies. Based on the state errors and the formation function errors among neighboring agents including the collaborative information of the leader, a new formation tracking protocol is presented, where an energy integral term is introduced to guarantee that the practical energy consumption is less than or equal to the whole energy supply. Especially, the dynamics of each agent is described by a high-order linear time-invariant model and the communication topology is depicted as a directed graph with a spanning tree. Furthermore, the whole dynamics of the leader-following swarm system is separated as the dynamics of the leader and the relative dynamics among all agents by constructing the specific nonsingular transformation matrix, and design and analysis criteria for time-varying formation tracking with the limited energy are proposed, respectively, where the relationship matrix between the matrix variable and the limited energy is associated with the Laplacian matrix of a star graph with the leader being the central vertex. Moreover, the time-varying formation tracking criteria are extended into the time-invariant ones by simplifying the different formation feasible condition.

The rest part of the current paper is arranged as follows. The communication topology of a leader-following swarm system is depicted and the problem description of formation tracking for a swarm system with the limited energy is revealed in Section 2. Sufficient conditions for formation tracking design and analysis are proposed in Section 3. In Section 4, two numerical examples are illustrated to demonstrate time-varying and time-invariant formation criteria, respectively. Finally, Section 5 summarizes the key features of our conclusions.

1.1. Notations. The symbols R^n and $R^{n\times n}$ denote the n-dimensional real column vector and the n-dimensional real matrix space, respectively. The symbol $\mathbf{1}_N$ represents the N-dimensional column vector with all components 1. The notation $\mathbf{0}$ denotes the zero vector or zero matrix with compatible dimensions, and the matrix I_r means an identity matrix with dimension r. The symbol $\otimes$ represents the Kronecker product. The notation $Q^T = Q > 0$ shows that the matrix Q is symmetric and positive definite.

2. Problem Description

2.1. Modeling Communication Topology. For a swarm system with N homogenous agents and a leader-following communication structure, the fixed topology can be described by a weighed directed graph $G = (V(G), E(G))$, where $V(G) = \{v_0, v_1, \ldots, v_N\}$ stands for the vertex set with the element $v_k (k \in \{0, 1, \ldots, N\})$ representing agent k and $E(G) = \{e_{ki} = (v_k, v_i)\}$ denotes the edge set with the element e_{ki} denoting the communication channel between agent k and agent i. Without loss of generality, agent 0 is set as the leader and agent k $(k = 1, 2, \ldots, N)$ are followers. Furthermore, it is assumed that the communication channels

from the leader to the followers are directional and the communication channels among followers are bidirectional, which means that the leader does not impacted by the followers and only some followers can obtain the collaborative information of the leader.

The symbol $N_k = \{i\colon (v_i, v_k) \in E(G)\}$ is used to represent the neighboring agent set of agent k, where each neighboring agent is called a neighbor. The symbol w_{ik} is applied to represent the communication weight between agents k and i, where $w_{ik} = 0$ if agents k and i are not connected, $w_{ik} > 0$ if agents k and i are connected. Especially, it is assumed that there does not exist self-loop; that is, $w_{kk} = 0$ $(k = 0, 1, \ldots, N)$. The matrix $L = [l_{ik}] \in R^{(N+1)\times(N+1)}$ is applied to denote the Laplacian matrix of the communication topology, where $l_{kk} = \sum_{i\in N_k} w_{ki}$ and $l_{ik} = -w_{ik}\,(k \neq i)$. The row sum of the Laplacian matrix is equal to zero; that is, $L\mathbf{1}_N = \mathbf{0}$. Moreover, it is supposed that the fixed communication topology has a spanning tree. In this case, zero is the simple eigenvalue of the Laplacian matrix and all the other eigenvalues are positive. One can find more basic concepts and conclusions on algebraic graph theory in [30].

2.2. Modeling Formation Control. The dynamics of each agent is modeled as the following high-order linear time-invariant system:

$$\begin{cases} \dot{x}_0(t) = Ax_0(t), \\ \dot{x}_k(t) = Ax_k(t) + Bu_k(t), \end{cases} \tag{1}$$

where $k = 1, 2, \ldots, N$, $A \in R^{n\times n}$, $B \in R^{n\times m}$, $x_0(t)$ is the collaborative state of the leader, and $x_k(t)$ and $u_k(t)$ represent the collaborative state and the control input, respectively. Furthermore, a vector-valued function $\eta(t) = [\eta_1^T(t), \eta_2^T(t), \ldots, \eta_N^T(t)]^T$ is applied to design a specific geometric structure for the followers to achieve and maintain, where the element $\eta_k(t)$ is the formation function of agent k $(k \in \{1, 2, \ldots, N\})$, which has the piecewise continuous differentiable property. The formation structure is time-varying if the vector-valued function $\eta(t)$ is time-varying, and the formation structure is fixed if the derivative of the vector-valued function $\eta(t)$ is zero.

In the sequel, a new formation tracking protocol with an energy integral term and a fixed communication topology is presented as

$$\begin{cases} u_k(t) = u_{k0}(t) + u_{kf}(t), \\ u_{k0}(t) = w_{k0}K\left(x_0(t) - x_k(t) + \eta_k(t)\right), \\ u_{kf}(t) = K\sum\limits_{i\in N_k} w_{ki}\left(x_i(t) - \eta_i(t) - x_k(t) + \eta_k(t)\right), \\ J_e = \sum\limits_{k=1}^{N}\int_0^{+\infty} u_k^T(t)Qu_k(t)\mathrm{d}t, \end{cases} \tag{2}$$

where $k = 1, 2, \ldots, N$, $K \in R^{m\times n}$, $Q^T = Q > 0$, and J_e denotes the practical energy consumption of swarm system (1) as a whole. Actually, the control input $u_k(t)$ contains two components $u_{k0}(t)$ and $u_{kf}(t)$, where the component $u_{k0}(t)$ denotes the impacts of the leader on the control input of agent k and the component $u_{kf}(t)$ represents the impacts of neighboring followers on the control input of agent k.

Let $J_{\max}$ be the energy supply of swarm system (1) as a whole; that is, the energy supply is limited. Now, we, respectively, give the analysis and design definitions of the limited-energy formation tracking of swarm system (1) with tracking protocol (2) as follows.

Definition 1. For any given $J_{\max} > 0$ and the control gain K, swarm system (1) with tracking protocol (2) is said to be limited-energy formation tracking achievable if $\lim_{t\longrightarrow+\infty}(x_k(t) - \eta_k(t) - x_0(t)) = \mathbf{0}$ $(k = 1, 2, \ldots, N)$ and $J_e \leq J_{\max}$ for any bounded disagreement initial conditions $x_k(0) - \eta_k(0)$ $(k = 1, 2, \ldots, N)$.

Definition 2. For any given $J_{\max} > 0$, swarm system (1) is said to be limited-energy formation tracking achievable by tracking protocol (2) if there exists a gain matrix K such that $\lim_{t\longrightarrow+\infty}(x_k(t) - \eta_k(t) - x_0(t)) = \mathbf{0}$ $(k = 1, 2, \ldots, N)$ and $J_e \leq J_{\max}$ for any bounded disagreement initial conditions $x_k(0) - \eta_k(0)$ $(k = 1, 2, \ldots \infty, N)$.

The key objective of the current paper is to give design and analysis criteria for limited-energy formation tracking of swarm system (1) with tracking protocol (2), where both time-varying formation and time-invariant formation are considered.

Remark 1. Tracking protocol (2) has three new properties. The first one is that tracking protocol (2) has an energy integral term associated with all following agents to guarantee that the practical energy consumption is less than or equal to the limited energy. In this case, it is the key challenge to introduce the limited energy into the dynamics of the whole system and to determine the constrained relationship between the energy supply and the control gain. The second one is that the impacts of the leader and neighboring followers are given, respectively, which can be used to decompose the collaborative state of the leader from the whole dynamics of swarm system (1) with tracking protocol (2). The third one is that the formation functions of neighboring agents are involved in tracking protocol (2), which can determine arbitrary piecewise continuous differentiable structures among followers to track the collaborative state of

the leader, but it should be pointed out that the formation feasibility is dependent on the dynamics of each agent.

3. Main Results

By the linear matrix inequality, this section, respectively, presents sufficient conditions for limited-energy formation tracking design and analysis of swarm system (1) with tracking protocol (2) and time-varying geometric structures. Then, limited-energy time-varying formation tracking criteria are extended into time-invariant formation tracking cases with limited energy.

Let $\zeta_0(t) = x_0(t) - \eta_0(t)$ with $\eta_0(t) \equiv \mathbf{0}$ and $\zeta_k(t) = x_k(t) - \eta_k(t)$, $(k = 1, 2, \ldots, N)$, then one can obtain by tracking protocol (2) that

$$u_k(t) = u_{k0}(t) + u_{kf}(t), \tag{3}$$

where $k = 1, 2, \ldots, N$, and

$$\begin{aligned} u_{k0}(t) &= w_{k0} K(\zeta_0(t) - \zeta_k(t)), \\ u_{kf}(t) &= K \sum_{i \in N_k} w_{ki}(\zeta_i(t) - \zeta_k(t)). \end{aligned} \tag{4}$$

Thus, the dynamics of follower k can be rewritten as

$$\dot{\zeta}_k(t) = A(\zeta_k(t) + \eta_k(t)) + Bu_k(t) - \dot{\eta}_k(t), \quad (k = 1, 2, \ldots, N). \tag{5}$$

Let $\zeta(t) = [\zeta_0^T(t), \zeta_1^T(t), \ldots, \zeta_N^T(t)]^T$ and $\widetilde{\eta}(t) = [\eta_0^T(t), \eta_1^T(t), \ldots, \eta_N^T(t)]^T$, and L_{ff} denotes the Laplacian matrix of the communication topology among followers; then, one can find by (3) and (5) that

$$\dot{\zeta}(t) = (I_{N+1} \otimes A - L \otimes BK)\zeta(t) + (I_{N+1} \otimes A)\widetilde{\eta}(t) - \dot{\widetilde{\eta}}(t), \tag{6}$$

where

$$\begin{aligned} L &= \begin{bmatrix} 0 & \mathbf{0} \\ -l_{fl} & L_{ff} + \Lambda_{fl} \end{bmatrix}, \\ l_{fl} &= [w_{10}, w_{20}, \ldots, w_{N0}]^T, \\ \Lambda_{fl} &= \operatorname{diag}\{w_{10}, w_{20}, \ldots, w_{N0}\}. \end{aligned} \tag{7}$$

Let $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ denote nonzero eigenvalues of the Laplacian matrix L, which are also the eigenvalues of the matrix $L_{ff} + \Lambda_{fl}$ by the structure of the Laplacian matrix L. By the linear matrix inequality technique, the following theorem presents a sufficient condition of limited-energy formation tracking design; that is, a design method of the gain matrix K is proposed to make swarm system (1) with tracking protocol (2) and a fixed communication topology achieve limited-energy formation tracking.

Theorem 1. *For any given $J_{\max} > 0$, if $A\eta_k(t) = \dot{\eta}_k(t)$, $(k = 1, 2, \ldots, N)$ and there exists $\widehat{P}^T = \widehat{P} > 0$ such that*

$$\begin{aligned} &\widehat{P}A^T + A\widehat{P} - 2BB^T + \lambda_1^{-2}\lambda_N^2 BQB^T < 0, \\ &(x(0) - \eta(0))^T \left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes I_n \right) (x(0) - \eta(0)) I_n \le J_{\max}\widehat{P}, \end{aligned} \tag{8}$$

then swarm system (1) is limited-energy time-varying formation tracking achievable by tracking protocol (2) with $K = \lambda_1^{-1} B^T \widehat{P}$.

Proof of Theorem 1. A nonsingular matrix is introduced as follows:

$$W = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{1}_N & I_N \end{bmatrix}. \tag{9}$$

Whose invertible matrix is

$$W^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ -\mathbf{1}_N & I_N \end{bmatrix}. \tag{10}$$

It can be shown that $\Lambda_{ff}\mathbf{1}_N = l_{fl}$, so one can deduce that

$$\begin{aligned} W^{-1}LW &= \begin{bmatrix} 0 & \\ & L_{ff} + \Lambda_{fl} \end{bmatrix}, \\ (W^{-1} \otimes I_n)\zeta(t) &= \left[x_0^T(t), \widetilde{\zeta}_1^T(t), \ldots, \widetilde{\zeta}_N^T(t) \right]^T, \end{aligned} \tag{11}$$

where $\widetilde{\zeta}_k(t) = \zeta_k(t) - x_0(t)$ $(k = 1, 2, \ldots, N)$. Because the communication channels among followers are bidirectional, the Laplacian matrix L_{ff} is symmetric. Since it is assumed that the whole communication topology has a spanning tree, the matrix Λ_{fl} is nonzero. In this case, the matrix $L_{ff} + \Lambda_{fl}$ is symmetric and positive definite. Thus, one can find an orthonormal matrix $\widehat{W} \in R^{N \times N}$ such that $\widehat{W}^T(L_{ff} + \Lambda_{fl})\widehat{W} = \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$, where $\lambda_k > 0$, $(k = 1, 2, \ldots, N)$ are nonzero eigenvalues of the Laplacian matrix L of the whole communication topology. Let $\widetilde{\zeta}(t) = [\widetilde{\zeta}_1^T(t), \widetilde{\zeta}_2^T(t), \ldots, \widetilde{\zeta}_N^T(t)]^T$ and $\widehat{\zeta}(t) = (\widehat{W}^T \otimes I_n)\widetilde{\zeta}(t) = [\widehat{\zeta}_1^T(t), \widehat{\zeta}_2^T(t), \ldots, \widehat{\zeta}_N^T(t)]^T$, then swarm system (1) with tracking protocol (2) can be transformed by (6) into

$$\begin{aligned} \dot{x}_0(t) &= Ax_0(t), \\ \dot{\widehat{\zeta}}_k(t) &= (A - \lambda_k BK)\widehat{\zeta}_k(t) + \left(e_k^T \widehat{W}^T [0, I_N] W^{-1} \otimes A \right) \widetilde{\eta}(t) - \left(e_k^T \widehat{W}^T [0, I_N] W^{-1} \otimes I_n \right) \dot{\widetilde{\eta}}(t), \end{aligned} \tag{12}$$

where $k = 1, 2, \ldots, N$ and the column vector e_k is N-dimensional and its k-th component is 1 and zero elsewhere. Because $\widetilde{\zeta}_k(t) = \zeta_k(t) - x_0(t)$ $(k = 1, 2, \ldots, N)$ and $\widehat{W}$ is orthonormal, it can be found that if $\lim_{t \longrightarrow +\infty} \widehat{\zeta}_k(t) = \mathbf{0}$ $(k = 1, 2, \ldots, N)$, then one can obtain that

$$\lim_{t\longrightarrow\infty}\left(x_k(t)-\eta_k(t)-x_0(t)\right)=0,\quad (k=1,2,\ldots,N), \tag{13}$$

which means that swarm system (1) with tracking protocol (2) achieves formation tracking.

Furthermore, we give an approach to design K such that $\lim_{t\longrightarrow+\infty}\widehat{\zeta}_k(t)=\mathbf{0},\quad (k=1,2,\ldots,N)$. Let $P^T=P>0$, then the following Lyapunov function candidate is adopted:

$$V_k(t)=\widehat{\zeta}_k^T(t)P\widehat{\zeta}_k(t),\quad (k=1,2,\ldots,N). \tag{14}$$

By taking the time derivative of $V_k(t)$, it can be derived that

$$\begin{aligned}\dot{V}_k(t)&=\widehat{\zeta}_k^T(t)\left(\left(A-\lambda_kBK\right)^TP+P\left(A-\lambda_kBK\right)\right)\widehat{\zeta}_k(t),\\&+2\widehat{\zeta}_k^T(t)P\left(e_k^T\widehat{W}^T[\mathbf{0},I_N]W^{-1}\otimes A\right)\widetilde{\eta}(t),\\&-2\widehat{\zeta}_k^T(t)P\left(e_k^T\widehat{W}^T[\mathbf{0},I_N]W^{-1}\otimes I_n\right)\dot{\widetilde{\eta}}(t).\end{aligned} \tag{15}$$

Let $K=\lambda_1^{-1}B^TP$, then one has

$$\left(A-\lambda_kBK\right)^TP+P\left(A-\lambda_kBK\right)\le A^TP+PA-2\lambda_1^{-1}\lambda_kPBB^TP. \tag{16}$$

Since the nonzero eigenvalues of the Laplacian matrix L satisfy that $0<\lambda_1\le\lambda_2\le\cdots\le\lambda_N$, one can deduce that $\lambda_1^{-1}\lambda_k\ge 1$ $(k=1,2,\ldots,N)$. Therefore, it can be derived by (16) that

$$A^TP+PA-2\lambda_1^{-1}\lambda_kPBB^TP\le A^TP+PA-2PBB^TP. \tag{17}$$

Due to

$$\begin{aligned}e_k^T\widehat{W}^T[\mathbf{0},I_N]W^{-1}\otimes A&=\left(e_k^T\widehat{W}^T[\mathbf{0},I_N]W^{-1}\otimes I_n\right)(I_{N+1}\otimes A),\\A\eta_k(t)-\dot{\eta}_k(t)&=0,\quad (k=1,2,\ldots,N),\end{aligned} \tag{18}$$

one can obtain that

$$\left(e_k^T\widehat{W}^T[\mathbf{0},I_N]W^{-1}\otimes A\right)\widetilde{\eta}(t)-\left(e_k^T\widehat{W}^T[\mathbf{0},I_N]W^{-1}\otimes I_n\right)\dot{\widetilde{\eta}}(t)=\mathbf{0}. \tag{19}$$

By (15), (17), and (19), if

$$A^TP+PA-2PBB^TP<0, \tag{20}$$

then one can deduce that

$$\lim_{t\longrightarrow\infty}\widehat{\zeta}_k(t)=0,\quad (k=1,2,\ldots,N). \tag{21}$$

That is, swarm system (1) with tracking protocol (2) achieves formation tracking. In the other work, swarm system (1) is formation tracking achievable by tracking protocol (2) with $K=\lambda_1^{-1}B^TP$.

In the sequel, the impacts of the limited energy are dealt with. One can show that

$$\sum_{k=1}^N u_k^T(t)Qu_k(t)=\widetilde{\zeta}^T(t)\left(\left(L_{ff}+\Lambda_{fl}\right)^2\otimes K^TQK\right)\widetilde{\zeta}(t). \tag{22}$$

Due to $\widehat{\zeta}(t)=(\widehat{W}^T\otimes I_n)\widetilde{\zeta}(t)$, one can deduce that

$$\widetilde{\zeta}^T(t)\left(\left(L_{ff}+\Lambda_{fl}\right)^2\otimes K^TQK\right)\widetilde{\zeta}(t)=\sum_{k=1}^N\lambda_k^2\widehat{\zeta}_k^T(t)K^TQK\widehat{\zeta}_k(t). \tag{23}$$

Since $K=\lambda_1^{-1}B^TP$, it can be found by (22) and (23) that

$$\sum_{k=1}^N u_k^T(t)Qu_k(t)=\lambda_1^{-2}\sum_{k=1}^N\lambda_k^2\widehat{\zeta}_k^T(t)PBQB^TP\widehat{\zeta}_k(t). \tag{24}$$

Let $\kappa\ge 0$, then one can see that

$$\begin{aligned}J_e^\kappa&=\int_0^\kappa\widetilde{\zeta}^T(t)\left(\left(L_{ff}+\Lambda_{fl}\right)^2\otimes K^TQK\right)\widetilde{\zeta}(t)\mathrm{d}t,\\&=\lambda_1^{-2}\sum_{i=1}^N\int_0^\kappa\lambda_k^2\hat{\zeta}_k^T(t)PBQB^TP\hat{\zeta}_k(t)\mathrm{d}t.\end{aligned} \tag{25}$$

According to $\int_0^\kappa\dot{V}_k(t)\mathrm{d}t=V_k(\kappa)-V_k(0)$, from (22) to (25), one can find that

$$\begin{aligned}J_e^\kappa&=\sum_{k=1}^N\int_0^\kappa\left(\left(\dot{V}_k(t)+\lambda_1^{-2}\lambda_k^2\widehat{\zeta}_k^T(t)PBQB^TP\widehat{\zeta}_k(t)\right)\mathrm{d}t-V_k(\kappa)\right)\\&+\sum_{k=1}^N V_k(0).\end{aligned} \tag{26}$$

Because λ_N is the maximum nonzero eigenvalue of the Laplacian matrix L of the whole communication topology, it can be obtained by (26) that

$$\begin{aligned}J_e^\kappa&\le\sum_{k=1}^N\int_0^\kappa\left(\left(\dot{V}_k(t)+\lambda_1^{-2}\lambda_N^2\widehat{\zeta}_k^T(t)PBQB^TP\widehat{\zeta}_k(t)\right)\right)\mathrm{d}t-V_k(\kappa)\right)\\&+\sum_{k=1}^N V_k(0)\cdot\sqrt{b^2-4ac}.\end{aligned} \tag{27}$$

Thus, one can find that if

$$A^TP+PA-2PBB^TP+\lambda_1^{-2}\lambda_N^2PBQB^TP<0, \tag{28}$$

then one has $\lim_{t\longrightarrow+\infty}\widehat{\zeta}_k(t)=\mathbf{0}$ and $\lim_{\kappa\longrightarrow+\infty}V_k(\kappa)=\mathbf{0},\quad (k=1,2,\ldots,N)$. In this case, one can deduce by (27) that

$$J_e\le\sum_{k=1}^N\widehat{\zeta}_k^T(0)P\widehat{\zeta}_k(0). \tag{29}$$

Since $\widehat{\zeta}_k(0)=e_k^T(\widehat{W}^T\otimes I_n)\widetilde{\zeta}(0),\quad (k=1,2,\ldots,N)$, one has

$$\sum_{k=1}^N\widehat{\zeta}_k^T(0)P\widehat{\zeta}_k(0)=\widetilde{\zeta}^T(0)\left(I_N\otimes P\right)\widetilde{\zeta}(0). \tag{30}$$

Due to

$$\tilde{\zeta}(0) = \left([\mathbf{0}, I_N]W^{-1} \otimes I_n\right)\zeta(0),$$
$$W^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ -\mathbf{1}_N & I_N \end{bmatrix}. \tag{31}$$

It can be found that

$$\left([\mathbf{0}, I_N]W^{-1}\right)^T\left([\mathbf{0}, I_N]W^{-1}\right) = \begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix}. \tag{32}$$

Thus, one can find by (30) that

$$\sum_{k=1}^{N} \tilde{\zeta}_k^T(0)P\tilde{\zeta}_k(0) = \tilde{\zeta}^T(0)\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes P\right)\tilde{\zeta}(0). \tag{33}$$

Because it is supposed that $x_k(0) - \eta_k(0)$, $(k = 1, 2, \ldots, N)$ are disagreement, there must exist $\tilde{\zeta}_k(0)$, $(k \in \{1, 2, \ldots, N\})$ is nonzero. In this case, one can show that

$$\zeta^T(0)\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes I_n\right)\zeta(0) = \sum_{k=1}^{N} \tilde{\zeta}_k^T(0)\tilde{\zeta}_k(0) > 0. \tag{34}$$

Hence, there exists $\gamma > 0$ such that

$$J_{\max} = \zeta^T(0)\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes \gamma I_n\right)\zeta(0). \tag{35}$$

Since the matrix $\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix}$ has a simple zero eigenvalue and $N-1$ positive eigenvalues, one can find by (33) and (35) that $P \le \gamma I_n$ can ensure that $J_e \le J_{\max}$; that is, it is required that

$$(x(0) - \eta(0))^T\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes I_n\right)(x(0) - \eta(0))P \le J_{\max} I_n. \tag{36}$$

Let $\hat{P} = P^{-1}$, then the conclusion of Theorem 1 can be deduced

Theorem 1 gives a sufficient condition for limited-energy formation design, which proposes a design method of the gain matrix of tracking protocol (2). For the case that the gain matrix is given, by the Schur lemma in [31] and the convex property of the linear matrix inequality, the following theorem can be obtained directly on the basis of the Proof of Theorem 1, which presents a sufficient condition for limited-energy formation tracking analysis. □

Theorem 2. *For any given $J_{\max} > 0$ and the gain matrix K, if $A\eta_k(t) = \dot{\eta}_k(t)$, $(k = 1, 2, \ldots, N)$ and there exists $P^T = P > 0$ such that*

$$\begin{bmatrix} A^TP + PA - \lambda_k PBK - \lambda_k K^T B^T P & \lambda_N K^T Q \\ \lambda_N QK & -Q \end{bmatrix} < 0, \quad (k = 1, N),$$
$$(x(0) - \eta(0))^T\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes I_n\right)(x(0) - \eta(0))P \le J_{\max} I_n, \tag{37}$$

then swarm system (1) with tracking protocol (2) achieves limited-energy time-varying formation tracking.

Moreover, if the formation structure among followers is time-invariant, then one can obtain the following two conclusions by the above analysis, which present limited-energy formation tracking design and analysis criteria, respectively.

Theorem 3. *For any given $J_{\max} > 0$, if $A\eta_k(t) = \mathbf{0}$, $\dot{\eta}_k(t) = \mathbf{0}$, $(k = 1, 2, \ldots, N)$ and there exists $\hat{P}^T = \hat{P} > 0$ such that*

$$\hat{P}A^T + A\hat{P} - 2BB^T + \lambda_1^{-2}\lambda_N^2 BQB^T < 0,$$
$$(x(0) - \eta(0))^T\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes I_n\right)(x(0) - \eta(0))I_n \le J_{\max}\hat{P}, \tag{38}$$

then swarm system (1) is limited-energy time-invariant formation tracking achievable by tracking protocol (2) with $K = \lambda_1^{-1}B^T\hat{P}^{-1}$.

Theorem 4. *For any given $J_{\max} > 0$, the gain matrix K, if $A\eta_k(t) = \mathbf{0}$, $\dot{\eta}_k(t) = \mathbf{0}$, $(k = 1, 2, \ldots, N)$ and there exists $P^T = P > 0$ such that*

$$\begin{bmatrix} A^TP + PA - \lambda_k PBK - \lambda_k K^T B^T P & \lambda_N K^T Q \\ \lambda_N QK & -Q \end{bmatrix} < 0, \quad (k = 1, N),$$
$$(x(0) - \eta(0))^T\left(\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix} \otimes I_n\right)(x(0) - \eta(0))P \le J_{\max} I_n, \tag{39}$$

then swarm system (1) with tracking protocol (2) achieves limited-energy time-invariant formation tracking.

Remark 2. By choosing different formation functions with the piecewise continuous differentiable property, the different formation structures can be achieved for followers in swarm systems with leader-following structures to maintain. From Theorems 1 to 4, it can be found that some formation structures may be unfeasible, and the feasible property depends on the dynamics of each follower; that is, the feasible condition $A\eta_k(t) = \dot{\eta}_k(t), \quad (k = 1, 2, \ldots, N)$ must be satisfied for time-varying formation cases and the feasible conditions $A\eta_k(t) = \mathbf{0}$ and $\dot{\eta}_k(t) = \mathbf{0}, \quad (k = 1, 2, \ldots, N)$ should hold for time-invariant formation cases. Furthermore, the relationship matrix $\begin{bmatrix} N & -\mathbf{1}_N^T \\ -\mathbf{1}_N & I_N \end{bmatrix}$ between the limited energy and the matrix variable essentially is the Laplacian matrix of a star graph with all edge weights one, where the leader is the central vertex and followers are not connected with each other. This is coincident with the leader-following communication structure of the whole swarm system. Especially, sufficient conditions in Theorems 1 to 4 include the minimum and maximum nonzero eigenvalues of the Laplacian matrix, whose precise values are difficult to be solved, but they can be estimated by the proposed methods in [32, 33], respectively. It should be pointed out that sufficient conditions in Theorems 1 to 4 can be checked by the FEASP solver of the Matlab's LMI toolbox in [34].

4. Numerical Examples

This section presents two numerical examples to demonstrate the effectiveness of main conclusions about limited-energy formation tracking for swarm systems with fixed communication topologies.

Example 1 (time-varying formation tracking). Consider a swarm system with six agents and a leader-following communication structure, where the dynamics of each agent is described by (1) with

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 2 & 3 \\ -2 & -2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ -1 & -1 \\ 0 & 0 \end{bmatrix}. \tag{40}$$

The communication topology with a leader-following structure is revealed by Figure 1, where agent 0 is the leader, agents 1 to 5 are followers, and the edge weight is set to be 0-1; that is, the weight of the connected edge is one and the weight of the unconnected edge is zero. The initial states of this swarm system are given as follows:

$$\begin{aligned} x_0(0) &= [1.50, 0.40, -3.20]^T, \\ x_1(0) &= [-1.51, 1.48, -6.85]^T, \\ x_2(0) &= [6.15, 0.73, 7.92]^T, \\ x_3(0) &= [5.99, 6.66, 2.62]^T, \\ x_4(0) &= [-4.38, -4.23, -0.37]^T, \\ x_5(0) &= [2.04, -3.27, 2.51]^T. \end{aligned} \tag{41}$$

The formation structure among followers is determined by

$$\eta_k(t) = \begin{bmatrix} \sin\left(t + \frac{2(k-1)\pi}{5}\right) - \cos\left(t + \frac{2(k-1)\pi}{5}\right) \\ -3\sin\left(t + \frac{2(k-1)\pi}{5}\right) - 3\cos\left(t + \frac{2(k-1)\pi}{5}\right) \\ 2\sin\left(t + \frac{2(k-1)\pi}{5}\right) + 2\cos\left(t + \frac{2(k-1)\pi}{5}\right) \end{bmatrix}, \quad (k = 1, 2, \ldots, 5). \tag{42}$$

It is clear that the formation functions satisfy that $A\eta_k(t) = \dot{\eta}_k(t), \quad (k = 1, 2, \ldots, 5)$ in Theorem 1. Let $Q = 0.1I_2$ and $J_{\max} = 40$. By the FEASP solver, one can obtain by Theorem 1 that

$$P = \begin{bmatrix} 285.68 & 299.00 & 394.00 \\ 299.00 & 357.38 & 434.39 \\ 394.00 & 434.39 & 597.76 \end{bmatrix}. \tag{43}$$

In this case, one can obtain that

$$K = \begin{bmatrix} -99.44 & -435.72 & -301.45 \\ -4364.10 & -4899.26 & -6183.20 \end{bmatrix}. \tag{44}$$

Figure 2 shows the trajectories of $\varsigma_k(t), \quad (k = 1, 2, \ldots, 5)$, where the curves formed by circle markers are the trajectories of the collaborative state of the leader.

In Figure 3, the state snapshots of the leader and five followers are shown at $t = 0$ s, $t = 16$ s, $t = 18$ s, and $t = 20$ s, where the leader is depicted by pentagrams and five

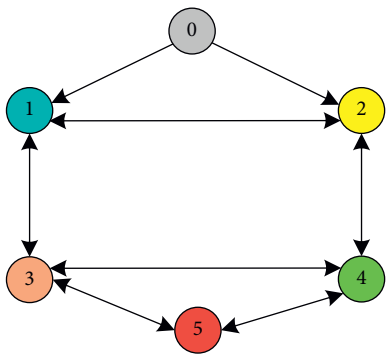


Figure 1: Communication topology for time-varying formation tracking case.

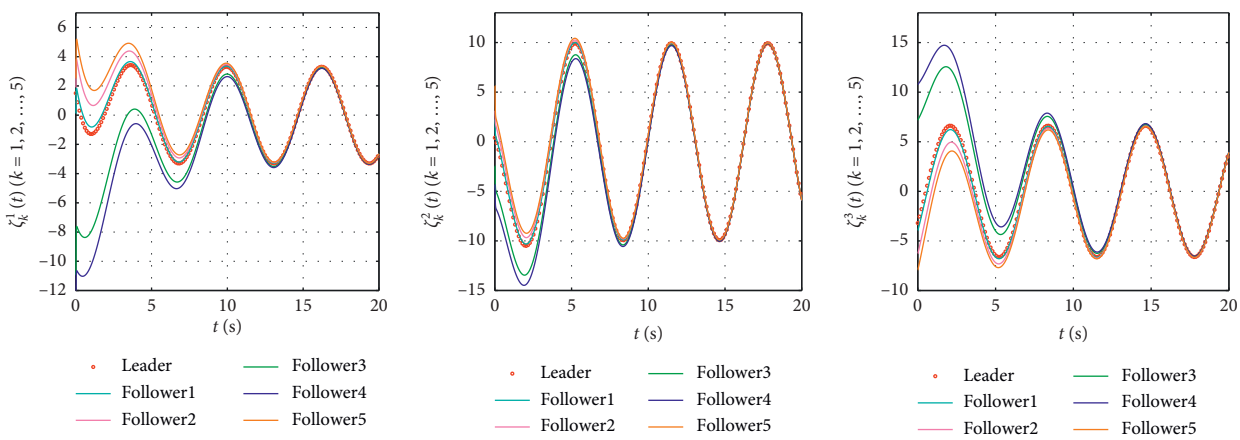


Figure 2: Trajectories of $\varsigma_k(t)$, $(k = 1, 2, \ldots, 5)$ for time-varying formation tracking case.

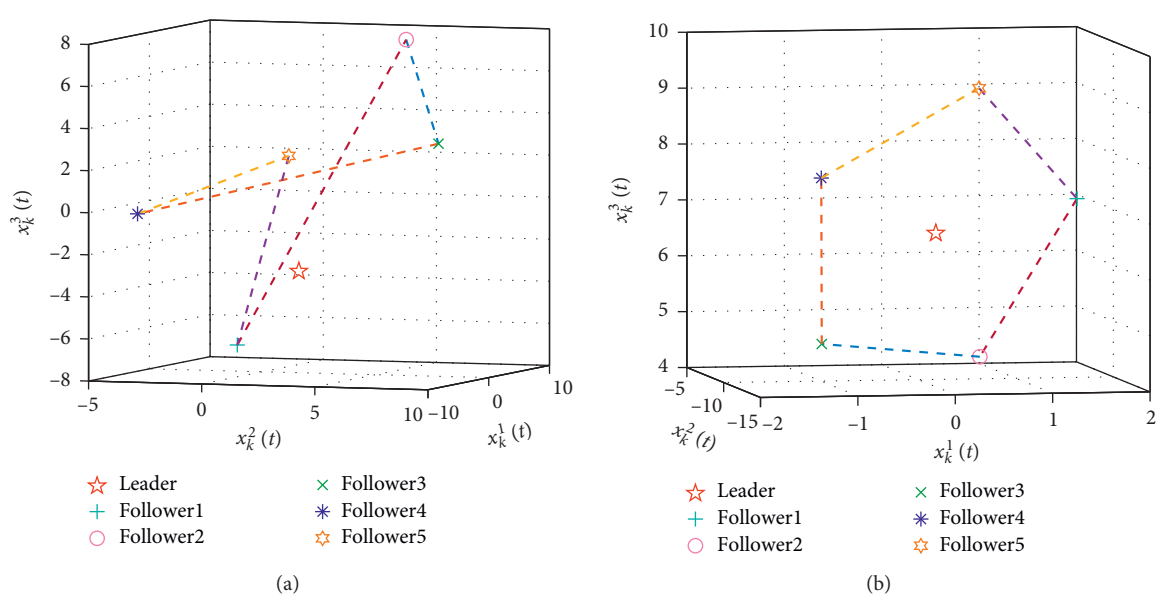


Figure 3: Continued.

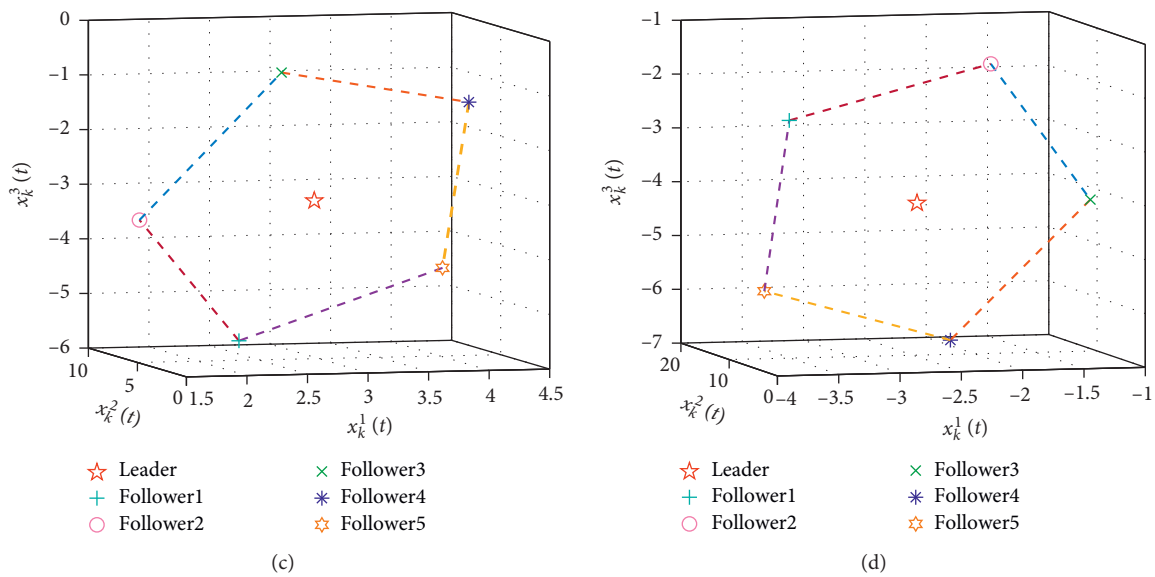


Figure 3: State snapshots of all the agents at different time for time-varying formation tracking case: (a) $t = 0$ s, (b) $t = 16$ s, (c) $t = 18$ s, and (d) $t = 20$ s.

followers are described by plusses, circles, x-marks, asterisks, and hexagrams, respectively, where it can be seen that five followers achieve a time-varying pentagon to track the collaborative state of the leader. Figure 4 reveals that the leader counterclockwise rotates about 3.18 cycles, where the initial state is represented by a blue pentagram and the final state is depicted by a black pentagram. In Figure 5, it is shown that the practical energy consumption $J_e(t)$ converges to a finite value less than $J_{\max}$. From Figures 2 to 5, it can be found that this swam system achieves limited-energy time-varying formation tracking.

Example 2 (time-invariant formation tracking). In this example, the leader-following swarm system consists of one leader and eight followers, whose communication topology is shown in Figure 6. In Figure 6, agent 0 is identified as the leader and agents 1–8 as followers, and without loss of generality, the edge weight among the nine agents is also set as 0-1.

The matrix pair (A, B) of the dynamics equation of each follower is given as

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{45}$$

where A is the system matrix of the leader. The initial values of the states for the leader are given as $x_0(0) = [3.5, -2, 2]^T$, and eight followers are initialized as

$$\begin{aligned} x_1(0) &= [12.04, 5.41, -4.98]^T, \\ x_2(0) &= [-8.11, 4.72, 3.04]^T, \\ x_3(0) &= [-5.64, -3.04, -2.63]^T, \\ x_4(0) &= [6.73, -4.68, 5.52]^T, \\ x_5(0) &= [-10.98, -2.22, -3.63]^T, \\ x_6(0) &= [-11.21, -1.17, 2.40]^T, \\ x_7(0) &= [2.95, 3.97, -4.60]^T, \\ x_8(0) &= [9.56, 2.60, 5.38]^T. \end{aligned} \tag{46}$$

The purpose of this example is to make eight followers track the leader with a time-invariant formation, which is described by a regular octagon geometric structure. The corresponding time-invariant formation function is described by

$$\eta_k(t) = \begin{bmatrix} \sqrt{2}\cos\left(\dfrac{(k-1)\pi}{4}\right) \\ \sin\left(\dfrac{(k-1)\pi}{4}\right) \\ \sin\left(\dfrac{(k-1)\pi}{4}\right) \end{bmatrix}, \quad (k = 1, 2, \ldots, 8). \tag{47}$$

One can find that the formation function $\eta_k(t)$ satisfies the condition that $A\eta_k(t) = \mathbf{0}$ and $\dot{\eta}_k(t) = 0$, $(k = 1, 2, \ldots, 8)$ in Theorem 2. Let $Q = 0.12I_2$ and $J_{\max} = 300$; then, by using the FEASP solver, it can be found by Theorem 2 that

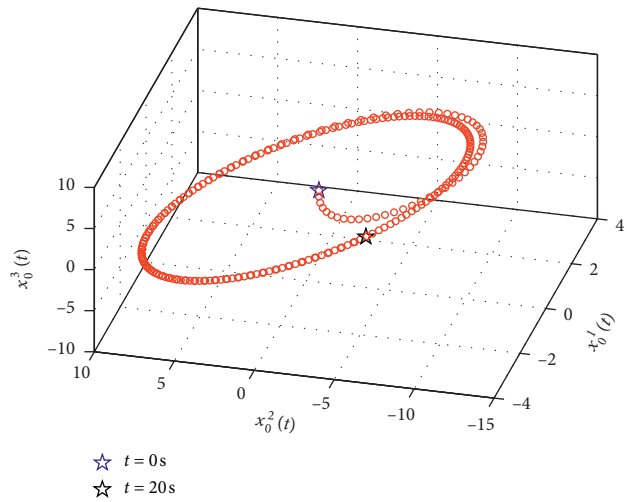


Figure 4: Movement track of the leader for time-varying formation tracking case.

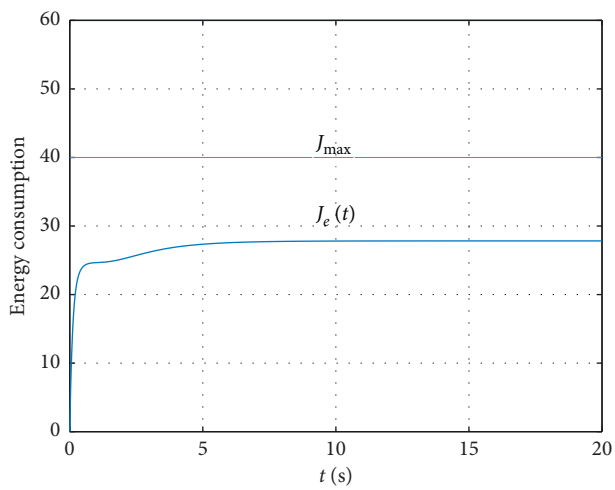


Figure 5: Energy consumption and the limited budget for time-varying formation tracking case.

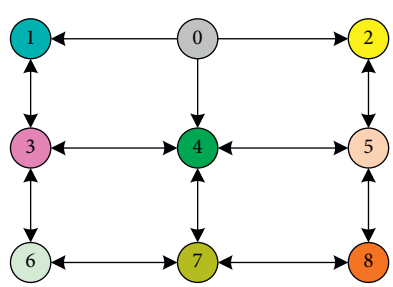


Figure 6: Communication topology for time-invariant formation tracking case.

$$P = \begin{bmatrix} 4.0614 & 4.0463 & -4.0463 \\ 4.0463 & 69.3470 & -8.0918 \\ -4.0463 & -8.0918 & 69.3470 \end{bmatrix}. \quad (48)$$

In this case, it can be deduced that

$$K = \begin{bmatrix} -1.5794 & -3.1584 & 27.0677 \\ 1.5794 & 27.0677 & -3.1584 \end{bmatrix}. \quad (49)$$

In Figure 7, the trajectories of $\varsigma_k(t)$, $(k = 1, 2, \ldots, 8)$ are depicted, where the curves formed by circle markers

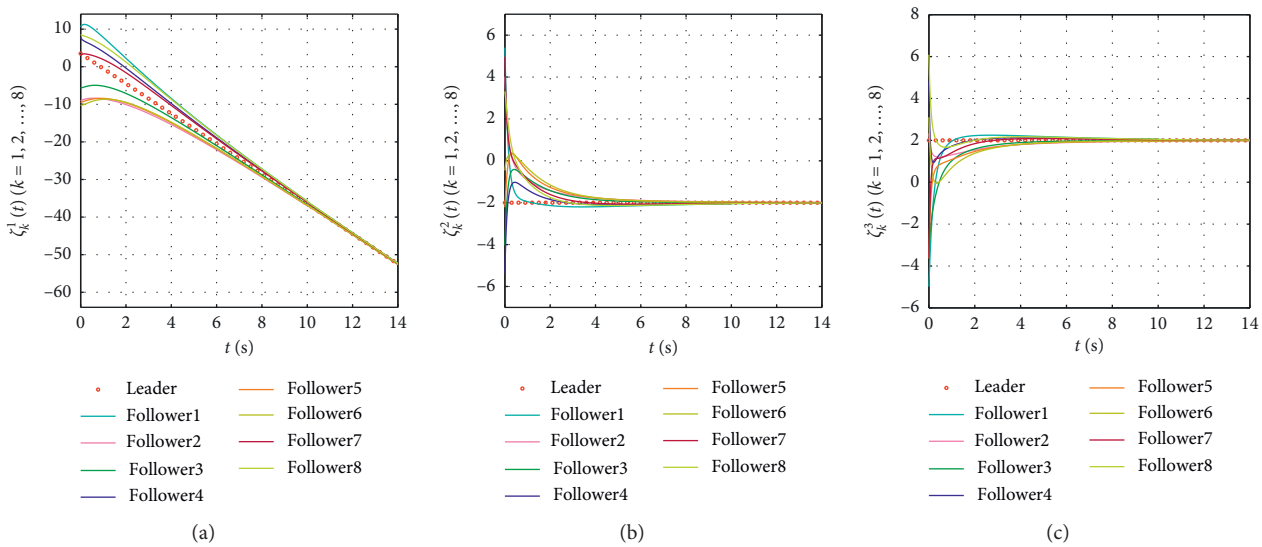


FIGURE 7: Trajectories of $\varsigma_k(t)$, $(k = 1, 2, \ldots, 8)$ for time-invariant formation tracking case.

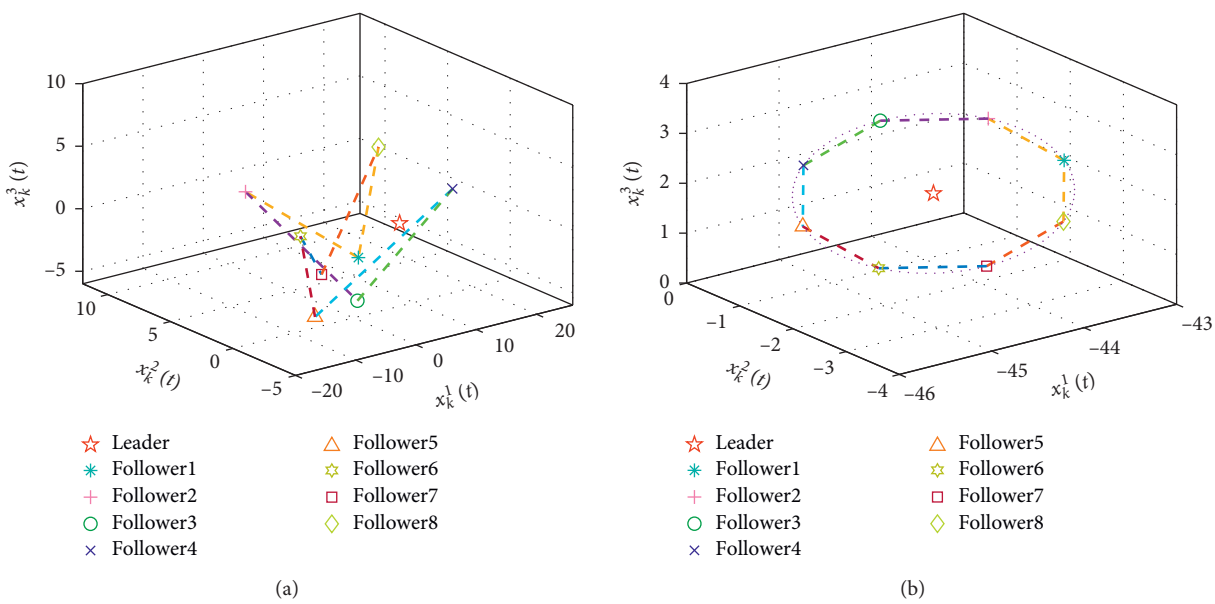


FIGURE 8: Continued.

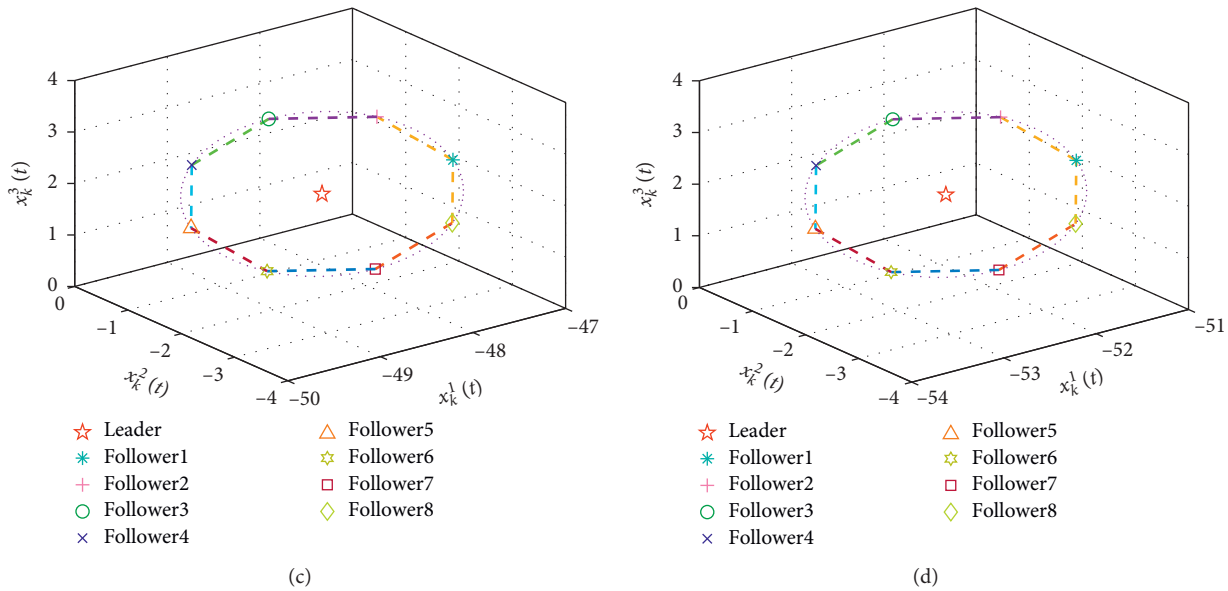


Figure 8: State snapshots of all the agents at different time for time-invariant formation tracking case: (a) $t = 0$ s, (b) $t = 12$ s, (c) $t = 13$ s, and (d) $t = 14$ s.

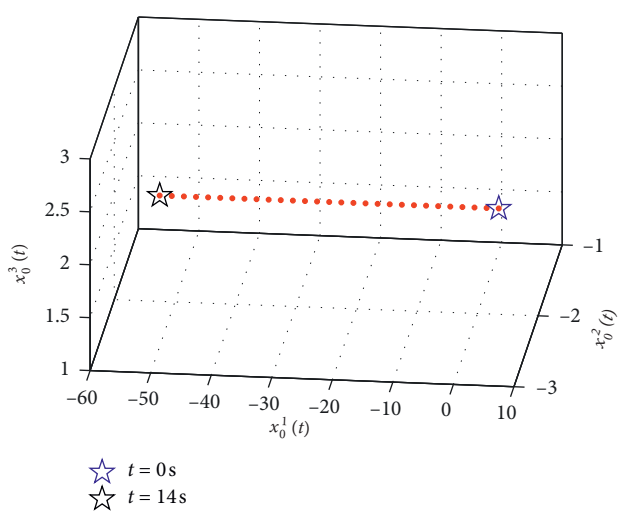


Figure 9: Movement track of the leader for time-invariant formation tracking case.

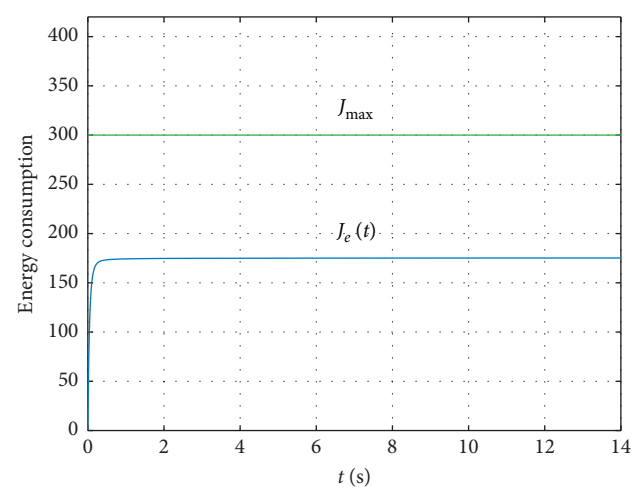


Figure 10: Energy consumption and the limited budget for time-invariant formation tracking case.

are the trajectories of the collaborative state of the leader. Figure 8 describes the state snapshots of the leader and eight followers at $t = 0$ s, $t = 12$ s, $t = 13$ s, and $t = 14$ s, and the leader and eight followers are depicted by pentagram, asterisks, plusses, circles, x-marks, triangles, hexagrams, squares, and diamonds, respectively, where eight followers achieve a regular time-invariant octagon to track the leader. Figure 9 shows that the leader moves along a straight line, where the initial and final states are depicted by a blue pentagram and a black pentagram, respectively. Figure 10 shows the trajectories of the energy consumption and the limited energy. From Figures 7 to 10, one can find that this swam system achieves limited-energy time-invariant formation tracking.

5. Conclusions

For high-order linear time-invariant swarm systems with the limited energy and fixed topologies, a new formation tracking protocol with an energy integral term was proposed to realize formation tracking under the condition that the whole energy consumption is less than or equal to the limited energy supply. By constructing some nonsingular transformation matrix, the whole dynamics of a swarm system with a leader-following topology structure was divided into two independent parts: the dynamics of the leader and the relative dynamics among all agents, and it was shown that the formation tracking is achieved if the relative dynamics is asymptotically stable. Furthermore, the limited energy was introduced into the formation tracking criteria

by the matrix variable and disagreement initial states and sufficient conditions for swam systems with the limited energy to achieve time-varying formation tracking and time-invariant formation tracking were proposed, respectively. Especially, these criteria are independent of the number of agents and only include linear matrix inequality constraints, so they are scalable and checkable; that is, the computation complexity does not increase as the number of agents of a swarm system is added. The further works will focus on extending main conclusions for swarm systems with the homogenous dynamics and the fixed topology to swarm systems with the heterogeneous dynamics and switching topologies.

Data Availability

The simulation data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Conceptualization was carried out by Jianye Yang and Xinzhong Han. Methodology was developed by Jianye Yang. Cheng Wang was responsible for formal analysis and validation. Investigation and project administration were performed by Hongtao Dang. Writing-original draft preparation and writing-review and editing were carried out by Cheng Wang and Jianye Yang. Xinzhong Han supervised the study. Cheng Wang and Xinzhong Han were involved in funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the Key Research and Development Program of Shaanxi (no. 2019GY-025).

References

[1] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.

[2] N. Sun, Y. Fu, T. Yang, J. Zhang, Y. Fang, and X. Xin, "Nonlinear motion control of complicated dual rotary crane systems without velocity feedback: design, analysis, and hardware experiments," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 2, pp. 1017–1029, 2020.

[3] J. Xi, C. Wang, H. Liu, and L. Wang, "Completely distributed guaranteed-performance consensualization for high-order multiagent systems with switching topologies," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1338–1348, 2019.

[4] J. Qu, Z. Ji, C. Lin, and H. Yu, "Fast consensus seeking on networks with antagonistic interactions," *Complexity*, vol. 2018, Article ID 7831317, 15 pages, 2018.

[5] J. Sun, Z. Geng, Y. Lv, Z. Li, and Z. Ding, "Distributed adaptive consensus disturbance rejection for multi-agent systems on directed graphs," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 202–212, 2018.

[6] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, In press.

[7] Y. Zhang, H. Li, J. Sun, and W. He, "Cooperative adaptive event-triggered control for multiagent systems with actuator failures," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1759–1768, 2019.

[8] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 2442–2454, 2020.

[9] X.-G. Guo, J.-L. Wang, F. Liao, and D. Wang, "Quantized H_∞ consensus of multiagent systems with quantization mismatch under switching weighted topologies," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 2, pp. 202–212, 2017.

[10] Z.-Y. Tan, N. Cai, J. Zhou, and S.-G. Zhang, "On performance of peer review for academic journals: analysis based on distributed parallel system," *IEEE Access*, vol. 7, pp. 19024–19032, 2019.

[11] L. Consolini, F. Morbidi, D. Prattichizzo, and M. Tosques, "Leader–follower formation control of nonholonomic mobile robots with input constraints," *Automatica*, vol. 44, no. 5, pp. 1343–1349, 2008.

[12] K.-K. Oh and H.-S. Ahn, "Formation control of mobile agents based on distributed position estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 737–742, 2013.

[13] H. Liu, T. Ma, F. L. Lewis, and Y. Wan, "Robust formation control for multiple quadrotors with nonlinearities and disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1362–1371, 2020.

[14] W. Ren, "Consensus strategies for cooperative control of vehicle formations," *IET Control Theory & Applications*, vol. 1, no. 2, pp. 505–512, 2007.

[15] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, In press.

[16] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[17] S. Liu, Z. Ji, and H. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, Article ID 7685460, 18 pages, 2020.

[18] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1465–1476, 2004.

[19] H. Du, G. Wen, Y. Cheng, Y. He, and R. Jia, "Distributed finite-time cooperative control of multiple high-order nonholonomic mobile robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2998–3006, 2016.

[20] W. Qin, Z. Liu, and Z. Chen, "A novel observer-based formation for nonlinear multi-agent systems with time delay and intermittent communication," *Nonlinear Dynamics*, vol. 79, no. 3, pp. 1651–1664, 2015.

[21] L. Brinon-Arranz, A. Seuret, and C. Canudas-de-Wit, "Cooperative control design for time-varying formations of multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2283–2288, 2014.

[22] R. Rahimi, F. Abdollahi, and K. Naqshi, "Time-varying formation control of a collaborative heterogeneous multi agent

system," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1799–1805, 2014.

[23] X. Dong and G. Hu, "Time-varying formation control for general linear multi-agent systems with switching directed topologies," *Automatica*, vol. 73, no. 73, pp. 47–55, 2016.

[24] F. Xiao, L. Wang, J. Chen, and Y. Gao, "Finite-time formation control for multi-agent systems," *Automatica*, vol. 45, no. 11, pp. 2605–2611, 2009.

[25] X. Dong, Y. Zhou, Z. Ren, and Y. Zhong, "Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5014–5024, 2017.

[26] X. Dong and G. Hu, "Time-varying formation tracking for linear multiagent systems with multiple leaders," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3658–3664, 2017.

[27] J. Xi, Z. Fan, H. Liu, and T. Zheng, "Guaranteed-cost consensus for multiagent networks with Lipschitz nonlinear dynamics and switching topologies," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 7, pp. 2841–2852, 2018.

[28] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[29] J. Yu, X. Dong, Q. Li, and Z. Ren, "Robust H_∞ guaranteed cost time-varying formation tracking for high-order multiagent systems with time-varying delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, pp. 1465–1475, 2020.

[30] C. Godsil and G. Royal, *Algebraic Graph Theory*, Springer, New York, NY, USA, 2001.

[31] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, PA, USA, 1994.

[32] A. Berman and X. D. Zhang, "Lower bounds for the eigenvalues of Laplacian matrices," *Linear Algebra and Its Applications*, vol. 316, no. 1–3, pp. 13–20, 2000.

[33] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1990.

[34] P. Gahinet, A. Nemirovskii, A. J. Laub, and M. Chilali, *LMI Control Toolbox User's Guide*, The Math Works, Natick, MA, USA, 1995.

Hindawi
Complexity
Volume 2020, Article ID 8844459, 9 pages
https://doi.org/10.1155/2020/8844459

Research Article

Classical Solutions to the Initial-Boundary Value Problems for Nonautonomous Fractional Diffusion Equations

Jia Mu,[1,2] **Yang Liu**,[2] **and Huanhuan Zhang**[2]

[1]*Key Laboratory of Streaming Data Computing Technologies and Application, Northwest Minzu University, Lanzhou 730000, China*
[2]*School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730000, China*

Correspondence should be addressed to Jia Mu; mujia88@163.com

Received 14 August 2020; Revised 12 September 2020; Accepted 29 September 2020; Published 13 October 2020

Academic Editor: Yi-Sheng Lv

In this paper, we investigate a class of nonautonomous fractional diffusion equations (NFDEs). Firstly, under the condition of weighted Hölder continuity, the existence and two estimates of classical solutions are obtained by virtue of the properties of the probability density function and the evolution operator family. Secondly, it focuses on the continuity and an estimate of classical solutions in the sense of fractional power norm. The results generalize some existing results on classical solutions and provide theoretical support for the application of NFDE.

1. Introduction

Due to the nonlocal kernel of fractional differential operators, the time fractional diffusion equations (FDEs) of order 0 to 1 can describe irregular diffusion phenomena with long tails. In real life, the regular diffusion phenomenon (integer order case) only occurs in a few special cases. Therefore, FDE has attracted the attention of many scholars.

Qualitative analysis on FDE is the premise of practical application. At present, the research in this area mainly includes the existence, regularity, and stability of solutions. El-Sayed and Herzallah [1] discussed the maximal regularity of strong solutions of the autonomous fractional nonhomogeneous evolution equations under the condition of Hölder continuity. The existence and uniqueness of the mild solution of the autonomous fractional evolution equations (AFEE) involving almost sectorial operators and the existence of classical solutions under the condition of Hölder continuity were researched by Wang et al. [2]. Other studies on the maximal regularity of classical solutions in the autonomous case in the function space of Hölder continuous functions can refer to [3–5]. It is known that Hölder continuity is a special case of weighted Hölder continuity. Mu et al. [6] studied the existence, maximum regularity, and spatial regularity of classical solutions to the autonomous fractional diffusion equations (AFDE) under the condition of weighted Hölder continuity and extended some results in existing research. Later, the Mittag-Leffler function and eigenfunction expansion were employed by Zhou et al. [7] to study the existence, uniqueness, and regularity of mild solutions to the backward problem of the AFDE in the function space of weighted Hölder continuous functions. For other relevant results, please refer to [8–15].

The FDE are often nonautonomous in practical problems, which makes it necessary to research the NFDE. The diffusion coefficient of the NFDE is related not only to the spatial variable but also to the time variable, which brings great difficulties to the research. For example, the diffusion term generates a continuous semigroup in the autonomous case, rather than a two-parameter family of evolution operators in the nonautonomous case. Nevertheless, El-Borai [16] obtained the existence of classical solutions of non-autonomous fractional evolution equations (NFEE) under the condition of Hölder continuity. A new resolvent family concept and a fixed point theorem were used by Debbouche and Baleanu [17] to establish some control results for nonlocal impulsive quasilinear

delay integrodifferential systems. Chalishajar et al. [18] used Sadovskii's fixed point theorem and Banach's fixed point theorem to study the existence of mild solutions to nonlocal problems of NFEE. In [19], the fractional resolvent family and the fixed point theorem are applied to investigate the global existence of mild solutions to NFEE. Chen et al. [20] applied noncompactness measure and Sadovskii's fixed point theorem to study the local existence and blow up of mild solutions to the Volterra-type NFEE. For other studies, see [21, 22]. On the basis of the above analysis, it can be found that the regularity of solutions to the NFDE and NFEE need to further study.

In this paper, we consider NFDE

$$\begin{cases} \partial_t^\alpha u(x,t) + \sum_{|p|\le 2m} b_p(x,t) D^p u(x,t) = f(x,t), & (x,t) \in \Omega \times I, \\ D^p u(x,t) = 0, \quad |p| \le m, & (x,t) \in \partial\Omega \times I, \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{1}$$

where ∂_t^α is the Caputo fractional partial derivative with respect to t, $\Omega \subset R^n$ is a bounded open domain, whose boundary $\partial\Omega$ is sufficiently smooth, $I = (0,T)\ T > 0$, u_0 is the initial data for u. $A(x,t) = \sum_{|p|\le 2m} b_p(x,t) D^p u(x,t)$, m is a positive integer, multi-index $x = (x_1, x_2, \ldots, x_n)$, $p = (p_1, p_2, \ldots, p_n)$, $|p| = \sum_{i=1}^n p_i$, $D^p = ((\partial^{p_1}/\partial x_1^{p_1})(\partial^{p_2}/\partial x_1^{p_2}) \cdots (\partial^{p_n}/\partial x_1^{p_n}))$. Additionally, the following hypotheses are satisfied:

(H_0)The operators $A(x,t)$ are uniformly elliptic operator in $\overline{\Omega} \times \overline{I}$. That is, there exists a constant C such that

$$C|\xi|^{2m} \le (-1)^m \mathrm{Re} \sum_{|p|=2m} b_p(x,t)\xi^p, \quad (x,t) \in \overline{\Omega} \times \overline{I}, \tag{2}$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_n) \in \mathbb{R}^n$, $|\xi|^2 = \sum_{i=1}^n \xi_i^2$;

(H_1)For $t \in \overline{I}$, the coefficients $b_q(x,t)$ are smooth functions with respect to$x \in \overline{\Omega}$, and there exists $0 < \sigma \le 1$ such that

$$\left| b_p(x,t) - b_p(x,s) \right| \le C|t-s|^\sigma, \quad \text{for } x \in \overline{\Omega}, \text{ and } t, s \in \overline{I}. \tag{3}$$

Firstly, when the inhomogeneous term satisfied the weighted Hölder continuity f and the initial value u_0 belonged to $D(A(0))$, the recursive method is applied to determine the representation of the solution to (1). The existence of a unique classical solution to (1) is proved by virtue of the properties of the probability density function and the evolution operator family. In addition, some estimates of the classical solution, directly connected with the regularity of u_0 and f, are carried out. Finally, the continuity of the classical solution to (1) in some fractional power norm is proved and a reasonable estimate is obtained. Theorem 1 extends Theorem 2.2 in [16], where f is Höler continuous.

The structure of this paper is as follows. The second section expounds the basic knowledge used later. In the third section, the existence and uniqueness of the classical solution to (1) and its continuity in the sense of fractional power are described, and the corresponding estimates are presented.

2. Preliminaries

Throughout this paper, the notation C represents a constant in a particular situation, $\Gamma(\cdot)$ denotes the Gamma function, and $B(\cdot,\cdot)$ denotes the Beta function. Let X be a Banach space with the norm $\|\cdot\|_X$ and γ, β are two constants satisfying $0 < \gamma < \beta \le 1$. We define the space of weighted Hölder continuous functions, which is Hölder continuous with the exponent γ and the weight $s^{1-\beta+\gamma}$, by $C^{\beta,\gamma}(I,X) = \{h\colon \overline{I} \longrightarrow X \mid h$ is continuous on I for $\beta \in (0,1)$(or $\overline{I}$ for $\beta = 1$), $\lim_{t\longrightarrow 0} t^{1-\beta} f(t)$ exsits for $\beta \in (0,1)$,

$$\sup_{0\le s<t\le T} \frac{s^{1-\beta+\gamma}\|h(t)-h(s)\|_X}{(t-s)^\gamma} < \infty, \qquad \lim_{t\longrightarrow 0} \sup_{0\le s<t} \frac{s^{1-\beta+\gamma}\|h(t)-h(s)\|_X}{(t-s)^\gamma} \longrightarrow \Big\}. \tag{4}$$

The norm is

$$\|h\|_{C^{\beta,\gamma}} = \sup_{0\le t\le T} t^{1-\beta}\|h(t)\|_X + \sup_{0\le s<t\le T} \frac{s^{1-\beta+\gamma}\|h(t)-h(s)\|_X}{(t-s)^\gamma}. \tag{5}$$

Remark 1. If h is Hölder continuous with exponent $\gamma(\gamma \in 0,1])$ on $\overline{I}$, then $h \in C^{1,\gamma}\ (I,X)$ [23]. Because of this, our results could generalize some existing conclusions which need Hölder continuity.

Set $u(t)(x) = u(x,t)$, $f(t)(x) = f(x,t)$. Define $A(t)\colon D \subset L^2(\Omega) \longrightarrow L^2(\Omega)$ by

$$A(t)u(t)(x) = A(x,t)u(x,t), \qquad D(A(t)) = D = H^{2m}(\Omega) \cap H_0^m(\Omega). \tag{6}$$

Then we turn (1) into the abstract fractional equations

$$\begin{cases} D_t^\alpha u(t) + A(t)u(t) = f(t), & t \in I, \\ u(0) = u_0, \end{cases} \tag{7}$$

in the Hilbert space $L^2(\Omega)$, where D_t^α is the Caputo fractional derivative. We say that $u\colon \overline{I} \longrightarrow L^2(\Omega)$ is a classical solution of (7), if $u(t)$ is continuous on $\overline{I}$, $A(t)u(t)$ and $D_t^\alpha u$ exist and are continuous on I, and (7) is satisfied on $\overline{I}$.

It is well known that each $-A(s)(s \in \overline{I})$ generates an analytic evolution family $\{T(t,s)\}_{t\ge 0}$. Under the assumptions (H_0) and (H_1), there exists a constant $k \ge 0$ such that $B(t) := A(t) + kI$ satisfies the following properties [24]:

(H_2) For all λ satisfying $\mathrm{Re}\lambda \le 0$, the resolvent $R(\lambda; B(t))$ of $B(t)$ exists and

$$\|R(\lambda; B(t))\|_{L^2(\Omega)} \le \frac{C}{|\lambda|+1}, \tag{8}$$

for each $t \in \overline{I}$.

(H_3)

$$\left\|(B(t)-B(s))B^{-1}(\tau)\right\|_{L^2(\Omega)} \le C|t-s|^\sigma, \tag{9}$$

for all $t, s, \tau \in \overline{I}$.

Without losing generality, we suppose $A(t)$ satisfies (H_2) and (H_3) in the following sections.

Set $U(t,s)=\alpha\int_0^\infty \theta\zeta_\alpha(\theta)T(t^\alpha\theta,s)\mathrm{d}\theta$, $\varphi_1(t,s)=(t-s)^{\alpha-1}[A(t)-A(s)]U(t-s,s)$, $\varphi_{n+1}(t,s)=\int_s^t\varphi_n(t,\tau)\varphi_1(\tau,s)\mathrm{d}\tau$, $n\in\mathbb{Z}^+$, $\varphi(t,s)=\sum_{n=1}^\infty\varphi_n(t,s)$, $V(t)=-A(t)A^{-1}(0)-\int_0^t\varphi(t,s)A(s)A^{-1}(0)\mathrm{d}s$, where $t,s\in\overline{I}$, ζ_α is probability density function defined on $(0,\infty)$:

$$\zeta_\alpha(\theta)=\frac{1}{\alpha}\theta^{-1-(1/\alpha)}\rho_\alpha\left(\theta^{(-1/\alpha)}\right),$$

$$\rho_\alpha(\theta)=\frac{1}{\pi}\sum_{n=0}^\infty(-1)^{n-1}\theta^{-\alpha n-1}\frac{\Gamma(n\alpha+1)}{n!}\sin(n\pi\alpha),\quad \theta\in(0,\infty), \tag{10}$$

and $\int_0^\infty\theta^\lambda\zeta_\alpha(\theta)\mathrm{d}\theta=(\Gamma(\lambda+1)/\Gamma(\alpha\lambda+1))$ for $\lambda\in(-1,\infty)$ [25]. In order to obtain the main results, we need to recall the fractional powers of $A(t)$ [24]. Let us denote by

$$A^{-q}(t)=\frac{1}{\Gamma(q)}\int_0^\infty s^{q-1}T(s,t)\mathrm{d}s,\quad \text{for } q>0,t\geq 0. \tag{11}$$

Then we define the fractional power of $A(t)$ by $A^q(t)=(A^{-q}(t))^{-1}$ for $q>0$, and $A^0(t)=I$. The following conclusions follow from some results of [16, 23, 24, 26].

Lemma 1

(i) $(t-s)^{\alpha-1}U(t-s,s)$ *and* $(t-s)^{\alpha-1}A(t)U(t-s,s)$ *are uniformly continuous, where* $t,s\in\overline{I}$, $t-s\geq\varepsilon$, *and* ϵ *is an arbitrary positive number;*

(ii) $\|A(t)(U(t-s,s)-U(t-s,t))\|_{L^2(\Omega)}\leq C(t-s)^{\sigma-\alpha}$ *for* $0\leq s<t\leq T$;

(iii) $\|A^q(s)U(t,s)\|_{L^2(\Omega)}\leq(C/t^{\alpha q})$ *for* $q\geq 0,t>0,s\in\overline{I}$;

(iv) $\|A^q(s)T(t,s)\|_{L^2(\Omega)}\leq(C/t^q)$ *for* $q\geq 0,t>0,s\in\overline{I}$;

(v) $\|\varphi_n(t,s)\|_{L^2(\Omega)}\leq(C^n(t-s)^{\sigma n-1}/\Gamma(\sigma n))$, $\|\varphi(t,s)\|_{L^2(\Omega)}\leq C(t-s)^{\sigma-1}$, $\|V(t)\|_{L^2(\Omega)}\leq C+Ct^\sigma$, $\varphi(t,s)$ *and* $V(t)$ *are uniformly continuous in* t,s, *and*

$$\begin{aligned}\|\varphi(t,\tau)-\varphi(s,\tau)\|_{L^2(\Omega)}&\leq C(t-s)^{\sigma-\nu}(s-\tau)^{\nu-1},\quad \text{for } 0\leq\tau<s<t\leq T,\\ \|V(t)-V(s)\|_{L^2(\Omega)}&\leq C(t-s)^\sigma+C(t-s)^{\sigma-\nu}s^\nu,\quad \text{for } 0\leq s<t\leq T,\end{aligned} \tag{12}$$

where $\nu\in(0,\sigma)$,

(vi) $\|A^{q_1}(t)A^{-q_2}(s)\|\leq C$, *where* $t,s\in\overline{I},0\leq q_1<q_2$;

(vii) $\|A(t)A^{-1}(s)\|\leq C$, *where* $t,s\in\overline{I}$;

(viii) $A^q(t)(t\geq 0)$ *is a closed operator, whose domain* $D(A^q(t))$ *is a Banach space;*

(ix) $A^{q_1+q_2}(t)=A^{q_1}(t)A^{q_2}(t)$, *for* $q_1,q_2\in\mathbb{R}$;

(x) $D(A^{q_1}(t))\subset D(A^{q_2}(t))$, *for* $0<q_2\leq q_1$.

3. Classical Solutions

In the following, we state the main results.

Theorem 1. *If* $f\in C^{\beta,\gamma}(I,L^2(\Omega))$, $u_0\in D(A(0))$, $0<\gamma<\beta\leq 1$, *and* $\alpha+\beta>1$, *then (7) has a unique classical solution:*

$$\begin{aligned}u(t)=u_0&+\int_0^t(t-s)^{\alpha-1}U(t-s,s)V(s)A(0)u_0\mathrm{d}s\\&+\int_0^t(t-s)^{\alpha-1}U(t-s,s)f(s)\mathrm{d}s\\&+\int_0^t\int_0^\tau(t-s)^{\alpha-1}U(t-s,s)\varphi(s,\tau)f(\tau)\mathrm{d}\tau\mathrm{d}s,\quad \text{for } t\in\overline{I},\end{aligned} \tag{13}$$

$$\begin{aligned}\|u(t)\|_{L^2(\Omega)}&\leq C\max\{1,t^{\alpha+\sigma}\}\|u_0\|_{L^2(\Omega)}\\&+C\max\{t^{\alpha+\beta-1},t^{\alpha+\sigma+\beta-1}\}\|f\|_{C^{\beta,\gamma}},\quad \text{for } t\in\overline{I},\end{aligned} \tag{14}$$

$$\begin{aligned}&\|A(t)u(t)\|_{L^2(\Omega)}+\|D_t^\alpha u(t)\|_{L^2(\Omega)}\\&\leq C\|A(0)u_0\|_{L^2(\Omega)}+C\max\{t^{\beta-1},t^{2\sigma+\beta-1}\}\|f\|_{C^{\beta,\gamma}},\quad \text{for } t\in I.\end{aligned} \tag{15}$$

Proof. We set

$$\begin{aligned}u(t)=u_0&+\int_0^t(t-s)^{\alpha-1}U(t-s,s)V(s)A(0)u_0\mathrm{d}s\\&+\int_0^t(t-s)^{\alpha-1}U(t-s,s)w(s)\mathrm{d}s.\end{aligned} \tag{16}$$

Substituting it into (7), by Theorem 2.2 in [16], we get

$$\begin{aligned}&D^\alpha\left(u_0+\int_0^t(t-s)^{\alpha-1}U(t-s,s)V(s)A(0)u_0\mathrm{d}s\right)\\&+A(t)\left(u_0+\int_0^t(t-s)^{\alpha-1}U(t-s,s)V(s)A(0)u_0\mathrm{d}s\right)=0.\end{aligned} \tag{17}$$

Then formally using Lemma 1 of [16], we obtain that

$$\begin{aligned} f(t) &= D^{\alpha}\int_0^t (t-s)^{\alpha-1}U(t-s,s)w(s)\mathrm{d}s \\ &\quad + A(t)\int_0^t (t-s)^{\alpha-1}U(t-s,s)w(s)\mathrm{d}s \\ &= w(t) - \int_0^t (t-s)^{\alpha-1}A(s)U(t-s,s)w(s)\mathrm{d}s \\ &\quad + A(t)\int_0^t (t-s)^{\alpha-1}U(t-s,s)w(s)\mathrm{d}s \\ &= w(t) + \int_0^t \varphi_1(t-s,s)w(s)\mathrm{d}s. \end{aligned} \tag{18}$$

Write $w_0(t) = f(t)$, $w_{n+1} = -\int_0^t \varphi_1(t,s)w_n(s)\mathrm{d}s$, and $w(t) = \sum_{n=0}^{\infty} w_n(t)$. According to the character of φ_n and Fubini's theorem, we have

$$w_n(t) = \int_0^t \varphi_n(t,s)f(s)\mathrm{d}s. \tag{19}$$

In view of Lemma 1, we obtain that

$$\begin{aligned} \left\|\sum_{n=0}^{\infty} w_n(t)\right\|_{L^2(\Omega)} &\le \sum_{n=0}^{\infty}\int_0^t \frac{C^n(t-s)^{\sigma n-1}}{\Gamma(\sigma n)}s^{\beta-1}\mathrm{d}s\|f\|_{C^{\beta,\gamma}} \\ &\le \|f\|_{C^{\beta,\gamma}}\sum_{n=0}^{\infty}\frac{C^n B(\sigma n,\beta)}{\Gamma(\sigma n)}T^{n\sigma+\beta-1} \\ &\le \|f\|_{C^{\beta,\gamma}}T^{\beta-1}\Gamma(\beta)\sum_{n=0}^{\infty}\frac{(CT^{\sigma})^n}{\Gamma(\sigma n+\beta)} \\ &= \|f\|_{C^{\beta,\gamma}}T^{\beta-1}\Gamma(\beta)E_{\sigma,\beta}(CT^{\sigma}), \end{aligned} \tag{20}$$

which implies that $\sum_{n=0}^{\infty} w_n(t)$ uniformly converges on $\overline{I}$ and

$$w(t) = f(t) + \int_0^t \varphi(t,s)f(s)\mathrm{d}s. \tag{21}$$

Besides,

$$\begin{aligned} \|w(t)-w(s)\|_{L^2(\Omega)} &\le \|f(t)-f(s)\|_{L^2(\Omega)} \\ &\quad + \int_0^s \|\varphi(t,\tau)-\varphi(s,\tau)\|_{L^2(\Omega)}\|f(\tau)\|_{L^2(\Omega)}\mathrm{d}\tau \\ &\quad + \int_s^t \|\varphi(t,\tau)f(\tau)\|_{L^2(\Omega)}\mathrm{d}\tau \\ &\le s^{-1+\beta-\gamma}(t-s)^{\gamma}\|f\|_{C^{\beta,\gamma}} + C\|f\|_{C^{\beta,\gamma}}\int_0^s (t-s)^{\sigma-\nu}(s-\tau)^{\nu-1}\tau^{\beta-1}\mathrm{d}\tau \\ &\quad + C\|f\|_{C^{\beta,\gamma}}\int_s^t (t-\tau)^{\sigma-1}\tau^{\beta-1}\mathrm{d}\tau \\ &\le s^{-1+\beta-\gamma}(t-s)^{\gamma}\|f\|_{C^{\beta,\gamma}} + CB(\nu,\beta)(t-s)^{\sigma-\nu}s^{\nu+\beta-1}\|f\|_{C^{\beta,\gamma}} \\ &\quad + \frac{C}{\sigma}s^{\beta-1}(t-s)^{\sigma}\|f\|_{C^{\beta,\gamma}}, \end{aligned} \tag{22}$$

provided $0 < s < t \le T$. Using Lemma 1, we deduce that u is continuous on I. In addition, we conclude from Lemma 1 that

$$\begin{aligned} \|u(t)-u(0)\|_{L^2(\Omega)} &\le C\int_0^t (t-s)^{\alpha-1}(1+s^{\sigma})\mathrm{d}s\|u_0\|_{L^2(\Omega)} + \left(C\int_0^t (t-s)^{\alpha-1}s^{\beta-1}\mathrm{d}s\right. \\ &\quad \left. + \int_0^t\int_0^s (t-s)^{\alpha-1}(s-\tau)^{\sigma-1}\tau^{\beta-1}\mathrm{d}\tau\mathrm{d}s\right)\|f\|_{C^{\beta,\gamma}} \\ &\le C\left(\frac{t^{\alpha}}{\alpha} + CB(\alpha,\sigma+1)t^{\alpha+\sigma}\right)\|u_0\|_{L^2(\Omega)} \\ &\quad + C\left(B(\alpha,\beta)t^{\alpha+\beta-1} + B(\sigma,\beta)B(\alpha,\sigma+\beta)t^{\alpha+\sigma+\beta-1}\right)\|f\|_{C^{\beta,\gamma}}, \end{aligned} \tag{23}$$

implies that u is continuous at $t = 0$. Next, we show that $u(t) \in D$ for $t \in I$. In view of

$$\begin{aligned}
&\int_0^t (t-s)^{\alpha-1} A(t) U(t-s,s) V(s) A(0) u_0 \mathrm{d}s \\
&= \int_0^t (t-s)^{\alpha-1} A(t) U(t-s,t)(V(s)-V(t)) A(0) u_0 \mathrm{d}s \\
&\quad + \int_0^t (t-s)^{\alpha-1} A(t)(U(t-s,s)-U(t-s,t)) V(s) A(0) u_0 \mathrm{d}s \\
&\quad - \int_0^\infty \zeta_\alpha(\theta) T(t^\alpha \theta, t) V(t) A(0) u_0 d\theta + V(t) A(0) u_0,
\end{aligned} \tag{24}$$

we conclude from Lemma 1 and (22) that

$$\begin{aligned}
&\left\| A(t)\left(u_0 + \int_0^t (t-s)^{\alpha-1} U(t-s,s) V(s) A(0) u_0 \mathrm{d}s \right) \right\|_{L^2(\Omega)} \\
&\le \left\| A(t) A^{-1}(0) A(0) u_0 \right\|_{L^2(\Omega)} + C \int_0^t (t-s)^{-1} \left((t-s)^\sigma + (t-s)^{\sigma-\nu} s^\nu \right) \mathrm{d}s \left\| A(0) u_0 \right\|_{L^2(\Omega)} \\
&\quad + C \int_0^t (t-s)^{\sigma-1} (1+s^\sigma) \mathrm{d}s \left\| A(0) u_0 \right\|_{L^2(\Omega)} + C(1+t^\sigma) \left\| A(0) u_0 \right\|_{L^2(\Omega)} \\
&\le C \left\| A(0) u_0 \right\|_{L^2(\Omega)} + C t^\sigma \left\| A(0) u_0 \right\|_{L^2(\Omega)} + C t^{2\sigma} \left\| A(0) u_0 \right\|_{L^2(\Omega)}.
\end{aligned} \tag{25}$$

That is, $u_0 + \int_0^t (t-s)^{\alpha-1} U(t-s,s) V(s) A(0) u_0 \mathrm{d}s \in D$ for $t \in I$. We also know that $A(t)(u_0 + \int_0^t (t-s)^{\alpha-1} U(t-s,s) V(s) A(0) u_0 ds)$ is continuous for $t \in I$ ([16]). Next, we write

$$\begin{aligned}
&\int_0^t (t-s)^{\alpha-1} A(t) U(t-s,s) w(s) \mathrm{d}s \\
&= \int_0^t (t-s)^{\alpha-1} A(t) U(t-s,t)(w(s)-w(t)) \mathrm{d}s \\
&\quad + \int_0^t (t-s)^{\alpha-1} A(t)(U(t-s,s)-U(t-s,t)) w(s) \mathrm{d}s \\
&\quad - \int_0^\infty \zeta_\alpha(\theta) T(t^\alpha \theta, t) w(t) d\theta + w(t) \\
&:= I_1(t) + I_2(t) + I_3(t) + I_4(t).
\end{aligned} \tag{26}$$

Thanks to (22) and

$$\begin{aligned}
\|w(t)\|_{L^2(\Omega)} &\le t^{\beta-1} \|f\|_{C^{\beta,\gamma}} + \int_0^t (t-s)^{\sigma-1} s^{\beta-1} \mathrm{d}s \|f\|_{C^{\beta,\gamma}} \\
&\le \left(t^{\beta-1} + B(\beta,\sigma) t^{\sigma+\beta-1} \right) \|f\|_{C^{\beta,\gamma}},
\end{aligned} \tag{27}$$

drawn from Lemma 1, we come to the conclusion that

$$\begin{aligned}
&\left\|\int_0^t (t-s)^{\alpha-1}A(t)U(t-s,s)w(s)\mathrm{d}s\right\|_{L^2(\Omega)}\\
&\le C\int_0^t (t-s)^{-1}\left(s^{-1+\beta-\gamma}(t-s)^{\gamma}+(t-s)^{\sigma-\nu}s^{\nu+\beta-1}+s^{\beta-1}(t-s)^{\sigma}\right)\mathrm{d}s\|f\|_{C^{\beta,\gamma}}\\
&\quad+C\int_0^t (t-s)^{\sigma-1}\left(s^{\beta-1}+s^{\sigma+\beta-1}\right)\mathrm{d}s\|f\|_{C^{\beta,\gamma}}+C\left(t^{\beta-1}+t^{\sigma+\beta-1}\right)\|f\|_{C^{\beta,\gamma}}\\
&\le C\Big(B(\gamma,\beta-\gamma)t^{\beta-1}+B(\sigma-\nu,\nu+\beta)t^{\sigma+\beta-1}+B(\sigma,\beta)t^{\sigma+\beta-1}\Big)\\
&\quad+t^{\beta-1}+t^{\sigma+\beta-1}+B(\sigma,\sigma+\beta)t^{2\sigma+\beta-1}\Big)\|f\|_{C^{\beta,\gamma}}.
\end{aligned}\tag{28}$$

That is $\int_0^t (t-s)^{\alpha-1}U(t-s,s)w(s)\mathrm{d}s \in D$ for $t\in[\varepsilon,T]$ $(\forall\varepsilon\in I)$. Next, we show that $A\int_0^t (t-s)^{\alpha-1}U(t-s,s)w(s)\mathrm{d}s \in C(I,L^2(\Omega))$. Moreover,

$$\begin{aligned}
I_1(t)-I_1(s)&=\int_s^t (t-\tau)^{\alpha-1}A(t)U(t-\tau,t)(w(\tau)-w(t))\mathrm{d}\tau\\
&\quad+\int_0^s\Big((t-\tau)^{\alpha-1}A(t)U(t-\tau,t)-(s-\tau)^{\alpha-1}A(s)U(s-\tau,s)\Big)\\
&\quad\cdot(w(\tau)-w(s))\mathrm{d}\tau+\int_0^s (t-\tau)^{\alpha-1}A(t)U(t-\tau,t)(w(s)-w(t))\mathrm{d}\tau\\
&:=J_1(t)+J_2(t)+J_3(t),
\end{aligned}\tag{29}$$

and for $t>s$, (22) and Lemma 1 imply that

$$\begin{aligned}
\|J_1(t)\|_{L^2(\Omega)}&\le C\int_s^t\left(\tau^{-1+\beta-\gamma}(t-\tau)^{\gamma-1}+(t-\tau)^{\sigma-\nu-1}\tau^{\nu+\beta-1}+\tau^{\beta-1}(t-\tau)^{\sigma-1}\right)\mathrm{d}\tau\\
&\le C\left(\frac{s^{-1+\beta-\gamma}}{\gamma}(t-s)^{\gamma}+\frac{\max\left\{s^{\nu+\beta-1},t^{\nu+\beta-1}\right\}}{\sigma-\nu}(t-s)^{\sigma-\nu}+\frac{s^{\beta-1}}{\sigma}(t-s)^{\sigma}\right).
\end{aligned}\tag{30}$$

Then $\lim_{t\longrightarrow s^+}J_1(t)=0$ $(s\in[\varepsilon,T],\forall\varepsilon\in I)$ holds. Furthermore, using Lemma 1 we have

$$\begin{aligned}
\left\|(t-\tau)^{\alpha-1}A(t)U(t-\tau,t)(w(\tau)-w(s))\right\|_{L^2(\Omega)}&\le C(t-\tau)^{-1}\|w(\tau)-w(s)\|_{L^2(\Omega)}\\
&\le C\left(\tau^{-1+\beta-\gamma}(s-\tau)^{\gamma-1}+(s-\tau)^{\sigma-\nu-1}\tau^{\nu+\beta-1}+\tau^{\beta-1}(s-\tau)^{\sigma-1}\right)\in L^1(I,\mathbb{R}).
\end{aligned}\tag{31}$$

Thus $\lim_{t\longrightarrow s^+}J_2(t)=0$ $(s\in[\varepsilon,T],\forall\varepsilon\in I)$ could be immediately gotten by Lebesgue's dominated convergence theorem. Applying Lemma 1, we also find that

$$\begin{aligned}\|J_3(t)\|_{L^2(\Omega)} &\le \left\|\left(\int_0^\infty \zeta_\alpha(\theta)T\left((t-s)^\alpha\theta,t\right)\mathrm{d}\theta - \int_0^\infty \zeta_\alpha(\theta)T\left(t^\alpha\theta,t\right)\mathrm{d}\theta\right)(w(t)-w(s))\right\|_{L^2(\Omega)} \\ &\le C\|w(t)-w(s)\|_{L^2(\Omega)}.\end{aligned} \tag{32}$$

Then $\lim_{t\longrightarrow s^+} J_3(t) = 0$ $(s \in [\varepsilon, T])$ follows from (22). Next, we write

$$\begin{aligned}I_2(t) - I_2(s) &= \int_s^t (t-\tau)^{\alpha-1}A(t)(U(t-\tau,\tau)-U(t-\tau,t))w(\tau)\mathrm{d}\tau \\ &\quad + \int_0^s \left((t-\tau)^{\alpha-1}A(t)(U(t-\tau,\tau)-U(t-\tau,t)) - (s-\tau)^{\alpha-1}A(s)(U(s-\tau,\tau)-U(s-\tau,s))\right)w(\tau)\mathrm{d}\tau =: K_1 + K_2.\end{aligned} \tag{33}$$

Then

$$\begin{aligned}\|K_1(t)\|_{L^2(\Omega)} &\le C\int_s^t (t-\tau)^{\sigma-1}\left(\tau^{\beta-1}+\tau^{\sigma+\beta-1}\right)\mathrm{d}\tau \\ &\le \frac{C}{\sigma}\left(s^{\beta-1}+\max\left\{s^{\sigma+\beta-1},t^{\sigma+\beta-1}\right\}\right)(t-s)^\sigma,\end{aligned} \tag{34}$$

derived by Lemma 1 and (27). That is, $\lim_{t\longrightarrow s^+} K_1(t) = 0$ $(s \in [\varepsilon, T], \forall \varepsilon \in I)$. Owing to

$$\begin{aligned}&\left\|(s-\tau)^{\alpha-1}A(s)(U(s-\tau,\tau)-U(s-\tau,t))w(\tau)\right\|_{L^2(\Omega)} \\ &\quad \le C(s-\tau)^{\sigma-1}\left(\tau^{\beta-1}+\tau^{\sigma+\beta-1}\right) \in L^1(I,\mathbb{R}),\end{aligned} \tag{35}$$

$\lim_{t\longrightarrow s^+} K_2(t) = 0$ $(s \in [\varepsilon, T]$, ε is arbitrary and $\varepsilon \in I)$ could be obtained by Lebesgue's dominated convergence theorem. In addition, Lemma 1 and the formula (22) imply that $\lim_{t\longrightarrow s^+} I_3(t) = \lim_{t\longrightarrow s^+} I_4(t) = 0$ $(s \in [\varepsilon, T])$. When $t < s$, the above limits are similar. Then by (16) and the properties of $w(t)$, using arguments similar to the ones in Theorem 2.2 and Lemma 1 in [16] one can easily obtain that $D_t^\alpha u$ exists and is continuous on I, and u satisfies (7). Therefore, one obtains u is a classical solution of (7). It is also easily seen that (14) and (15) hold, by (23), (25), and (28). □

Remark 2. Theorem 1 extends Theorem 2.2 in [16], where f is Höler continuous.

Theorem 2. *If $f \in C^{\beta,\gamma}(I, L^2(\Omega))$, $u_0 \in D(A(0))$, $0 < \gamma < \beta \le 1$, $\alpha + \beta > 1$, and $0 < \beta_1 < 1 - (1 - \beta/\alpha)$, then the classical solution to (7) has the property: $A^{\beta_1}(t)u(t)$ is continuous on $\overline{I}$, and*

$$\left\|A_1^\beta(t)u(t)\right\|_{L^2(\Omega)} \le C\left\|D\left(A^\beta(0)\right)u_0\right\|_{L^2(\Omega)} + C\|f\|_{C^{\beta,\gamma}}. \tag{36}$$

Proof. The existence of the classical solution could be gotten immediately from Theorem 1. Since $D(A(0)) \subset D(A^\beta(0))$, $u_0 \in D(A^\beta(0))$. In view of $A^{\beta_1}(t) = A^{\beta_1-1}(t)A(t)$ and $A^{\beta_1-1}(t)$ is a bounded linear operator for $t \in \overline{I}$, using Theorem 1 we may find $A^{\beta_1}(t)u(t)$ is continuous on I. If $\alpha + \beta > 1$, then $0 < \beta_1 < 1 - (1 - \beta/\alpha) < \beta$. Next we show that $A^{\beta_1}(t)u(t)$ is continuous at $t = 0$. In fact, note that

$$\begin{aligned}A^{\beta_1}(t)u(t) - A^{\beta_1}(0)u_0 &= \left(A^{\beta_1}(t)A^{-\beta}(0) - A^{\beta_1-\beta}(0)\right)A^\beta(0)u_0 \\ &\quad + \int_0^t (t-s)^{\alpha-1}A^\mu(s)U(t-s,s)V(s)A^{-\mu}(t)A^{\beta_1}(t)A^{1-\beta}(0)A^\beta(0)u_0\mathrm{d}s \\ &\quad + \int_0^t (t-s)^{\alpha-1}A^{\beta_1-1}(t)A(t)U(t-s,s)w(s)\mathrm{d}s \\ &=: P_1(t) + P_2(t) + P_3(t),\end{aligned} \tag{37}$$

where $1 - \beta + \beta_1 < \mu < 1$.

It is clear that (H_1) implies that $\lim_{t\longrightarrow 0} P_1(t) = 0$. We now estimate, using Lemma 1 and (27),

$$\begin{aligned}
\left\|P_2(t)\right\|_{L^2(\Omega)} &\le C\int_0^t (t-s)^{\alpha-\alpha\mu-1}\left(1+s^{\sigma}\right)\mathrm{d}s\left\|A^{\beta}(0)u_0\right\|_{L^2(\Omega)} \\
&= C\left(t^{\alpha(1-\mu)}+B(\alpha(1-\mu),\sigma+1)t^{\alpha(1-\mu)+\sigma}\right)\left\|A^{\beta}(0)u_0\right\|_{L^2(\Omega)}, \\
\left\|P_3(t)\right\|_{L^2(\Omega)} &\le \left\|\int_0^t (t-s)^{\alpha-1}A^{\beta_2}(s)A^{-\beta_2}(s)A^{\beta_1}(t)U(t-s,s)w(s)\mathrm{d}s\right\| \\
&\le C\int_0^t (t-s)^{\alpha-\alpha\beta_2-1}\left(s^{\beta-1}+B(\beta,\sigma)s^{\sigma+\beta-1}\right)\|f\|_{C^{\beta,\gamma}}\mathrm{d}s \\
&\le C\left(B(\alpha-\alpha\beta_2,\beta)t^{\alpha-\alpha\beta_2+\beta-1}+B(\beta,\sigma)B(\alpha-\alpha\beta_2,\beta+\sigma)t^{\alpha-\alpha\beta_2+\beta+\sigma-1}\right)\|f\|_{C^{\beta,\gamma}},
\end{aligned} \tag{38}$$

where $\beta_1<\beta_2<1-(1-\beta/\alpha)$. These prove $\lim_{t\longrightarrow 0}P_2(t)=\lim_{t\longrightarrow 0}P_3(t)=0$. From the above, we see that $A^{\beta_1}(t)u(t)$ is continuous on $\overline{I}$ and (36) holds. □

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This research was supported by Scientific Research Foundation of the Higher Education Institutions of Gansu Province (2019B-022), Fundamental Research Funds for the Central Universities (31920180047), Gansu Provincial First-Class Discipline Program of Northwest Minzu University (11080305), and Innovation Team of Intelligent Computing and Dynamical System Analysis and Application of Northwest Minzu University.

References

[1] A. M. A. El-Sayed and M. A. E. Herzallah, "Continuation and maximal regularity of fractional-order evolution equation," *Journal of Mathematical Analysis and Applications*, vol. 296, no. 1, pp. 340–350, 2004.

[2] R.-N. Wang, D.-H. Chen, and T.-J. Xiao, "Abstract fractional Cauchy problems with almost sectorial operators," *Journal of Differential Equations*, vol. 252, no. 1, pp. 202–235, 2012.

[3] C. Li and M. Li, "Holder regularity for abstract fractional cauchy problems with order in (0, 1)," *Journal of Applied Mathematics and Physics*, vol. 6, no. 1, pp. 310–319, 2018.

[4] L. Liu, Z. Fan, G. Li, and S. Piskarev, "Maximal regularity for fractional cauchy equation in hölder space and its approximation," *Computational Methods in Applied Mathematics*, vol. 19, no. 4, pp. 779–796, 2019.

[5] R. Ponce, "Hölder continuous solutions for fractional differential equations and maximal regularity," *Journal of Differential Equations*, vol. 255, no. 10, pp. 3284–3304, 2013.

[6] J. Mu, B. Ahmad, and S. Huang, "Existence and regularity of solutions to time-fractional diffusion equations," *Computers & Mathematics with Applications*, vol. 73, no. 6, pp. 985–996, 2017.

[7] Y. Zhou, J. Wei He, B. Ahmad, and N. Huy Tuan, "Existence and regularity results of a backward problem for fractional diffusion equations," *Mathematical Methods in the Applied Sciences*, vol. 42, no. 18, pp. 6775–6790, 2019.

[8] X. Su and M. Li, "The regularity of fractional stochastic evolution equations in Hilbert space," *Stochastic Analysis and Applications*, vol. 36, no. 4, pp. 639–653, 2018.

[9] G. A. Zou, B. Wang, and Y. Zhou, "Existence and regularity of mild solutions to fractional stochastic evolution equations," *Mathematical Modelling of Natural Phenomena*, vol. 13, no. 1, 2018.

[10] Y. Sarol, F. Viens, and F. Viens, "Time regularity of the evolution solution to fractional stochastic heat equation," *Discrete & Continuous Dynamical Systems-B*, vol. 6, no. 4, pp. 895–910, 2006.

[11] Z. Fan, "Existence and regularity of solutions for evolution equations with Riemann-Liouville fractional derivatives," *Indagationes Mathematicae*, vol. 25, no. 3, pp. 516–524, 2014.

[12] J. L. Vázquez, A. de Pablo, F. Quirós, and A. Rodríguez, "Classical solutions and higher regularity for nonlinear fractional diffusion equations," *Journal of the European Mathematical Society*, vol. 19, no. 7, pp. 1949–1975, 2017.

[13] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, 2020.

[14] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[15] S. Liu, Z. Ji, and H. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, Article ID 7685460, 18 pages, 2020.

[16] M. El-Borai, "The fundamental solutions for fractional evolution equations of parabolic type," *International Journal of Stochastic Analysis*, vol. 11, no. 1, pp. 29–43, 2004.

[17] A. Debbouche and D. Baleanu, "Controllability of fractional evolution nonlocal impulsive quasilinear delay integro-differential systems," *Computers & Mathematics with Applications*, vol. 62, no. 3, pp. 1442–1450, 2011.

[18] D. Chalishajar, D. Raja, K. Karthikeyan, and P. Sundararajan, "Existence results for nonautonomous impulsive fractional evolution equations," *Results in Nonlinear Analysis*, vol. 1, no. 3, pp. 133–147, 2018.

[19] B. Zhu, L. Liu, and Y. Wu, "Local and global existence of mild solutions for a class of nonlinear fractional reaction-diffusion equations with delay," *Applied Mathematics Letters*, vol. 61, pp. 73–79, 2016.

[20] P. Chen, X. Zhang, X. Zhang, and Y. Li, "A blowup alternative result for fractional nonautonomous evolution equation of Volterra type," *Communications on Pure & Applied Analysis*, vol. 17, no. 5, pp. 1975–1992, 2018.

[21] M. Mahmoud, E. Khairia, and G. Eman, "On some fractional evolution equations," *Computers & Mathematics with Applications*, vol. 59, no. 3, pp. 1352–1355, 2010.
[22] P. Chen, X. Zhang, and Y. Li, "Cauchy problem for fractional non-autonomous evolution equations," *Banach Journal of Mathematical Analysis*, vol. 14, no. 2, pp. 559–584, 2020.
[23] A. Yagi, *Abstract Parabolic Evolution Equations and Their Applications*, Springer Science & Business Media, Berlin, Germany, 2009.
[24] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer Science & Business Media, Berlin, Germany, 1992.
[25] M. M. El-Borai, "Some probability densities and fundamental solutions of fractional evolution equations," *Chaos, Solitons & Fractals*, vol. 14, no. 3, pp. 433–440, 2002.
[26] A. Friedman, *Partial Differential Equations*, Holt, Rinehat and Winston, New York, NY, USA, 1969.

Hindawi
Complexity
Volume 2020, Article ID 8825301, 11 pages
https://doi.org/10.1155/2020/8825301

Research Article

Limited-Budget Formation Control for High-Order Linear Swarm Systems with Fix Topologies

Hongtao Dang,[1] Le Wang,[2] Yan Zhang,[1] and Jianye Yang[1]

[1]*Shannxi Engineering Research Center of Controllable Neutron Source, School of Science, Xijing University, Xi'an 710123, China*
[2]*High-Tech Institute of Xi'an, Xi'an 710025, China*

Correspondence should be addressed to Le Wang; wangaz14@163.com

Received 31 August 2020; Revised 13 September 2020; Accepted 25 September 2020; Published 8 October 2020

Academic Editor: Ning Cai

This paper discusses limited-budget time-varying formation design and analysis problems for a high-order linear swarm system with a fixed communication topology. Firstly, the communication topology among agents is modeled as an undirected and connected graph, and a new formation control protocol with an energy integral term is proposed to realize formation control and to guarantee the practical energy assumption is less than the limited energy budget. Then, by the matrix inequality tool, sufficient conditions for limited-budget formation design and analysis are proposed, respectively, which are scalable and checkable since they are independent of the number of agents of a swarm system and can be transformed into linear matrix inequality constraints. Moreover, an explicit expression of the formation center function is given, which contains the formation function part and the cooperative state part and is not associated with the derivatives of the formation functions. Finally, a numerical simulation is shown to demonstrate the effectiveness of theoretical results.

1. Introduction

Recently, many researchers from system and control fields paid their attentions to the distributed coordination of swarm systems, which have potential applications in many aspects, such as flocking algorithm and theory [1, 2], synchronization analysis and design [3–11], formation control [12–14], and distributed parallel computation [15–19]. Formation control of a distributed swarm system is inspired by the biological formation behaviors without the superintend node, where it is needed that some specific geometric structure is achieved and maintained by local interactions among all animals. Formation control of distributed swarm systems has extensive application in the coordinated maneuver of multiple androids and coordinated attack of multiple unmanned surface ships. In [20], formation control methods were divided into three classes: the virtual structure method, the behavior-based method, and the synchronization-based method, and it was pointed out that the synchronization-based method is distributed and can overcome some limitations of the virtual structure method and the behavior-based method.

According to the time-varying property of geometric structures, formation is divided into time-invariant geometric structures and time-varying geometric structures. For time-invariant geometric structures, the time derivative of the formation function of a swarm system as a whole is zero; that is, the formation structure is time-invariant once a geometric structure is formed by all agents. For time-varying geometric structures, the time derivative of the formation function of a swarm system as a whole is nonzero, which means that the formation structure is time-varying even if a geometric structure is formed by all agents. By the Nyquist stability criterion, time-invariant formation criteria were proposed in [21]. Du et al. [22] presented time-invariant formation control criteria by using nonlinear formation protocol and applied theoretical results into multiple androids systems. Qin et al. [23] investigated the influences of time delays on the time-invariant formation of swarm systems and gave several sufficient conditions for time-invariant formation control. In [24–26], time-varying formation analysis and design for swarm systems were addressed and some important conclusions for time-varying

formation were presented. Overall, time-varying formation control is more complex than time-invariant formation control, and some specific formation feasible conditions are needed.

In [27–29], formation control design and analysis problems for swarm systems with different topology structures were investigated, where the influences of the limited energy supply were not dealt with, which is critically important in practical applications. In [30, 31], the concept of the guaranteed-cost control was introduced to analyze the impacts of the energy constraint and gave several guaranteed-cost formation criteria. Yu et al. [32] investigated the guaranteed-cost time-varying formation control problem, where the impacts of unknown external disturbances and time delays were considered and the guaranteed-cost time-varying formation criteria were proposed, but the performance cost cannot be determined by their methods. In [30–32], the energy supply of a swarm system as a whole cannot be given previously. However, the whole energy supply is usually limited in practical swarm systems and it is critically important to discuss the influences of the whole energy supply on time-varying formation design and analysis problems. To the best of our knowledge, limited-budget formation control for high-order linear swarm systems with fixed communication topologies is still open and is not comprehensively addressed and analyzed.

This paper deals with time-varying formation design and analysis for high-order linear swarm systems with the limited energy budget. A formation control protocol is proposed on the basis of the state errors and the formation function errors among neighboring agents, where an energy integral term is introduced to ensure that the whole energy budget is larger than or equal to the practical energy assumption. Furthermore, the dynamics of a swarm system with a fixed communication topology is divided into two parts by the state space decomposition, which can be used to describe the macroscopic motion as a whole and the microcosmic motions between any two agents. By constructing the relationship between the matrix variable and the energy budget, limited-budget formation design and analysis criteria are proposed, respectively, where the scalability property and the checkable property are discussed in detail. Moreover, the formation center function is determined, which involves two independent parts associated with the average value of the time-varying formation functions and the average value of the initial conditions of cooperative states, respectively.

Compared with the recent works on the time-varying formation control, the innovation of this paper is twofold. Firstly, this paper considers the limited-budget time-varying formation control, where the energy budget of the whole swarm system is given and is limited previously. The energy budget was not considered and the practical energy consumption cannot be limited by the energy budget in the time-varying formation as shown in [24–29]. Secondly, the effect of the energy budget to the matrix variable of the matrix inequality conditions is considered and determined, which can introduce the energy constraint into the design of the gain matrix of the time-varying formation control protocol. In contrast, the formation control methods in [24–29] cannot give the specific interaction mechanism for the energy budget.

The main arrangement of this paper is given as follows. The communication topology of a swarm system is modeled and the problem description of limited-budget formation for a swarm system is shown in Section 2. The main results of limited-budget formation are presented in Section 3. A numerical simulation is given to demonstrate the limited-budget formation criterion in Section 4. Finally, Section 5 summarizes the main conclusions of this paper.

Symbols R^n and $R^{n\times n}$, respectively, represent the n-dimensional real column vector and the n-dimensional real matrix space. 1_N denotes the N-dimensional column vector with all components 1. 0 represents the zero vector or zero matrix with compatible dimensions. The symbol $\otimes$ represents the Kronecker product. The notation $Q^T = Q > 0$ shows that the matrix Q is symmetric and positive definite.

2. Problem Description

2.1. Communication Topology Model. For a swarm system with N identical agents, the fixed communication topology is depicted a weighed graph $G = (V(G), E(G))$, where $V(G) = \{v_1, v_2, \cdots, v_N\}$ is the node set with $v_k (k = 1, 2, \cdots, N)$ representing agent k and $E(G) = \{e_{ki} = (v_k, v_i)\}$ is the edge set with (v_k, v_i) denoting the interaction channel from agent k to agent i. The notation $N_k = \{i\colon (v_i, v_k) \in E(G)\}$ is used to denote the neighbor set of agent k. The symbol w_{ik} denotes the communication weight from agent k to agent i, where $w_{ii} = 0$ and $w_{ik} = 0$ when agent k and agent i are not connected and $w_{ik} > 0$ when agent k and agent i are connected. The Laplacian matrix of the communication topology is denoted as $L = [l_{ik}] \in R^{N\times N}$ with $l_{kk} = \sum_{i\in N_k} w_{ki}$ and $l_{ik} = -w_{ik} (k \neq i)$. If the fixed communication topology is connected, then the Laplacian matrix L is symmetric and positive semidefinite; that is, zero is its simple eigenvalue and all its nonzero eigenvalues are positive. More basic concepts and conclusions on graph theory can be found in [33].

2.2. Swarm System Model. The dynamics of each agent is depicted in the following high-order linear system:

$$\dot{x}_k(t) = Ax_k(t) + Bu_k(t), \tag{1}$$

where $k = 1, 2, \cdots, N$, $A \in R^{n\times n}$, and $B \in R^{n\times m}$, and $x_k(t)$ and $u_k(t)$, respectively, denote the cooperative state and the control input. Moreover, a vector-valued function $\eta(t) = [\eta_1^T(t), \eta_2^T(t), \cdots, \eta_N^T(t)]^T$ is used to depict a specific formation structure for swarm system (1) to maintain, where the piecewise continuous differentiable components $\eta_k(t)$ are formation functions of agent $k (k = 1, 2, \cdots, N)$. The formation is time-varying when $\eta(t)$ is time-varying and the formation is fixed when $\eta(t)$ is time-invariant.

In the following, a limited-budget formation control protocol with a fixed communication topology is proposed as follows:

$$\begin{cases} u_k(t) = K\sum_{i\in N_k} w_{ki}\left(x_i(t) - \eta_i(t) - x_k(t) + \eta_k(t)\right), \\ J_e = \sum_{k=1}^{N}\int_0^{+\infty} u_k^T(t)Qu_k(t)\mathrm{d}t, \end{cases} \tag{2}$$

where $k = 1, 2, \cdots, N$, $K \in R^{m\times n}$, and $Q^T = Q > 0$, and J_e denotes the practical energy consumption of swarm system (1) as a whole. Let $J_{\max}$ be the maximum energy supply of swarm system (1); that is, the energy budget is limited. The definition of the limited-budget formation control of a swarm system is proposed as follows.

Definition 1. For any given $J_{\max} > 0$, swarm system (1) is said to be *limited-budget formation achievable* by control protocol (2) if there exists a gain matrix K such that $\lim_{t\longrightarrow+\infty}(x_k(t) - \eta_k(t) - \eta_c(t)) = 0$ $(k = 1, 2, \cdots, N)$ and $J_e \le J_{\max}$ for any bounded disagreement initial conditions $x_k(0) - \eta_k(0)$ $(k = 1, 2, \cdots, N)$, where $\eta_c(t)$ is said to be *the formation center function*.

It can be found from Definition 1 that the formation control is equivalent to the consensus control if $\eta_k(t) \equiv 0$ $(k = 1, 2, \cdots, N)$, which means that the consensus control can be regarded as a special case of the formation control. Besides, the vector $\eta_k(t)$ denotes the desired formation shape of the swarm systems and the state of all agents should track $\eta_k(t)$ with the formation center function.

The main objective of this paper contains two aspects. The first one is to design the gain matrix K such that swarm system (1) with control protocol (2) achieves limited-budget formation. The second one is to determine an explicit expression of the formation center function.

Remark 1. It should be pointed out that the vector-valued function $\eta(t)$ can be used to design the arbitrary geometric structure with the piecewise continuous differentiable property, but it may be unfeasible for some agent dynamics of a swarm system; that is, the formation achievable property is associated with the mechanism structure of practical dynamic agents in a swarm system. Moreover, the energy consumption of a practical swarm system is limited and control protocol (2) imposes an integral term to guarantee that the actual energy consumption is less than the maximum energy supply. This makes it challenging to construct the relationship between the gain matrix K and the limited-budget $J_{\max}$, that is, how to design the gain matrix K such that swarm system (1) with control protocol (2) achieves formation under the condition that the maximum energy supply is larger than or equal to the actual energy consumption.

Remark 2. Notice that the practical energy consumption J_e is constructed according to the weight matrix Q and is an integral quadratic associated with Q. In the design of the gain matrix, the matrix Q is often considered as a diagonal matrix, whose kth element in the diagonal line denotes the weight to the kth channel of the control protocol. By adjusting the value of the matrix Q, the contribution of the control protocol to the practical energy consumption J_e will be changed, and Q is utilized to determine the matrix variable of the matrix inequality conditions. Moreover, compared with the formation control protocols of [24–29], control protocol (2) introduces the practical energy consumption J_e to save the control energy under the limited-budget constraint; that is, the control gain K should be designed to satisfy that $J_e \le J_{\max}$ with the weight matrix Q.

3. Main Results

Based on the matrix inequality tool, this section gives sufficient conditions for limited-budget formation design and analysis, respectively. Then, an explicit expression of the formation center function is proposed, which contains two parts: the formation function one and the cooperative state one.

For $k = 1, 2, \cdots, N$, let $\varsigma_k(t) = x_k(t) - \eta_k(t)$, then it can be obtained by (1) and (2) that

$$\dot{\varsigma}_k(t) = A\left(\varsigma_k(t) + \eta_k(t)\right) + BK\sum_{i\in N_k} w_{ki}\left(\varsigma_i(t) - \varsigma_k(t)\right) - \dot{\eta}_k(t). \tag{3}$$

Let $\varsigma(t) = [\varsigma_1^T(t), \varsigma_2^T(t), \cdots, \varsigma_N^T(t)]^T$, then one can transform swarm system (3) into

$$\dot{\varsigma}(t) = \left(I_N \otimes A - L\otimes BK\right)\varsigma(t) + \left(I_N \otimes A\right)\eta(t) - \dot{\eta}(t). \tag{4}$$

It is assumed that the communication topology of swarm system (1) is undirected and connected, so the Laplacian matrix L is symmetric and positive semidefinite, which owns a simple zero eigenvalue and $N - 1$ positive eigenvalues. Hence, one can find an orthonormal matrix $U = [1_N/\sqrt{N}, \widetilde{U}]$ such that $U^T LU = \mathrm{diag}\{0, \lambda_2, \cdots, \lambda_N\}$, where $0 < \lambda_2 \le \lambda_3 \le \cdots \le \lambda_N$ denote the positive eigenvalues of the Laplacian matrix L. One can set that $\widetilde{\varsigma}(t) = (U^T \otimes I_n)\varsigma(t) = [\widetilde{\varsigma}_1^T(t), \widetilde{\varsigma}_2^T(t), \cdots, \widetilde{\varsigma}_N^T(t)]^T$, then one can transform swarm system (4) into

$$\dot{\widetilde{\varsigma}}_1(t) = A\widetilde{\varsigma}_1(t) + \left(\frac{1}{\sqrt{N}}1_N^T \otimes A\right)\eta(t) - \left(\frac{1}{\sqrt{N}}1_N^T \otimes I_n\right)\dot{\eta}(t), \tag{5}$$

$$\dot{\widetilde{\varsigma}}_i(t) = \left(A - \lambda_i BK\right)\widetilde{\varsigma}_i(t) + \left(e_i^T U^T \otimes A\right)\eta(t) - \left(e_i^T U^T \otimes I_n\right)\dot{\eta}(t), \tag{6}$$

where e_i $(i = 2, 3, \cdots, N)$ represent N-dimensional column vectors with the ith part being 1 and zero elsewhere.

The following theorem shows a sufficient condition of limited-budget formation design for swarm system (1) with a fixed communication topology on the basis of the matrix inequality tool, which proposes a design approach of the gain matrix K such that swarm system (1) with control protocol (2) achieves limited-budget formation.

Theorem 1. *For any given $J_{\max} > 0$ and $W = I_N - N^{-1}1_N 1_N^T$, if $A\eta_k(t) - \dot{\eta}_k(t) = 0$ $(k = 1, 2, \cdots, N)$ and there exists $P^T = P > 0$ such that*

$$\begin{aligned}&(x(0)-\eta(0))^T(W\otimes I_n)(x(0)-\eta(0))P\le J_{\max}I_n,\\&A^TP+PA-PBB^TP+0.25\lambda_2^{-2}\lambda_N^2PBQB^TP<0,\end{aligned}\tag{7}$$

then swarm system (1) is limited-budget formation achievable by control protocol (2) with $K=0.5\lambda_2^{-1}B^TP$.

Proof of Theorem 1. Define

$$\varsigma_c(t)\triangleq(U\otimes I_n)\left[\tilde{\varsigma}_1^T(t),0\right]^T,\tag{8}$$

$$\varsigma_{\tilde{c}}(t)\triangleq(U\otimes I_n)\left[0,\tilde{\varsigma}_2^T(t),\cdots,\tilde{\varsigma}_N^T(t)\right]^T.\tag{9}$$

One can find that

$$(U\otimes I_n)\left[\tilde{\varsigma}_1^T(t),0\right]^T=\frac{1}{\sqrt{N}}1_N\otimes\tilde{\varsigma}_1(t),\tag{10}$$

$$(U\otimes I_n)\left[0,\tilde{\varsigma}_2^T(t),\cdots,\tilde{\varsigma}_N^T(t)\right]^T=\sum_{i=2}^{N}Ue_i\otimes\tilde{\varsigma}_i(t).\tag{11}$$

Since $\tilde{\varsigma}(t)=(U^T\otimes I_n)\varsigma(t)$, it can be shown that

$$\varsigma(t)=(U\otimes I_n)\tilde{\varsigma}(t),\tag{12}$$

that is,

$$\varsigma(t)=\varsigma_c(t)+\varsigma_{\tilde{c}}(t).\tag{13}$$

In this case, the transformation matrix U is orthonormal and $\varsigma_k(t)=x_k(t)-\eta_k(t)$ $(k=1,2,\ldots N)$, so it can be found from (8)–(11) that the component $\tilde{\varsigma}_1(t)/\sqrt{N}$ can be used to describe the formation center function $\eta_c(t)$ as the time tends to infinity and swarm system (1) with control protocol (2) achieves formation if and only if all the components $\tilde{\varsigma}_i(t)\,(i=2,3,\cdots,N)$ tend to zero as the time tends to infinity.

In the following, an approach is shown to design the gain matrix K such that all the components $\tilde{\varsigma}_i(t)\,(i=2,3,\cdots,N)$ tend to zero as the time tends to infinity. Let $P^T=P>0$, then one can construct a Lyapunov function candidate as follows:

$$V_i(t)=\tilde{\varsigma}_i^T(t)P\tilde{\varsigma}_i(t)\,(i=2,3,\cdots,N).\tag{14}$$

By taking the time derivative of $V_i(t)$ along the trajectories of subsystems (6), one can deduce that

$$\begin{aligned}\dot{V}_i(t)=&\,\tilde{\varsigma}_i^T(t)\left((A-\lambda_iBK)^TP+P(A-\lambda_iBK)\right)\tilde{\varsigma}_i(t)\\&+2\tilde{\varsigma}_i^T(t)P\left(e_i^TU^T\otimes A\right)\eta(t)-2\tilde{\varsigma}_i^T(t)P\left(e_i^TU^T\otimes I_n\right)\dot{\eta}(t).\end{aligned}\tag{15}$$

Let $K=0.5\lambda_2^{-1}B^TP$, then it can be found that

$$(A-\lambda_iBK)^TP+P(A-\lambda_iBK)\le A^TP+PA-\lambda_i\lambda_2^{-1}PBB^TP.\tag{16}$$

Due to $0<\lambda_2\le\lambda_3\le\cdots\le\lambda_N$, it can be shown that $\lambda_i\lambda_2^{-1}\ge1\,(i=2,3,\cdots,N)$. Hence, it can be found by (16) that

$$A^TP+PA-\lambda_i\lambda_2^{-1}PBB^TP\le A^TP+PA-PBB^TP.\tag{17}$$

One can set that

$$\begin{aligned}\Pi(t)=\Big[&\left(A\eta_1(t)-\dot{\eta}_1(t)\right)^T,\left(A\eta_2(t)-\dot{\eta}_2(t)\right)^T,\cdots,\\&\left(A\eta_N(t)-\dot{\eta}_N(t)\right)^T\Big]^T.\end{aligned}\tag{18}$$

Since

$$e_i^TU^T\otimes A=\left(e_i^TU^T\otimes I_n\right)(I_N\otimes A),\tag{19}$$

it can be derived by (18) that

$$\left(e_i^TU^T\otimes A\right)\eta(t)-\left(e_i^TU^T\otimes I_n\right)\dot{\eta}(t)=\left(e_i^TU^T\otimes I_n\right)\Pi(t).\tag{20}$$

If

$$A\eta_k(t)-\dot{\eta}_k(t)=0\ (k=1,2,\cdots,N),\tag{21}$$

then one can show by (18) that

$$\Pi(t)=0.\tag{22}$$

In this case, since $U^T\otimes I_n$ is invertible, one can find that

$$\left(e_i^TU^T\otimes A\right)\eta(t)-\left(e_i^TU^T\otimes I_n\right)\dot{\eta}(t)=0.\tag{23}$$

Thus, if $A^TP+PA-PBB^TP<0$, then $\lim_{t\longrightarrow+\infty}\tilde{\varsigma}_i(t)=0$ $(i=2,3,\cdots,N)$ by (15), (17), and (23). Hence, swarm system (1) with a fixed communication topology is formation achievable by control protocol (2) with $K=0.5\lambda_2^{-1}B^TP$.

Finally, the influences of the limited budget are analyzed. By (2), one can find that

$$J_e=\int_0^{+\infty}\varsigma^T(t)\left(L^2\otimes K^TQK\right)\varsigma(t)\mathrm{d}t.\tag{24}$$

By $K=0.5\lambda_2^{-1}B^TP$ and $\tilde{\varsigma}(t)=(U^T\otimes I_n)\varsigma(t)$, it can be deduced by (27) that

$$\varsigma^T(t)\left(L^2\otimes K^TQK\right)\varsigma(t)\le0.25\lambda_2^{-2}\sum_{i=2}^{N}\lambda_i^2\tilde{\varsigma}_i^T(t)PBQB^TP\tilde{\varsigma}_i(t).\tag{25}$$

Let $\phi\ge0$, then one has

$$\begin{aligned}J_e^{\phi}=&\int_0^{\phi}\varsigma^T(t)\left(L^2\otimes K^TQK\right)\varsigma(t)\mathrm{d}t\\&\le0.25\lambda_2^{-2}\sum_{i=2}^{N}\int_0^{\phi}\lambda_i^2\tilde{\varsigma}_i^T(t)PBQB^TP\tilde{\varsigma}_i(t)\mathrm{d}t.\end{aligned}\tag{26}$$

By

$$\int_0^{\phi}\dot{V}_i(t)\mathrm{d}t=V_i(\phi)-V_i(0),\tag{27}$$

it can be derived from (27)–(29) that

$$\begin{aligned}J_e^{\phi}\le&\sum_{i=2}^{N}\int_0^{\phi}\Big(\left(\dot{V}_i(t)+0.25\lambda_2^{-2}\lambda_i^2\tilde{\varsigma}_i^T(t)PBQB^TP\tilde{\varsigma}_i(t)\right)\mathrm{d}t\\&-V_i(\phi)\Big)+\sum_{i=2}^{N}V_i(0).\end{aligned}\tag{28}$$

Since λ_N is the maximum eigenvalue of the Laplacian matrix L and $\lambda_i\lambda_2^{-1} \geq 1$ $(i = 2, 3, \cdots, N)$, one can derive by (31) that

$$J_e^{\phi} \leq \sum_{i=2}^{N} \int_0^{\phi} \Big(\big(\dot{V}_i(t) + 0.25\lambda_2^{-2}\lambda_N^2 \tilde{\varsigma}_i^T(t) PBQB^T P \tilde{\varsigma}_i(t) \big) \mathrm{d}t - V_i(\phi) \Big) + \sum_{i=2}^{N} V_i(0). \tag{29}$$

If

$$A^T P + PA - PBB^T P + 0.25\lambda_2^{-2}\lambda_N^2 PBQB^T P < 0, \tag{30}$$

then all the components $\tilde{\varsigma}_i(t)$ $(i = 2, 3, \cdots, N)$ tend to zero as the time t tends to infinity, and $V_i(\phi)$ $(i = 2, 3, \cdots, N)$ tend to zero as the parameter ϕ the time tends to infinity; that is, $\lim_{t \longrightarrow +\infty} \tilde{\varsigma}_i(t) = 0$ and $\lim_{\phi \longrightarrow +\infty} V_i(\phi) = 0$ $(i = 2, 3, \cdots, N)$. Thus, it can be found by (27) and (30) that

$$J_e \leq \sum_{i=2}^{N} \tilde{\varsigma}_i^T(0) P \tilde{\varsigma}_i(0). \tag{31}$$

Due to

$$\tilde{\varsigma}_i(0) = \left(e_i^T U^T \otimes I_n \right) \varsigma(0) \, (i = 2, 3, \cdots, N), \tag{32}$$

one can derive that

$$\sum_{i=2}^{N} \tilde{\varsigma}_i^T(0) P \tilde{\varsigma}_i(0) = \varsigma^T(0) \left(\begin{bmatrix} e_2^T U^T \\ \vdots \\ e_N^T U^T \end{bmatrix}^T \otimes I_n \right) \cdot (I_N \otimes P) \left(\begin{bmatrix} e_2^T U^T \\ \vdots \\ e_N^T U^T \end{bmatrix} \otimes I_n \right) \varsigma(0). \tag{33}$$

Since

$$U = \left[1_N / \sqrt{N}, \tilde{U} \right], \tag{34}$$

it can be shown that

$$UU^T = I_N = \frac{1}{N} 1_N 1_N^T + \tilde{U}\tilde{U}^T, \tag{35}$$

that is,

$$\tilde{U}\tilde{U}^T = I_N - \frac{1}{N} 1_N 1_N^T. \tag{36}$$

Since $[Ue_2, Ue_3, \cdots, Ue_N] = \tilde{U}$, from (34)–(36), it can be found that

$$J_e \leq \varsigma^T(0)(W \otimes P)\varsigma(0), \tag{37}$$

where $W = I_N - N^{-1} 1_N 1_N^T$. It is assumed that $x_k(0) - f_k(0)$ where $(k = 1, 2, \cdots, N)$ are disagreement, so there always exists some $\tilde{\varsigma}_i(0) \neq 0$ $(i \in \{2, 3, \cdots, N\})$. Hence, it can be obtained that

$$\varsigma^T(0)(W \otimes I_n)\varsigma(0) = \sum_{i=2}^{N} \tilde{\varsigma}_i^T(0) \tilde{\varsigma}_i(0) > 0. \tag{38}$$

In this case, there exists a positive scalar ξ such that

$$J_{\max} = \varsigma^T(0)(W \otimes \xi I_n)\varsigma(0). \tag{39}$$

The matrix W has a simple zero eigenvalue and $N - 1$ nonzero eigenvalues are positive, so it can be derived by (37) and (39) that $P \leq \xi I_n$ can guarantee that $J_e \leq J_{\max}$, which means that

$$(x(0) - \eta(0))^T (W \otimes I_n)(x(0) - \eta(0)) P \leq J_{\max} I_n. \tag{40}$$

Thus, the Proof of Theorem 1 is completed.

In Theorem 1, a limited-budget formation design criterion is proposed, which gives a design approach of the gain matrix K of control protocol (2) to make swarm system (1) achieve limited-budget formation. If the gain matrix K is given, then the following theorem shows a limited-budget formation analysis criterion, which can be obtained directly according to the Proof of Theorem 1. □

Theorem 2. *For any given $J_{\max} > 0$, K and $W = I_N - N^{-1} 1_N 1_N^T$, swarm system (1) with control protocol (2) achieves limited-budget formation if $A\eta_k(t) - \dot{\eta}_k(t) = 0$ $(k = 1, 2, \cdots, N)$ and there exists $P^T = P > 0$ such that*

$$(x(0) - \eta(0))^T (W \otimes I_n)(x(0) - \eta(0)) P \leq J_{\max} I_n, \tag{41}$$

$$A^T P + PA - \lambda_i PBK - \lambda_i K^T B^T P + \lambda_N^2 K^T QK < 0 \, (i = 2, 3, \cdots, N). \tag{42}$$

There exist two difficulties to check the matrix inequalities in Theorem 2. The first one is that it is not scalable since the number of constraint conditions $A^T P + PA - \lambda_i PBK - \lambda_i K^T B^T P + \lambda_N^2 K^T QK < 0 \, (i = 2, 3, \cdots, N)$ increases as the number of agents increases. The second one is that the term $\lambda_N^2 K^T QK$ makes matrix inequalities $A^T P + PA - \lambda_i PBK - \lambda_i K^T B^T P + \lambda_N^2 K^T QK < 0 \, (i = 2, 3, \cdots, N)$ not linear, which are difficult to be checked. By the convex feature of linear matrix inequalities and the Schur lemma in [34], the following theorem can be obtained on the basis of Theorem 2, which is scalable and linear and can be checked by the feasp solver in the Matlab's LMI toolbox given in [35] and will bring no conservatism from Theorem 2 to Theorem 3.

Theorem 3. *For any given $J_{\max} > 0$, K and $W = I_N - N^{-1} 1_N 1_N^T$, swarm system (1) with control protocol (2) achieves limited-budget formation if $A\eta_k(t) - \dot{\eta}_k(t) = 0$ $(k = 1, 2, \cdots, N)$ and there exists $P^T = P > 0$ such that*

$$(x(0) - \eta(0))^T (W \otimes I_n)(x(0) - \eta(0)) P \leq J_{\max} I_n, \tag{43}$$

$$\begin{bmatrix} A^T P + PA - \lambda_i PBK - \lambda_i K^T B^T P & \lambda_N K^T Q \\ \lambda_N QK & -Q \end{bmatrix} < 0 \, (i = 2, N). \tag{44}$$

Remark 3. Intuitively speaking, by designing piecewise continuous differentiable $\eta_k(t)$ $(k = 1, 2, \cdots, N)$, arbitrarily required formation structures for swarm system (1) with control protocol (2) to maintain can be obtained. However,

the achievability of a geometric structure is dependent on the structure property of a swarm system; that is, the achievability is closely related to the system matrix A. If the formation functions $\eta_k(t)(k=1,2,\cdots,N)$ are not time-varying, that is, $\eta_k(t)(k=1,2,\cdots,N)$ are constant, then the condition $A\eta_k(t)=0$ $(k=1,2,\cdots,N)$ is needed for swarm system (1) with control protocol (2) to achieve time-variant formation structures as shown in [15]. If the formation functions $\eta_k(t)(k=1,2,\cdots,N)$ are time-varying, then Theorem 1 requires that the condition $A\eta_k(t)-\dot{\eta}_k(t)=0$ $(k=1,2,\cdots,N)$ for swarm system (1) with control protocol (2) to achieve time-varying formation structures. The literature studies [25, 26] showed some similar feasible conditions for swarm systems to achieve time-varying formation structures, where feasible conditions are different from the one in Theorem 1 since the limited budget was not taken into consideration.

Remark 4. Because the whole energy budget must be limited in practical engineering swarm systems, it is critically important to analyze the impacts of the limited budget for swarm system (1) with control protocol (2) to achieve specific formation; that is, swarm system (1) with control protocol (2) can obtain and maintain specific geometric structures under the condition that the whole energy consumption is restricted. Note that the condition $(x(0)-\eta(0))^T(W\otimes I_n)(x(0)-\eta(0))P\le J_{\max}I_n$ introduces the maximum energy supply $J_{\max}$ into the gain matrix design, where the initial condition $x(0)-\eta(0)$ and the interaction relationship matrix W are used. The initial condition is often available in practical applications and the interaction relationship matrix shows the impact of the topology. By the structure property of the transformation matrix U, Theorem 1 proposes an approach to construct the interaction relationship between the limited budget and the matrix variable, which introduces the limited budget into the formation criterion. It should be pointed out that the interaction relationship matrix $W=I_N-N^{-1}1_N1_N^T$ can be regarded as the Laplacian matrix of a complete graph with all edge weights being N^{-1}. In this case, the interaction relationship matrix has a simple zero eigenvalue and $N-1$ identical nonzero eigenvalues. Moreover, the sufficient conditions in Theorem 1 for swarm system (1) with control protocol (2) to achieve limited-budget formation are scalable since they are independent of the number of agents and the two eigenvalues λ_2 and λ_N of the Laplacian matrix can be estimated by the approaches in [36, 37].

By the analysis of Theorem 1, the component $\tilde{\varsigma}_1(t)/\sqrt{N}$ can be applied to depict the formation center function $\eta_c(t)$ as the time tends to infinity, which is associated with subsystem (5). The following theorem shows a method to give an explicit expression of the formation center function.

Theorem 4. *If swarm system (1) with control protocol (2) achieves limited-budget formation, then the formation center function satisfies that* $\lim_{t\longrightarrow+\infty}(\eta_c(t)-\eta_{cf}(t)-\eta_{cx}(t))=0$, *where* $\eta_{cf}(t)=-N^{-1}\sum_{k=1}^{N}\eta_k(t)$ *and* $\eta_{cx}(t)=N^{-1}e^{At}\sum_{k=1}^{N}x_k(0)$.

Proof of Theorem 4. Due to

$$\tilde{\varsigma}(0)=\left(U^T\otimes I_n\right)\varsigma(0), \tag{45}$$

$$\varsigma_k(0)=x_k(0)-\eta_k(0)(k=1,2,\cdots,N), \tag{46}$$

it can be shown that

$$\tilde{\varsigma}_1(0)=\left(e_1^TU^T\otimes I_n\right)\varsigma(0)=\frac{1}{\sqrt{N}}\left(\sum_{k=1}^{N}x_k(0)-\sum_{k=1}^{N}\eta_k(0)\right). \tag{47}$$

Furthermore, the dynamic response of subsystem (5) introduced by initial conditions is that

$$\varsigma_0(t)=e^{At}\tilde{\varsigma}_1(0). \tag{48}$$

By (47) and (48), one can see that

$$\varsigma_0(t)=\frac{1}{\sqrt{N}}e^{At}\left(\sum_{k=1}^{N}x_k(0)-\sum_{k=1}^{N}\eta_k(0)\right). \tag{49}$$

Moreover, the dynamic response of subsystem (5) introduced by the formation function $\eta(t)$ is that

$$\varsigma_\eta(t)=\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\int_0^t e^{A(t-\varphi)}A\eta_k(\varphi)\mathrm{d}\varphi. \tag{50}$$

In particular, the dynamic response of subsystem (5) introduced by the derivative of the formation function $\dot{\eta}(t)$ is that

$$\begin{aligned}\varsigma_{\dot{\eta}}(t)&=-\frac{1}{\sqrt{N}}\int_0^t e^{A(t-\varphi)}\left(1_N^T\otimes I_n\right)\dot{\eta}(\varphi)\mathrm{d}\varphi\\&=-\frac{1}{\sqrt{N}}\left(\sum_{k=1}^{N}\eta_k(t)+e^{At}\sum_{k=1}^{N}\eta_k(0)\right)\\&\quad-\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\int_0^t Ae^{A(t-\varphi)}\eta_k(\varphi)\mathrm{d}\varphi.\end{aligned} \tag{51}$$

The component $\tilde{\varsigma}_1(t)/\sqrt{N}$ can be applied to depict the formation center function $\eta_c(t)$ as the time tends to infinity, so from (50) and (51), one can show that

$$\lim_{t\longrightarrow+\infty}\left(\eta_c(t)-\eta_{cf}(t)-\eta_{cx}(t)\right)=0, \tag{52}$$

where

$$\eta_{cf}(t)=-\frac{1}{N}\sum_{k=1}^{N}\eta_k(t), \tag{53}$$

$$\eta_{cx}(t)=\frac{1}{N}e^{At}\sum_{k=1}^{N}x_k(0). \tag{54}$$

Thus, the Proof of Theorem 4 is completed. □

Remark 5. Limited-budget formation motions of a swarm system with a fixed communication topology contain two parts. The first part is the macroscopic motion as a whole and

the second part is the microcosmic motions between any two agents. Limited-budget formation is obtained if and only if the microcosmic motions between any two agents are asymptotically stable. If limited-budget formation is achieved, then all agents in a swarm system move along the trajectories of the formation center function, which can be used to depict the macroscopic motion as a whole. From Theorem 4, the formation center function involves two parts, which are associated with the average value of the time-varying formation functions and the average value of the initial conditions of cooperative states, which are called the formation function part and the cooperative state part, but it is not associated with the derivatives of the formation functions; that is, the variance of the formation structure does not impact the macroscopic motion as a whole. In particular, according to the explicit expression of the formation center function in Theorem 4, the limited budget of energy assumptions does not also influence the macroscopic motion as a whole.

4. Numerical Simulations

This section provides the numerical simulation to demonstrate the correctness of our main conclusions about limited-budget formation control for swarm systems with fixed communication topologies.

Consider a swarm system with five agents, where the dynamics of each agent is modeled by (1) with

$$A = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & -2 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ -0.5 \\ -1 \end{bmatrix}. \tag{55}$$

The communication topology of this swarm system is depicted in Figure 1, where the edge weight is set to be 0–1; that is, the weight of the connected edge is one and the weight of the unconnected edge is zero. The initial conditions of the cooperative states are presented as follows:

$$\begin{aligned} x_1(0) &= [3.2, 1.8, -2.1]^T, \\ x_2(0) &= [4.4, -2.7, 0.2]^T, \\ x_3(0) &= [-3.3, 1.5, 6.1]^T, \\ x_4(0) &= [-0.8, 2.9, 1.4]^T, \\ x_5(0) &= [-1.1, 3.9, -3.8]^T. \end{aligned} \tag{56}$$

The formation functions are chosen as follows:

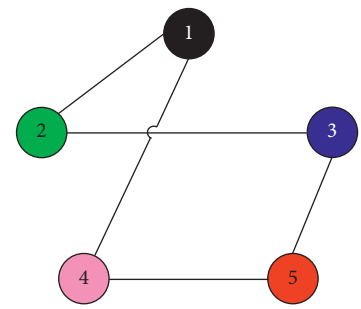


Figure 1: Communication topology.

$$\eta_k(t) = \begin{bmatrix} \sin\left(t + \frac{2(k-1)\pi}{5}\right) \\ -\sin\left(t + \frac{2(k-1)\pi}{5}\right) \\ -\cos\left(t + \frac{2(k-1)\pi}{5}\right) \end{bmatrix} (k = 1, 2, \cdots, 5). \tag{57}$$

One can find that the formation functions $\eta_k(t)\,(k = 1, 2, \cdots, 5)$ satisfy the feasibility conditions $A\eta_k(t) - \dot{\eta}_k(t) = 0\,(k = 1, 2, \cdots, 5)$ in Theorem 1. Let $Q = 0.1$ and $J_{\max} = 1000$. By the feasp solver in the Matlab's LMI toolbox, it can be found from Theorem 1 that

$$P = \begin{bmatrix} 7.0480 & 0.9924 & 0.0699 \\ 0.9924 & 8.4394 & 1.2820 \\ 0.0699 & 1.2820 & 3.3207 \end{bmatrix}. \tag{58}$$

In this case, the gain matrix is $K = [2.3452, -1.6315, -1.4081]$.

In Figure 2, the trajectories of $\varsigma_k(t)\,(k = 1, 2, \cdots, 5)$ for this swarm system are presented, where the curves depicted by blue star markers represent the trajectories of the formation center function. Figure 3 shows the state snapshots of five agents and the formation center at $t = 0$ s, $t = 8$ s, $t = 9$ s, and $t = 10$ s, where five agents are depicted by black circles, green plusses, blue asterisks, pink x-marks, and red hexagrams, orderly. By Figure 3, one can find that all the agents achieve a time-varying pentagon around the formation center at different times.

The formation center of this swam system counterclockwise rotates about 1.62 cycles as shown in Figure 4, where the initial state is denoted by a red pentacle and the final state is represented by a black pentacle.

Figure 5 shows that the practical energy consumption $J_e(t)$ converges to a finite value less than $J_{\max}$. From Figures 2–5, one can see that this swam system achieves the desired pentagonal formation in the form of the circular movement and the limited budget is guaranteed.

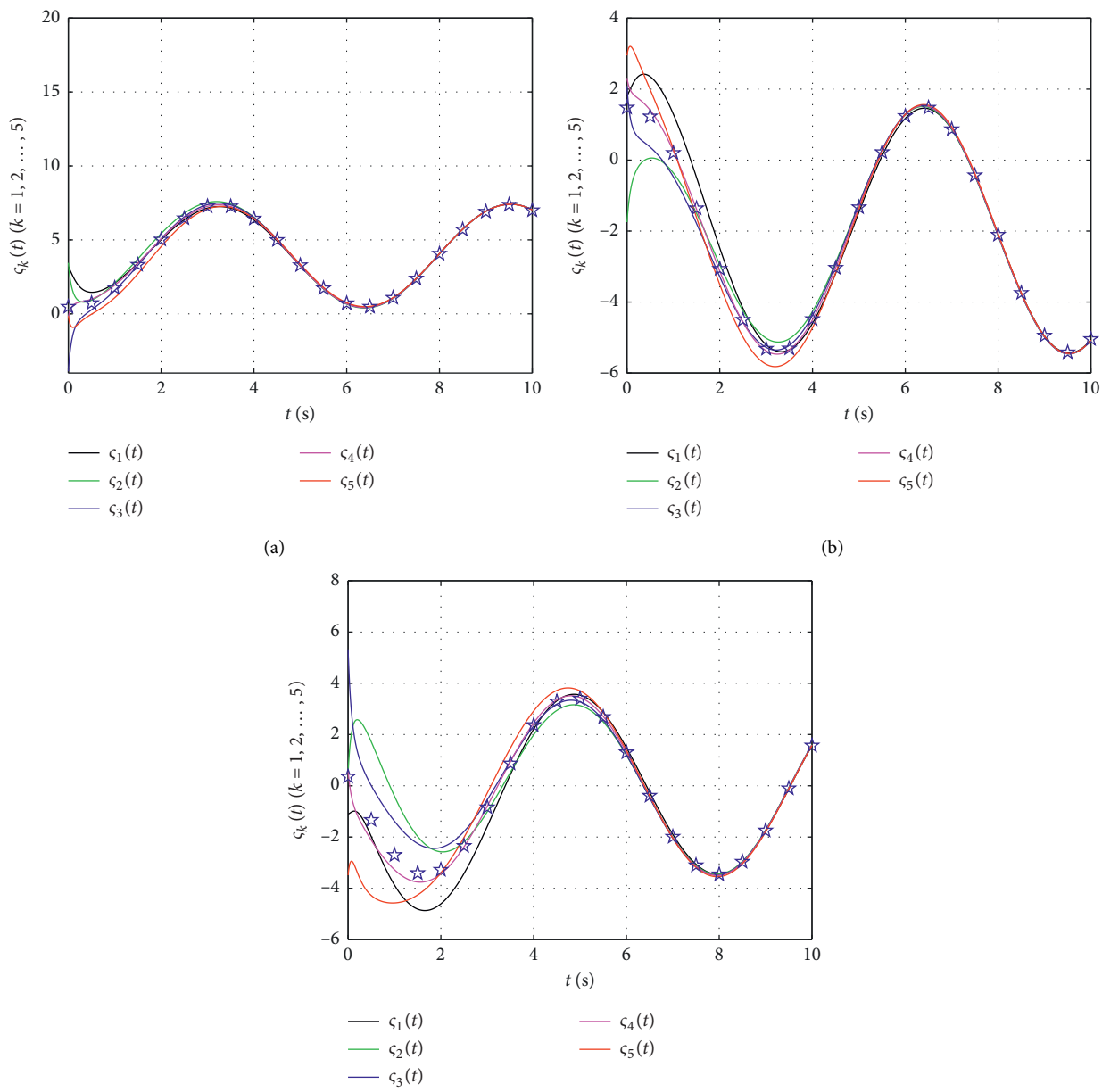


Figure 2: Trajectories of $\varsigma_k(t)\,(k = 1, 2, \cdots, 5)$.

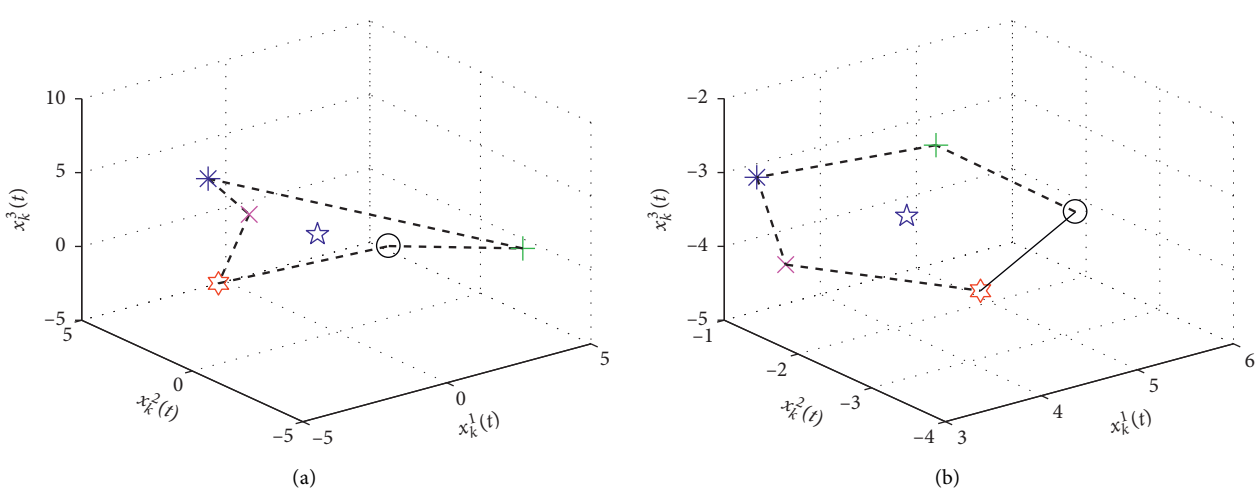


Figure 3: Continued.

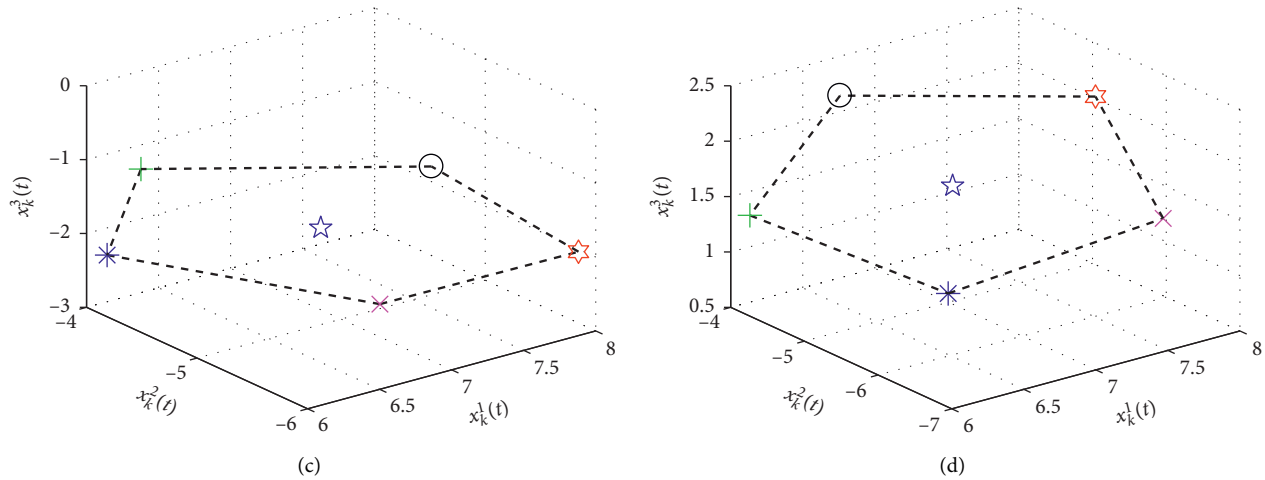


Figure 3: State snapshots of all the agents at different times: (a) $t = 0$ s, (b) $t = 8$ s, (c) $t = 9$ s, and (d) $t = 10$ s.

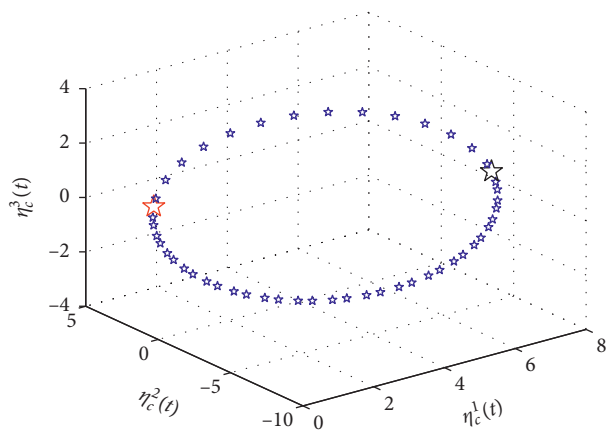


Figure 4: Movement track of the formation center function.

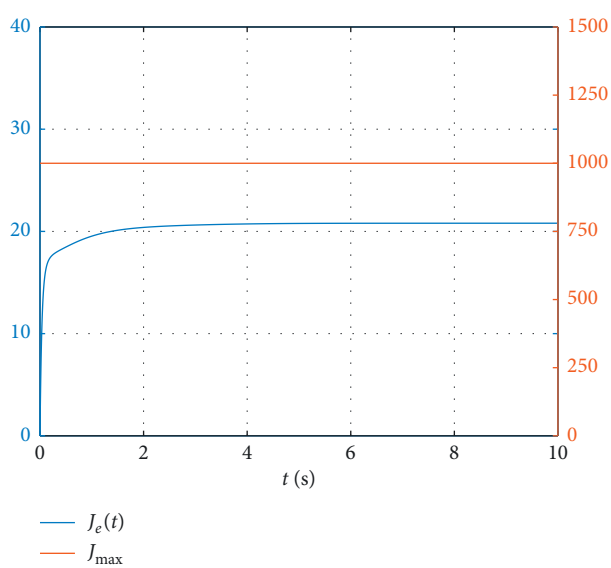


Figure 5: Practical energy consumption and the limited budget.

5. Conclusions

For high-order linear swarm systems with fixed communication topologies, a new formation control protocol with an energy integral term was proposed by using the state errors and the formation function errors among neighboring agents, which can guarantee that the practical energy assumption is less than the whole energy budget. Furthermore, by an orthonormal transformation associated with the Laplacian matrix of the fixed communication topology, limited-budget formation control problems were converted into asymptotic stability ones, and sufficient conditions for swam systems to achieve limited-budget formation were proposed, where the relationship between the matrix variable and the limited budget was constructed on the basis of the specific structure feature of the orthonormal transformation matrix. In particular, those criteria are scalable and checkable since their constraints are independent of the number of agents and can be converted into linear matrix inequalities. Moreover, an explicit expression of the formation center function was presented, which contains two independent parts associated with the formation functions and the cooperative states, but it is independent of the variances of the formation functions.

The further works will focus on extending the current results to the swarm systems with heterogeneous dynamics, and the fixed topology in this paper will be changed to the switching connected topologies and the jointly switching topologies.

Data Availability

The data used to support the findings of the study are available in Section 4 of this paper.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Hongtao Dang and Le Wang conceptualized the study and prepared the original draft; Hongtao Dang investigated the data and was responsible for methodology and funding acquisition and involved in project administration; Yan Zhang validated the data; Jianye Yang performed formal analysis; Yan Zhang and Jianye Yang reviewed and edited the manuscript; Le Wang supervised the study. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the Key Research and Development Program of Shaanxi (no. 2019GY-025), also funded by the National Natural Science Foundation of China under Grants 61867005.

References

[1] J. Zhou, X. Wu, W. Yu, M. Small, and J.-A. Lu, "Flocking of multi-agent dynamical systems based on pseudo-leader mechanism," *Systems & Control Letters*, vol. 61, no. 1, pp. 195–202, 2012.

[2] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.

[3] J. Xi, C. Wang, H. Liu, and L. Wang, "Completely distributed guaranteed-performance consensualization for high-order multiagent systems with switching topologies," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1338–1348, 2019.

[4] J. Qu, Z. Ji, C. Lin, and H. Yu, "Fast consensus seeking on networks with antagonistic interactions," *Complexity*, vol. 78, 2018.

[5] J. Sun, Z. Geng, Y. Lv, Z. Li, and Z. Ding, "Distributed adaptive consensus disturbance rejection for multi-agent systems on directed graphs," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 202–212, 2018.

[6] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 1, 2020.

[7] Y. Zhang, H. Li, J. Sun, and W. He, "Cooperative adaptive event-triggered control for multiagent systems with actuator failures," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1759–1768, 2019.

[8] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 2442–2454, 2020.

[9] Y. Zhu, S. Li, J. Ma, and Y. Zheng, "Bipartite consensus in networks of agents with antagonistic interactions and quantization," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 12, pp. 2012–2016, 2018.

[10] Y. Zheng, Q. Zhao, J. Ma, and L. Wang, "Second-order consensus of hybrid multi-agent systems," *Systems & Control Letters*, vol. 125, pp. 51–58, 2019.

[11] Y. Zheng, J. Ma, and L. Wang, "Consensus of hybrid multi-agent systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 1359–1365, 2018.

[12] L. Consolini, F. Morbidi, D. Prattichizzo, and M. Tosques, "Leader-follower formation control of nonholonomic mobile robots with input constraints," *Automatica*, vol. 44, no. 5, pp. 1343–1349, 2008.

[13] K.-K. Oh and H.-S. Ahn, "Formation control of mobile agents based on distributed position estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 737–742, 2013.

[14] H. Liu, T. Ma, F. L. Lewis, and Y. Wan, "Robust formation control for multiple quadrotors with nonlinearities and disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1362–1371, 2020.

[15] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, vol. 50, 2020.

[16] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[17] S. Liu, Z. Ji, and H. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, 2020.

[18] X.-G. Guo, J.-L. Wang, F. Liao, and D. Wang, "Quantized H_{∞} consensus of multiagent systems with quantization mismatch under switching weighted topologies," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 2, pp. 202–212, 2017.

[19] Z.-Y. Tan, N. Cai, J. Zhou, and S.-G. Zhang, "On performance of peer review for academic journals: analysis based on distributed parallel system," *IEEE Access*, vol. 7, pp. 19024–19032, 2019.

[20] W. Ren, "Consensus strategies for cooperative control of vehicle formations," *IET Control Theory & Applications*, vol. 1, no. 2, pp. 505–512, 2007.

[21] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1465–1476, 2004.

[22] H. Du, G. Wen, Y. Cheng, Y. He, and R. Jia, "Distributed finite-time cooperative control of multiple high-order nonholonomic mobile robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2998–3006, 2016.

[23] W. Qin, Z. Liu, and Z. Chen, "A novel observer-based formation for nonlinear multi-agent systems with time delay and intermittent communication," *Nonlinear Dynamics*, vol. 79, no. 3, pp. 1651–1664, 2015.

[24] L. Brinon-Arranz, A. Seuret, and C. Canudas-de-Wit, "Cooperative control design for time-varying formations of multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2283–2288, 2014.

[25] R. Rahimi, F. Abdollahi, and K. Naqshi, "Time-varying formation control of a collaborative heterogeneous multi agent system," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1799–1805, 2014.

[26] X. Dong and G. Hu, "Time-varying formation control for general linear multi-agent systems with switching directed topologies," *Automatica*, vol. 73, no. 73, pp. 47–55, 2016.

[27] F. Xiao, L. Wang, J. Chen, and Y. Gao, "Finite-time formation control for multi-agent systems," *Automatica*, vol. 45, no. 11, pp. 2605–2611, 2009.

[28] X. W. Dong, Y. Zhou, Z. Ren, and Y. S. Zhong, "Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5014–5024, 2016.

[29] X. Dong and G. Hu, "Time-varying formation tracking for linear multiagent systems with multiple leaders," *IEEE*

Transactions on Automatic Control, vol. 62, no. 7, pp. 3658–3664, 2017.

[30] J. Xi, Z. Fan, H. Liu, and T. Zheng, "Guaranteed-cost consensus for multiagent networks with Lipschitz nonlinear dynamics and switching topologies," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 7, pp. 2841–2852, 2018.

[31] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[32] J. Yu, X. Dong, Q. Li, and Z. Ren, "Robust ${H_\infty}$ guaranteed cost time-varying formation tracking for high-order multiagent systems with time-varying delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, pp. 1465–1475, 2020.

[33] C. Godsil and G. Royal, *Algebraic Graph Theory*, Springer-Verlag, New York, NY, USA, 2001.

[34] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, PA, USA, 1994.

[35] P. Gahinet, A. Nemirovskii, A. J. Laub, and M. Chilali, *LMI Control Toolbox User's Guide*, The Math Works, Natick, MA, USA, 1995.

[36] A. Berman and X. D. Zhang, "Lower bounds for the eigenvalues of Laplacian matrices," *Linear Algebra and its Applications*, vol. 316, no. 1–3, pp. 13–20, 2000.

[37] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University, London, UK, 1990.

Hindawi
Complexity
Volume 2020, Article ID 1045760, 15 pages
https://doi.org/10.1155/2020/1045760

Research Article

Existence and Stability of Square-Mean S-Asymptotically Periodic Solutions to a Fractional Stochastic Diffusion Equation with Fractional Brownian Motion

Jia Mu,[1,2] Jiecuo Nan,[2] and Yong Zhou[3,4]

[1]*Key Laboratory of Streaming Data Computing Technologies and Application, Northwest Minzu University, Lanzhou 730000, China*
[2]*School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730000, China*
[3]*School of Mathematics and Computer Science, Xiangtan University, Xiangtan, Hunan 411105, China*
[4]*Nonlinear Analysis and Applied Mathematics Research Group, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

Correspondence should be addressed to Yong Zhou; yzhou@xtu.edu.cn

Received 27 July 2020; Revised 29 August 2020; Accepted 18 September 2020; Published 5 October 2020

Academic Editor: Jianxiang Xi

In this paper, a generalized Gronwall inequality is demonstrated, playing an important role in the study of fractional differential equations. In addition, with the fixed-point theorem and the properties of Mittag–Leffler functions, some results of the existence as well as asymptotic stability of square-mean S-asymptotically periodic solutions to a fractional stochastic diffusion equation with fractional Brownian motion are obtained. In the end, an example of numerical simulation is given to illustrate the effectiveness of our theory results.

1. Introduction

Originated in 1695, fractional calculus has been widely applied in physics, chemistry, economics, biology, and other fields. Recent decades have witnessed the rapid development of fractional calculus, with the emergence of many related researches [1–9]. The dynamic behavior of some complex processes in reality can be explained by fractional differential equations. For example, anomalous diffusion phenomena can be described with fractional diffusion equations. Compared with the traditional diffusion equations (first order), with fractional diffusion equations, subdiffusion or supdiffusion phenomenon can be described when its order is between 0 and 1 or between 1 and 2, respectively.

In addition, the diffusion phenomenon in real life is often affected by random factors, promoting the generation of fractional stochastic diffusion equations. However, this research has not aroused much concern until recent years. In [10], a class of nonautonomous fractional stochastic reaction-diffusion equations was studied, obtaining the regularity of random attractors. The Galerkin method was applied by Wang [11] to investigate the existence of tempered pullback random attractors for nonautonomous fractional reaction-diffusion equations with multiplicative noise. Chen [12] studied the stochastic time-fractional diffusion equations with multiplicative white noise, obtaining the Hölder continuity of the solution. Peng and Huang [13] established the existence of mild solutions for a nonlocal backward problem for fractional stochastic diffusion equations.

We also notice that some researches focus on the stability of solutions to fractional stochastic differential equations of order $\alpha \in (0, 1)$. Li and Wang [14] studied the existence and asymptotic behavior of solutions to fractional stochastic delay evolution equations with integral term and Wiener process by using fractional resolvent operator theory and the Schauder fixed-point theorem. Mathiyalagan and Balachandran [15] studied the finite-time stochastic stability of fractional-order singular systems with time delay and white noise utilizing the Gronwall approach and stochastic analysis technique. In [16], applying the Laplace transform

method, the authors obtained the existence, uniqueness, and Hyers–Ulam stability of solutions to a class of linear fractional differential equations involving Mittag–Leffler kernel. The resolvent operator technique and contraction mapping principle were used in [17] to study the existence and uniqueness of mild solution to fractional neutral stochastic integrodifferential equations involving impulses driven by fractional Brownian motion (FBM), and a new impulsive-integral inequality was used to obtain the exponential stability for these equations. Moreover, the existence and asymptotic stability in the p-th moment of mild solutions to a class of fractional stochastic partial differential equations with Wiener process was investigated by Zhang et al. [18]. Because the form of the equations in this paper is different from those in the above studies, the methods to prove the stability in these studies cannot be directly applied to this paper.

Inspired by the above researches, in order to study the stability and periodicity of anomalous diffusion phenomena affected by random factors, we consider the fractional stochastic diffusion equation involving Dirichlet boundary conditions:

$$\begin{cases} \partial_{0,t}^{\alpha} u(x,t) + u(x,t) = f(t,u(x,t)) + D(t,u(x,t))\dfrac{\partial B_Q^H(x,t)}{\partial t}, & \text{in } \Omega \times (0,\infty), \\ u = 0, & \text{on } \partial\Omega \times (0,\infty), \\ u(x,0) = u_0(x), & \text{in } \Omega, \end{cases} \tag{1}$$

where $\partial_{0,t}^{\alpha}$ denotes the Caputo fractional derivative, $\alpha \in ((1/2), (3/2) - H)$, and $\Omega \subset R^n$ is a bounded open domain, whose boundary $\partial\Omega$ is sufficiently smooth. Functions f and D satisfy some appropriate conditions, with the initial data u_0 for u, $(\partial B_Q^H/\partial t)$ is the fractional Brownian motion (FBM), and $H \in ((1/2), 1)$. In addition,

$$A\,u(x,t) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\sum_{j=1}^{n} a_{ij}(x)\frac{\partial}{\partial x_j}u(x,t)\right) + b(x)u(x,t), \tag{2}$$

and the functions a_{ij} satisfy

$$\begin{gathered} a_{ij} \in L^{\infty}(\mathbb{R}^n), \quad 1 \le i,\ j \le n, \\ C_0 \sum_{i=1}^{n} \xi_i^2 \le \sum_{i,j=1}^{n} a_{ij}(x)\xi_i\xi_j, \quad \text{a.e. } x \in \mathbb{R}^n,\ \xi \in \mathbb{R}^n, \end{gathered} \tag{3}$$

where $C_0 > 0$ represents a constant, and $b(x)$ satisfies

$$\begin{gathered} b \in L^{\infty}(\mathbb{R}^n), \\ b(x) \ge B_0 > 0, \text{ a.e. } x \end{gathered} \tag{4}$$

Firstly, a new generalized Gronwall inequality is given. Then, we obtain the existence and uniqueness of the S-asymptotically periodic solutions for problem (1) based on the characteristics of Mittag–Leffler functions, Hölder inequality, and the inequality for fractional stochastic integral with FBM and Banach's fixed-point theorem. In addition, with the help of the generalized Gronwall inequality, some conditions are given to ensure the asymptotic stability of the S-asymptotically periodic solutions for problem (1). We notice that the generalized Gronwall inequality in [19] cannot be applied to Theorem 2 because the estimation obtained by that method is not stable.

Compared with previous research results, the innovations of this paper include the following: (1) the equations studied contain both fractional differential operators and FBM. It is worth mentioning that the standard Brownian motion, without long memory, cannot represent all types of noise. A good long-term memory noise could be described by FBM of Hurst parameter $H \in ((1/2), 1)$ [20]. For instance, the continuous disturbance and long-term dependence in the financial market model can be considered as a kind of FBM [21], with the impact of nuclear waste on the environment being seen as FBM in ecological models. Other research studies on FBM can be referred to [22–30]. (2) Stability of the S-asymptotically periodic solutions is studied by means of a new generalized Gronwall inequality.

The paper is organized as follows: the readers are allowed to review Section 2 for the necessary basic knowledge, followed by some results of the existence and uniqueness of S-asymptotically periodic solutions in Section 3. Subsequently, the asymptotic stability of S-asymptotically periodic solutions is studied in Section 4, with a numerical simulation example in Section 5.

2. Preliminaries

For the sake of convenience in writing, throughout this paper, by ∞, we mean $+\infty$. $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, P)$ denotes a complete filtered probability space, and U and $\mathbb{H}$ are two separable Hilbert spaces. The space of bounded linear operators from U into $\mathbb{H}$ is written as $L(\mathrm{U}, \mathbb{H})$. For convenience, the same notation $\|\cdot\|$ is applied to denote norms in $\mathrm{U}, \mathbb{H}$ and $L(\mathrm{U}, \mathbb{H})$; $(\cdot,\cdot)$ is applied to denote the inner produce of U and $\mathbb{H}$. Moreover, $L^2(\Omega; \mathbb{H})$ is the space of all strongly measurable and square-integrable $\mathbb{H}$-valued random variables under the Banach norm $(E\|\cdot\|^2)^{1/2}$.

A stochastic process u: $[0,\infty) \longrightarrow L^2(\Omega; \mathbb{H})$ is called stochastically bounded if $\sup_{t\ge 0} E\|u(t)\|^2 < +\infty$, called stochastically continuous if $\lim_{t \longrightarrow s} E\|u(t) - u(s)\|^2 = 0$ for all $t, s \ge 0$ and called square-mean S-asymptotically ω-periodic if $\lim_{t\longrightarrow\infty} E\|u(t+\omega) - u(t)\|^2 = 0$, where $\omega > 0$ is a constant.

Denote by $\mathrm{SBC}([0,\infty); L^2(\Omega;\mathbb{H}))$ the space of all stochastically bounded and continuous processes from $[0,\infty)$ into $L^2(\Omega;\mathbb{H})$ and its norm

$$\|u\|_\infty = \left(\sup_{t\geq 0} E\|u(t)\|^2\right)^{1/2}, \tag{5}$$

where $E\|u(t)\|^2 = \int_\Omega \|u(t)\|^2 \mathrm{d}P$. Then, $\mathrm{SBC}([0,\infty); L^2(\Omega;\mathbb{H}))$ is a Banach space.

We use $\mathrm{SAP}_w([0,\infty); L^2(\Omega;\mathbb{H}))$ to denote the space of square-mean S-asymptotically ω-periodic stochastic process from $[0,\infty)$ into $L^2(\Omega;\mathbb{H})$. Then, $\mathrm{SAP}_w([0,\infty); L^2(\Omega;\mathbb{H}))$ is a Banach space with the sup norm $\|\cdot\|_\infty$ and is a linear closed subspace of $\mathrm{SBC}([0,\infty); L^2(\Omega;\mathbb{H}))$.

In the following, we introduce the definition and properties of FBM. We denote by $\{\beta^H(t)\}_{t\in\mathbb{R}}$ $(H\in(0,1))$ a two-sided one-dimensional FBM [23]. Then, β^H is a continuous-centered Gaussian process, whose variance function is

$$\begin{aligned} R_H(t,s) &= E\left[\beta^H(t)\beta^H(s)\right] \\ &= \frac{1}{2}\left(t^{2H}+s^{2H}-|t-s|^{2H}\right), \quad t,\, s\in\mathbb{R}. \end{aligned} \tag{6}$$

In addition, if W is a Wiener process, then

$$\begin{aligned} \beta^H(t) &= \int_0^t K_H(t,s)\mathrm{d}W(s), \\ K_H(t,s) &= C_H s^{(1/2)-H}\int_s^t (\tau-s)^{H-(3/2)}\tau^{H-(1/2)}\mathrm{d}\tau, \end{aligned} \tag{7}$$

for $t>s$, where $C_H = \sqrt{H(2H-1)/B(2-2H, H-(1/2))}$, with B denoting the beta function.

Let $Q\in L(\mathbb{H},\mathbb{H})$ be an operator with $T_r(Q) = \sum_{n=1}^\infty \lambda_n < \infty$ and $Qe_n = \lambda_n e_n$ for constants $\lambda_n \geq 0\,(n=1,2,\ldots)$ and a complete orthonormal basis $\{e_n\}_{n=1}^\infty$ in $\mathbb{H}$. The infinite dimensional FBM on $\mathbb{H}$ can be expressed by

$$B_Q^H(t) = \sum_{n=1}^\infty \beta_n^H(t)Q^{1/2}e_n = \sum_{n=1}^\infty \sqrt{\lambda_n}\beta_n^H(t)e_n, \quad t\geq 0, \tag{8}$$

where Q is the covariance operator and $\{\beta_n^H(t)\}_{n=1}^\infty$ are two-sided one-dimensional FBMs, which are mutually independent on $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, P)$.

Let $L_2^0(\mathrm{U},\mathbb{H})$ be the collection of all Q-Hilbert–Schmidt operators ξ: $\mathrm{U}\longrightarrow\mathbb{H}$, where $\xi Q^{1/2}$ is a Hilbert–Schmidt operator, and the norm is

$$\|\xi\|_{L_2^0}^2 := \mathrm{tr}\left(\xi Q\xi^*\right) = \sum_{n=1}^\infty \left\|\sqrt{\lambda_n}\,\xi e_n\right\|^2 < \infty. \tag{9}$$

For convenience, set $L_2^0(\mathbb{H}) := L_2^0(\mathbb{H},\mathbb{H})$. The space $L_2^0(\mathrm{U},\mathbb{H})$ is a separable Hilbert space whose inner product $\langle\varphi,\phi\rangle_{L_2^0} = \sum_{n=1}^\infty \langle\varphi e_n, \phi e_n\rangle$. Then, we define the stochastic integral of ϕ with regard to B_Q^H by

$$\begin{aligned} \int_0^t \phi(s)\mathrm{d}B_Q^H(s) &= \sum_{n=1}^\infty \int_0^t \sqrt{\lambda_n}\phi(s)e_n\mathrm{d}\beta_n^H(s) \\ &= \sum_{n=1}^\infty \int_0^t \phi(s)Q^{1/2}e_n\mathrm{d}\beta_n^H(s), \end{aligned} \tag{10}$$

where $\{\phi(t)\}_{t\in[0,T]}$ is the deterministic function with values in $L_2^0(\mathrm{U},\mathbb{H})$.

Now, we recall the Mittag–Leffler function and the probability density functions which play important roles in fractional differential equations [31].

Lemma 1. *The Mittag–Leffler functions*

$$\begin{aligned} E_\alpha(t) &= \sum_{k=0}^\infty \frac{t^k}{\Gamma(\alpha k+1)}, \\ E_{\alpha,\alpha}(t) &= \sum_{k=0}^\infty \frac{t^k}{\Gamma(\alpha(k+1))}, t\in\mathbb{R}, \end{aligned} \tag{11}$$

where Γ is the Gamma function, have the following properties:

(1) $E_\alpha(t), E_{\alpha,\alpha}(t) > 0$ [32, 33].

(2) $(E_\alpha(t))' = (1/\alpha)E_{\alpha,\alpha}(t)$ [34].

(3) $\lim_{t\longrightarrow-\infty}E_\alpha(t) = \lim_{t\longrightarrow-\infty}E_{\alpha,\alpha}(t) = 0$ [32, 35].

We notice that the Mittag–Leffler function $E_\alpha(t)$ is a generalization of exponential function e^t, which is $E_1(t)$.

Set $u(t)(x) = u(x,t)$, $f(t,u(t))(x) = f(t,x,u(x,t))$, and $(B_Q^H u)(t)(x) = B_Q^H u(x,t)$. Let $\mathbb{H} = L^2(\Omega)$, and we define A: $D(A)\subset\mathbb{H}\longrightarrow\mathbb{H}$ by

$$\begin{aligned} D(A) &= H^2(\Omega)\cap H_0^1(\Omega), \\ (Au)(t)x &= \mathscr{A}u(x,t). \end{aligned} \tag{12}$$

We suppose that $-A$ generates an exponentially stable C_0-semigroup $\{T(t)\}_{t\geq 0}$ satisfying

$$\|T(t)\| \leq Me^{-\delta t}, \quad \forall t\in[0,\infty), \tag{13}$$

where $M>0$ and $\delta>0$ are constants.

Thus, (1) can be transformed into

$$\begin{cases} D_0^\alpha u(t) + Au(t) = f(t,u(t)) + D(t,u(t))\dfrac{\mathrm{d}B_Q^H(t)}{\mathrm{d}t}, & t\in(0,\infty), \\ u(0) = u_0, \end{cases} \tag{14}$$

where D_0^α is the Caputo-fractional derivative. Later, in the paper, $u_0\in L^2(\Omega;\mathbb{H})$.

Remark 1. We see that (13) is much easy to be verified. For example, let

$$\begin{aligned}
&\Omega = [0,\pi],\\
&\mathbb{H} = L^2[0,\pi],\\
&A = -\frac{\partial^2}{\partial x^2},\\
&D(A) = \{u \in \mathbb{H} \mid u, u' \text{ are absolutely continuous, } u'' \in \mathbb{H},\ u(0) = u(\pi) = 0\}.
\end{aligned} \tag{15}$$

Then, A has eigenvalues n^2 $(n \in \mathbb{N})$, whose normalized eigenvectors $w_n(t) = \sqrt{2/\pi}\sin(nt)$ $(n \in \mathbb{N})$, $-A$ generates an analytic, compact, and exponentially stable semigroup $\{T(t)\}_{t\geq 0}$, and

$$\begin{aligned}
&T(t)u = \sum_{n=1}^{\infty} e^{-n^2 t}(u, w_n)w_n,\\
&\|T(t)\| \leq e^{-t},\ t \in [0,\infty).
\end{aligned} \tag{16}$$

Definition 1. A stochastic continuous process u: $[0,\infty) \longrightarrow L^2(\Omega;\mathbb{H})$ is said to be a mild solution of equation (14) if

$$\begin{aligned}
u(t) = U(t)u_0 &+ \int_0^t (t-s)^{\alpha-1}V(t-s)f(s,u(s))\mathrm{d}s\\
&+ \int_0^t (t-s)^{\alpha-1}V(t-s)D(s,u(s))\mathrm{d}B_Q^H(s),
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
U(t) &= \int_0^{\infty} \zeta_\alpha(\theta)T(t^\alpha\theta)\mathrm{d}\theta,\\
V(t) &= \alpha\int_0^{\infty} \theta\zeta_\alpha(\theta)T(t^\alpha\theta)\mathrm{d}\theta,
\end{aligned} \tag{18}$$

and the probability density function [36, 37]

$$\begin{aligned}
&\zeta_\alpha(\theta) = \frac{1}{\alpha}\theta^{-1-(1/\alpha)}\rho_\alpha\left(\theta^{-(1/\alpha)}\right),\\
&\rho_\alpha(\theta) = \frac{1}{\pi}\sum_{n=0}^{\infty}(-1)^{n-1}\theta^{-\alpha n-1}\frac{\Gamma(n\alpha+1)}{n!}\sin(n\pi\alpha), \quad \theta \in (0,\infty).
\end{aligned} \tag{19}$$

Later, in this paper, we need the following results.

Remark 2

(1) $\zeta_\alpha(\theta) \geq 0$, for $\theta \in (0,\infty)$.
(2) $\int_0^\infty \theta^\nu\zeta_\alpha(\theta)\mathrm{d}\theta = (\Gamma(1+\nu)/\Gamma(1+\alpha\nu))$, for $\nu \in (-1,+\infty)$ [36, 38].
(3) $\int_0^\infty e^{-t\theta}\theta\zeta_\alpha(\theta)\mathrm{d}\theta = (1/\alpha)E_{\alpha,\alpha}(-t)$, for $t \in \mathbb{R}$ [37].
(4) $\int_0^\infty e^{-t\theta}\theta\zeta_\alpha(\theta)\mathrm{d}\theta = (1/\alpha)E_{\alpha,\alpha}(-t)$, for $t \in \mathbb{R}$ [36].

Lemma 2

(1) $U(t)$ and $V(t)$ are strongly continuous for $t \geq 0$ [38].
(2) If $\{T(t)\}_{t\geq 0}$ satisfy (14), then $\|U(t)\| \leq ME_\alpha(-\delta t^\alpha)$ and $\|V(t)\| \leq ME_{\alpha,\alpha}(-\delta t^\alpha)$ for $t \geq 0$ [2].
(3) $\|V(t)\| \leq (M\alpha/\delta t^\alpha)$ for $t > 0$.

Proof. The proof of (3) is as follows. In fact, for $t > 0$, in view of $e^{-\nu} < (1/\nu)$ with $\nu > 0$, we have

$$\begin{aligned}
V(t)\| &\leq \alpha\left\|\int_0^\infty \theta\zeta_\alpha(\theta)T(t^\alpha\theta)\mathrm{d}\theta\right\|\\
&\leq M\alpha\int_0^\infty \theta\zeta_\alpha(\theta)e^{-\delta t^\alpha\theta}\mathrm{d}\theta\\
&\leq M\alpha\int_0^\infty \frac{\theta\zeta_\alpha(\theta)}{\delta t^\alpha\theta}\mathrm{d}\theta\\
&\leq \frac{M\alpha}{\delta t^\alpha}.
\end{aligned} \tag{20}$$

In order to get the stability of the square-mean S-asymptotically ω-periodic solution, we need the following generalized Gronwall inequality for fractional differential equations. □

Lemma 3. *Let u_0, λ_1, $\lambda_2 \in \mathbb{R}$ be two constants. If a continuous function u: $[0,+\infty) \longrightarrow \mathbb{R}$ satisfies*

$$u(t) \leq E_\alpha(\lambda_1 t^\alpha)u_0 + \lambda_2\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(\lambda_1(t-s)^\alpha)u(s)\mathrm{d}s, \tag{21}$$

then

$$u(t) \leq E_\alpha((\lambda_1+\lambda_2)t^\alpha)u_0. \tag{22}$$

Proof. We find that the solution of the equation [31]

$$\begin{cases} D_0^\alpha u(t) = \lambda_1 u(t) + \lambda_2 u(t), & t \in (0,+\infty),\\ u(0) = u_0, \end{cases} \tag{23}$$

is given by

$$\begin{aligned}
&u(t) = E_\alpha(\lambda_1 t^\alpha)u_0 + \lambda_2\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(\lambda_1(t-s))^\alpha u(s)\mathrm{d}s,\\
&\text{or}\\
&u(t) = E_\alpha((\lambda_1+\lambda_2)t^\alpha)u_0.
\end{aligned} \tag{24}$$

In view of the uniqueness of solution to (23), we get (22). □

Remark 3. Compared with the generalized Gronwall inequality in [19], $E_\alpha(\lambda_1 t^\alpha)u_0$ does not have to be a nondecreasing function, and λ_2 is not necessarily nonnegative. This is very important to prove the stability of the solution.

Next, we give a result which is very useful for the estimations of fractional stochastic integral with FBM.

Lemma 4 (see [39]). *Let $H \in (1/2, 1)$ and $h\colon [0,T] \longrightarrow L_2^0(\mathbb{U},\mathbb{V})$ satisfy*

$$\int_a^b \|\phi(s)\|_{L_2^0}^2 \mathrm{d}s < \infty, \quad \text{for } \forall a, b \in [0,T] \text{ with } b > a, \tag{25}$$

then the corresponding sum given in (10) is well defined and we obtain

$$E\left\|\int_a^b h(s)\mathrm{d}B_Q^H(s)\right\|^2 \le 2H(b-a)^{2H-1}\int_a^b h(s)_{L_2^0}^2 \mathrm{d}s. \tag{26}$$

3. Existence Uniqueness of Square-Mean S-Asymptotically ω-Periodic Solutions

Lemma 5. *Let $u \in SAP_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$,*

$$(\Gamma_1 u)(t) = U(t)u_0 + \int_0^t (t-s)^{\alpha-1} V(t-s) f(s, u(s))\mathrm{d}s, \tag{27}$$

and if $f\colon [0,\infty) \times L^2(\Omega;\mathbb{H}) \longrightarrow L^2(\Omega;\mathbb{H})$ satisfies.

(H_1) For x in every bounded subset K of $L^2(\Omega;\mathbb{H})$, $f(t,x)$ is S-asymptotically ω-periodic in t. Moreover, there exists a positive constant L such that

$$E\|f(t,y) - f(t,z)\|^2 \le LE\|y-z\|^2, \tag{28}$$

for $t \in [0,\infty)$ and $y, z \in L^2(\Omega;\mathbb{H})$.

Then, $\Gamma_1 u \in SAP_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$.

Proof

Step 1: for $\forall u(t) \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$, we prove that

$$\lim_{t\longrightarrow\infty} E\left\|(\Gamma_1 u)(t+\omega) - (\Gamma_1 u)(t)\right\|^2 = 0. \tag{29}$$

Firstly, we have

$$\begin{aligned}
&(\Gamma_1 u)(t+\omega) - (\Gamma_1 u)(t) \\
&\quad = (U(t+\omega) - U(t))u_0 + \int_{-\omega}^0 (t-s)^{\alpha-1} V(t-s) f(s+\omega, u(s+\omega))\mathrm{d}s \\
&\qquad + \int_0^t (t-s)^{\alpha-1} V(t-s)(f(s+\omega, u(s+\omega)) - f(s, u(s+\omega)))\mathrm{d}s \\
&\qquad + \int_0^t (t-s)^{\alpha-1} V(t-s)(f(s, u(s+\omega)) - f(s, u(s)))\mathrm{d}s.
\end{aligned} \tag{30}$$

Then,

$$\begin{aligned}
&E\left\|(\Gamma_1 u)(t+\omega) - (\Gamma_1 u)(t)\right\|^2 \\
&\quad \le 4E\left\|[U(t+\omega) - U(t)]u_0\right\|^2 \\
&\qquad + 4E\left\|\int_{-\omega}^0 (t-s)^{\alpha-1} V(t-s) f(s+\omega, u(s+\omega))\mathrm{d}s\right\|^2 \\
&\qquad + 4E\left\|\int_0^t (t-s)^{\alpha-1} V(t-s)(f(s+\omega, u(s+\omega)) - f(s, u(s+\omega)))\mathrm{d}s\right\|^2 \\
&\qquad + 4E\left\|\int_0^t (t-s)^{\alpha-1} V(t-s)(f(s, u(s+\omega)) - f(s, u(s)))\mathrm{d}s\right\|^2 \\
&\quad =: I_1(t) + I_2(t) + I_3(t) + I_4(t).
\end{aligned} \tag{31}$$

In view of Lemma 2, it is obvious that

$$\lim_{t\longrightarrow\infty} I_1(t) = 0. \tag{32}$$

By combining Hölder inequality with (H_1) and using Lemma 2, we have

$$\begin{aligned}
I_2(t) &\le 4\int_{-\omega}^{0}(t-s)^{\alpha-1}\|V(t-s)\|\mathrm{d}s\int_{-\omega}^{0}(t-s)^{\alpha-1}\|V(t-s)\|E\|f(s+\omega,u(s+\omega))\|^2\mathrm{d}s\\
&\le 8\left(LE\|u\|_{\infty}^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)\left(\int_{-\omega}^{0}(t-s)^{\alpha-1}\|V(t-s)\|\mathrm{d}s\right)^2\\
&\le 8M^2\left(LE\|u\|_{\infty}^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)\left(\int_{-\omega}^{0}(t-s)^{\alpha-1}E_{\alpha,\alpha}\left(-\delta(t-s)^{\alpha}\right)\mathrm{d}s\right)^2\\
&\le \frac{8M^2}{\delta^2}\left(LE\|u\|_{\infty}^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)\left(E_{\alpha}\left(-\delta(t-s)^{\alpha}\right)\Big|_{-\omega}^{0}\right)^2\\
&= \frac{8M^2}{\delta^2}\left(LE\|u\|_{\infty}^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)\left(E_{\alpha}\left(-\delta t^{\alpha}\right)-E_{\alpha}\left(-\delta(t+\omega)^{\alpha}\right)\right)^2.
\end{aligned} \tag{33}$$

The last formula and Lemma 1 yield that

$$\lim_{t\longrightarrow\infty} I_2(t) = 0. \tag{34}$$

Since (H_1) implies $\lim_{t\longrightarrow+\infty}E\|f(t+\omega,u(t+\omega))-f(t,u(t+\omega))\|^2=0$, then for $\varepsilon>0$, there exists $T_\varepsilon>0$ such that $E\|f(t+\omega,u(t+\omega))-f(t,u(t+\omega))\|^2\le\varepsilon$ whenever $u>T_\varepsilon$. Then, we have

$$\begin{aligned}
I_3(4) &\le 4\int_0^t(t-s)^{\alpha-1}\|V(t-s)\|\mathrm{d}s\int_0^t(t-s)^{\alpha-1}\|V(t-s)\|E\|f(s+\omega,u(s+\omega))-f(s,u(s+\omega))\|^2\mathrm{d}s\\
&\le 4M^2\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}\left(-\delta(t-s)^{\alpha}\right)\mathrm{d}s\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}\left(-\delta(t-s)^{\alpha}\right)\cdot E\|f(s+\omega,u(s+\omega))-f(s,u(s+\omega))\|^2\mathrm{d}s\\
&\le \frac{4M^2}{\delta}\left(1-E_{\alpha}\left(-\delta t^{\alpha}\right)\right)\left(\int_0^{T_\varepsilon}(t-s)^{\alpha-1}E_{\alpha,\alpha}\left(-\delta(t-s)^{\alpha}\right)\cdot E\|f(s+\omega,u(s+\omega))-f(s,u(s+\omega))\|^2\mathrm{d}s\right.\\
&\quad\left.+\int_{T_\varepsilon}^t(t-s)^{\alpha-1}E_{\alpha,\alpha}\left(-\delta(t-s)^{\alpha}\right)E\|f(s+\omega,u(s+\omega))-f(s,u(s+\omega))\|^2\mathrm{d}s\right)\\
&\le \frac{4M^2}{\delta^2}\left(1-E_{\alpha}\left(-\delta t^{\alpha}\right)\right)\left(4\left(LE\|u\|_{\infty}^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)\left(E_{\alpha}\left(-\delta(t-T_\varepsilon)^{\alpha}\right)-E_{\alpha}\left(-\delta t^{\alpha}\right)\right)+\varepsilon\left(1-E_{\alpha}\left(-\delta(t-T_\varepsilon)^{\alpha}\right)\right)\right).
\end{aligned} \tag{35}$$

Due to Lemma 1, it is obvious that

$$\lim_{t\longrightarrow\infty} I_3(t) = 0. \tag{36}$$

(H_1) implies that

$$I_4(4) \le 4L\int_0^t(t-s)^{\alpha-1}\|V(t-s)\|\mathrm{d}s\int_0^t(t-s)^{\alpha-1}\|V(t-s)\|E\|u(s+\omega)-u(s)\|^2\mathrm{d}s. \tag{37}$$

Using a strategy similar to the one in the proof of (36), we get

$$\lim_{t\longrightarrow\infty} I_4(t) = 0. \tag{38}$$

Then, $\lim_{t\longrightarrow\infty} E\|(\Gamma_1 u)(t+\omega) - (\Gamma_1 u)(t)\|^2 = 0$.

Step 2: For $\forall u(t) \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$, we prove that $(\Gamma_1 u)(t)$ is stochastically bounded and continuous.

On the one hand, for a given $t_0 \geq 0$, we have

$$\begin{aligned} E\|(\Gamma_1 u)(t) - (\Gamma_1 u)(t_0)\|^2 &\leq 3E\|[U(t) - U(t_0)]u_0\|^2 \\ &\quad + 3E\left\|\int_{t_0-t}^{0} (t_0-s)^{\alpha-1} V(t_0-s) f(s+t-t_0, u(s+t-t_0))\mathrm{d}s\right\|^2 \\ &\quad + 3E\left\|\int_{0}^{t_0} (t_0-s)^{\alpha-1} V(t_0-s)\left(f(s+t-t_0, u(s+t-t_0)) - f(s,u(s))\right)\mathrm{d}s\right\|^2 \\ &=: K_1(t) + K_2(t) + K_3(t). \end{aligned} \tag{39}$$

Lemma 2 (1) implies that $\lim_{t\longrightarrow t_0} K_1(t) = 0$. Arguing similarly as in (33), we see that

$$\begin{aligned} K_2(t) &\leq \frac{6M^2}{\delta^2}\left(LE\|u\|_\infty^2 + \sup_{t\geq 0} E\|f(t,0)\|^2\right) \\ &\quad \cdot \left(E_\alpha(-\delta t_0^\alpha) - E_\alpha(-\delta t^\alpha)\right)^2, \end{aligned} \tag{40}$$

which means that $\lim_{t\longrightarrow t_0} K_2(t) = 0$.

$$\begin{aligned} K_3(t) &\leq 6M^2 \int_0^{t_0} (t_0-s)^{\alpha-1} E_{\alpha,\alpha}(-\delta(t_0-s)^\alpha)\mathrm{d}s \int_0^{t_0} (t_0-s)^{\alpha-1} E_{\alpha,\alpha}(-\delta(t_0-s)^\alpha) \\ &\quad \cdot \left(E\|f(s+t-t_0, u(s+t-t_0)) - f(s,u(s+t-t_0))\|^2 + E\|f(s,u(s+t-t_0)) - f(s,u(s))\|^2\right)\mathrm{d}s \\ &\leq \frac{6M^2}{\delta}\left(1 - E_\alpha(-\delta t_0^\alpha)\right) \int_0^{t_0} (t_0-s)^{\alpha-1} E_{\alpha,\alpha}(-\delta(t_0-s)^\alpha)\Big(E\|f(s+t-t_0, u(s+t-t_0)) - f(s,u(s+t-t_0))\|^2 \\ &\quad + LE\|u(s+t-t_0) - u(s)\|^2\Big)\mathrm{d}s. \end{aligned} \tag{41}$$

For an arbitrary sequence of real numbers $\{t_n\}$ with $t_n \longrightarrow t_0$ as $n \longrightarrow \infty$, for $\forall u(t) \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$, we have

$$\begin{aligned} &E\|f(s+t_n-t_0, u(s+t_n-t_0)) - f(s,u(s+t_n-t_0))\|^2 \\ &\quad + LE\|u(s+t_n-t_0) - u(s)\|^2 \longrightarrow 0, \quad \text{as } n \longrightarrow \infty, \end{aligned} \tag{42}$$

which is due to (H_1). Hence,

$$\begin{aligned} &E\|f(s+t_n-t_0, u(s+t_n-t_0)) - f(s,u(s+t_n-t_0))\|^2 \\ &\quad + LE\|u(s+t_n-t_0) - u(s)\|^2 < 1, \end{aligned} \tag{43}$$

for every n sufficiently large. In view of

$$\int_0^{t_0} (t_0-s)^{\alpha-1} E_{\alpha,\alpha}(-\delta(t_0-s)^\alpha)\mathrm{d}s = \frac{1}{\delta}\left(1 - E_\alpha(-\delta t_0^\alpha)\right) < \frac{1}{\delta} < \infty, \tag{44}$$

then it follows from Lebesgue's dominated convergence theorem that

$$\lim_{n\longrightarrow\infty}\int_0^{t_0}(t_0-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t_0-s)^{\alpha})\Big(E\|f(s+t_n-t_0,u(s+t_n-t_0))-f(s,u(s+t_n-t_0))\|^2+LE\|u(s+t_n-t_0)-u(s)\|^2\Big)\mathrm{d}s=0. \tag{45}$$

Additionally, according to the arbitrariness of t_n, we have that

$$\lim_{t\longrightarrow t_0}\int_0^{t_0}(t_0-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t_0-s)^{\alpha})\Big(E\|f(s+t-t_0,u(s+t-t_0))-f(s,u(s+t-t_0))\|^2+LE\|u(s+t-t_0)-u(s)\|^2\Big)\mathrm{d}s=0, \tag{46}$$

which gives $\lim_{t\longrightarrow t_0}K_3(t)=0$. Then, we know that $(\Gamma_1u)(t)$ is stochastically continuous.

On the other hand,

$$\begin{aligned}E\|(\Gamma_1u)(t)\|^2&\le 2E\|U(t)u_0\|^2+2E\left\|\int_0^t(t-s)^{\alpha-1}V(t-s)f(s,u(s))\mathrm{d}s\right\|^2\\&\le 2M^2E\|u_0\|^2+2M^2\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t-s)^{\alpha})\mathrm{d}s\cdot\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t-s)^{\alpha})E\|f(s,u(s))\|^2\mathrm{d}s\\&\le 2M^2E\|u_0\|^2+\frac{4M^2}{\delta^2}(1-E_\alpha(-\delta t^\alpha))\left(LE\|u\|_\infty^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)\\&\le 2M^2E\|u_0\|^2+\frac{4M^2}{\delta^2}\left(LE\|u\|_\infty^2+\sup_{t\ge0}E\|f(t,0)\|^2\right)<\infty,\end{aligned} \tag{47}$$

which implies that $(\Gamma_1u)(t)$ is stochastically bounded.

By Steps 1-2, we obtain $\Gamma_1u\in\mathrm{SAP}_\omega([0,\infty),L^2(\Omega;\mathbb{H}))$. □

Lemma 6. *Let $u\in SAP_\omega([0,\infty),L^2(\Omega;\mathbb{H}))$ and*

$$(\Gamma_2u)(t)=\int_0^t(t-s)^{\alpha-1}V(t-s)D(s,u(s))\mathrm{d}B_Q^H(s). \tag{48}$$

If D: $[0,\infty)\times L^2(\Omega;\mathbb{H})\longrightarrow L_2^0(\mathbb{H})$ satisfies.

(H_2) For x in every bounded subset K of $L^2(\Omega;\mathbb{H})$, $D(t,x)$ is stochastically bounded and continuous in t. Furthermore, for $\forall\varepsilon>0$ and K, there exists $T_\varepsilon>0$ such that $t^{2(\alpha+H-1)}\|D(t+\omega,y)-D(t,y)\|_{L_2^0}^2<\varepsilon$ for $t>T_\varepsilon$ and $y\in K$.

(H_3) There exists a positive constant L_1 such that

$$t^{2(\alpha+H-1)}\|D(t,y)-D(t,z)\|_{L_2^0}^2\le L_1E\|y-z\|^2, \tag{49}$$

for $t\in[0,\infty)$ and $y,z\in L^2(\Omega;\mathbb{H})$.

(H_4) $D(t,0)=0$ for $t\in[0,\infty)$.

Then, $\Gamma_2u\in SAP_\omega([0,\infty),L^2(\Omega;\mathbb{H}))$.

Proof

Step 1: for $\forall u(t)\in\mathrm{SAP}_\omega([0,\infty),L^2(\Omega;\mathbb{H}))$, we prove that

$$\lim_{t\longrightarrow\infty}E\|(\Gamma_2u)(t+\omega)-(\Gamma_2u)(t)\|^2=0. \tag{50}$$

Let $\widetilde{B}_Q^H(\tau)=B_Q^H(\tau+\omega)-B_Q^H(\omega)$ for each $\tau\in R$. Then, it is easy to find that $\widetilde{B}_Q^H(\tau)$ is identically distributed like $B_Q^H(\tau)$. Next, for $u\in\mathrm{SAP}_\omega([0,\infty),L^2(\Omega;\mathbb{H}))$, we see that

$$\begin{aligned}(\Gamma_2 u)(t+\omega)-(\Gamma_2 u)(t) &= \int_0^{t+\omega}(t+\omega-s)^{\alpha-1}V(t+\omega-s)D(s,u(s))\mathrm{d}B_Q^H(s)\\ &\quad-\int_0^t(t-s)^{\alpha-1}V(t-s)D(s,u(s))\mathrm{d}B_Q^H(s)\\ &= \int_{-\omega}^t(t-\tau)^{\alpha-1}V(t-\tau)D(\tau+\omega,u(\tau+\omega))\mathrm{d}B_Q^H(\tau+\omega)\\ &\quad-\int_0^t(t-s)^{\alpha-1}V(t-s)D(s,u(s))\mathrm{d}B_Q^H(s)\\ &= \int_{-\omega}^t(t-\tau)^{\alpha-1}V(t-\tau)D(\tau+\omega,u(\tau+\omega))\mathrm{d}\widetilde{B}_Q^H(\tau)\\ &\quad-\int_0^t(t-s)^{\alpha-1}V(t-s)D(s,u(s))\mathrm{d}\widetilde{B}_Q^H(s).\end{aligned} \tag{51}$$

Then,

$$\begin{aligned}E\left\|(\Gamma_2 u)(t+\omega)-(\Gamma_2 u)(t)\right\|^2 &\le 2E\left\|\int_{-\omega}^0(t-s)^{\alpha-1}V(t-s)D(s+\omega,u(s+\omega))\mathrm{d}\widetilde{B}_Q^H(s)\right\|^2\\ &\quad+2E\left\|\int_0^t(t-s)^{\alpha-1}V(t-s)(D(s+\omega,u(s+\omega))-D(s,u(s)))\mathrm{d}\widetilde{B}_Q^H(s)\right\|^2\\ &=: 2I(t)+2J(t).\end{aligned} \tag{52}$$

Moreover, $(H_3)-(H_4)$, Lemmas 2 and 4 yield that

$$\begin{aligned}I(t) &\le 2Hw^{2H-1}\int_{-\omega}^0(t-s)^{2\alpha-2}\|V(t-s)\|^2\|D(s+\omega,u(s+\omega))\|_{L_2^0}^2\mathrm{d}s\\ &\le \frac{2Hw^{2H-1}M^2\alpha^2}{\delta^2}\int_{-\omega}^0(t-s)^{-2}\|D(s+\omega,u(s+\omega))\|_{L_2^0}^2\mathrm{d}s\\ &= \frac{2Hw^{2H-1}M^2\alpha^2}{\delta^2t^2}\int_0^\omega\|D(s,u(s))\|_{L_2^0}^2\mathrm{d}s\\ &\le \frac{4Hw^{2H-1}M^2\alpha^2}{\delta^2t^2}\int_0^\omega\left(\|D(s,0)\|_{L_2^0}^2+L_1s^{2(1-H-\alpha)}E\|u(s)\|^2\right)\mathrm{d}s\\ &\le \frac{4Hw^{2H-1}M^2\alpha^2}{\delta^2t^2}\cdot\left(\omega\sup_{t\ge0}\|D(t,0)\|_{L_2^0}^2+\frac{L_1\omega^{3-2H-2\alpha}}{3-2H-2\alpha}\|u\|_\infty^2\right).\end{aligned} \tag{53}$$

Then, and (H_2), (H_3), and Lemma 4 imply that

$$\lim_{t\longrightarrow\infty}I(t)=0, \tag{54}$$

$$\begin{aligned}J(t) &\le 4HT_\varepsilon^{2H-1}\int_0^{T_\varepsilon}(t-s)^{2\alpha-2}\|V(t-s)\|^2\|D(s+\omega,u(s+\omega))-D(s,u(s))\|_{L_2^0}^2\mathrm{d}s\\ &\quad+4H(t-T_\varepsilon)^{2H-1}\int_{T_\varepsilon}^t(t-s)^{2\alpha-2}\|V(t-s)\|^2\|D(s+\omega,u(s+\omega))-D(s,u(s))\|_{L_2^0}^2\mathrm{d}s\\ &= 4HT_\varepsilon^{2H-1}J_1(t)+4HJ_2(t),\end{aligned} \tag{55}$$

where $\lim_{t\longrightarrow\infty} J_1(t) = 0$ can be gotten similarly to (53). Moreover, without loss of generality, we may suppose for $\forall \varepsilon > 0$, $E\|u(t+\omega) - u(t)\|^2 < \varepsilon$, for $t > T_\varepsilon$, owing to $u \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$:

$$\begin{aligned}
J_2(t) &\le 2(t-T_\varepsilon)^{2H-1}\int_{T_\varepsilon}^{t}(t-s)^{2\alpha-2}\|V(t-s)\|^2\Big(\|D(s+\omega,u(s+\omega)) - D(s,u(s+\omega))\|_{L_2^0}^2 + \|D(s,u(s+\omega)) - D(s,u(s))\|_{L_2^0}^2\Big)\mathrm{d}s \\
&\le 2M^2(t-T_\varepsilon)^{2H-1}\int_{T_\varepsilon}^{t}(t-s)^{2\alpha-2}\Big(E_{\alpha,\alpha}(-\delta(t-s)^\alpha)\Big)^2\Big(\|D(s+\omega,u(s+\omega)) - D(s,u(s+\omega))\|_{L_2^0}^2 \\
&\quad + \|D(s,u(s+\omega)) - D(s,u(s))\|_{L_2^0}^2\Big)\mathrm{d}s \\
&= \frac{2M^2\varepsilon(1+L_1)}{(\Gamma(\alpha))^2}(t-T_\varepsilon)^{2H-1}\int_0^{t-T_\varepsilon}\tau^{2\alpha-2}(t-\tau)^{2(1-\alpha-H)}\mathrm{d}\tau \\
&= \frac{2M^2\varepsilon(1+L_1)}{(\Gamma(\alpha))^2}\int_0^1\mu^{2\alpha-2}(1-\mu)^{2(1-\alpha-H)}\mathrm{d}\mu \\
&= \frac{2M^2(1+L_1)B(2\alpha-1,3-2(\alpha+H))}{(\Gamma(\alpha))^2}\varepsilon,
\end{aligned} \tag{56}$$

where B is the beta function. Thus, $\lim_{t\longrightarrow\infty} J_2(t) = 0$. Therefore, $\lim_{t\longrightarrow\infty} E\|(\Gamma_2 u)(t+\omega) - (\Gamma_2 u)(t)\|^2 = 0$.

Step 2: For $\forall u \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$, we prove that $(\Gamma_2 u)(t)$ is stochastically bounded and continuous.

For a given number $t_0 \ge 0$ and $t > t_0$, we get

$$\begin{aligned}
E\|(\Gamma_2 u)(t) - (\Gamma_2 u)(t_0)\|^2 &\le 2E\left\|\int_{t_0-t}^{0}(t_0-s)^{\alpha-1}V(t_0-s)D(s+t-t_0,u(s+t-t_0))\mathrm{d}\widetilde{B}_Q^H(s)\right\|^2 \\
&\quad + 2E\left\|\int_0^{t_0}(t_0-s)^{\alpha-1}V(t_0-s)\big(D(s+t-t_0,u(s+t-t_0)) - D(s,u(s))\big)\mathrm{d}\widetilde{B}_Q^H(s)\right\| \\
&:= 2N_1(t) + 2N_2(t).
\end{aligned} \tag{57}$$

It follows from (H_3) and Lemmas 2 and 4 that

$$\begin{aligned}
N_1(t) &\le \frac{2HM^2}{(\Gamma(\alpha)^2)}(t-t_0)^{2H-1}\int_{t_0-t}^{0}(t_0-s)^{2\alpha-2}\|D(s+t-t_0,u(s+t-t_0))\|_{L_2^0}^2\mathrm{d}s \\
&\le \frac{2HM^2}{(\Gamma(\alpha)^2)}(t-t_0)^{2H-1}\int_0^{t-t_0}(t-s)^{2\alpha-2}\|D(s,u(s))\|_{L_2^0}^2\mathrm{d}s \\
&\le \frac{4HM^2}{(\Gamma(\alpha)^2)}(t-t_0)^{2H-1}\int_0^{t-t_0}(t-s)^{2\alpha-2}\mathrm{d}s\sup_{t\ge0}\|D(t,0)\|_{L_2^0}^2 \\
&\quad + \frac{4HM^2L_1}{(\Gamma(\alpha)^2)}(t-t_0)^{2H-1}\int_0^{t-t_0}(t-s)^{2\alpha-2}s^{2(1-H-\alpha)}\mathrm{d}s\|u\|_\infty^2 \\
&\le \frac{4HM^2\sup_{t\ge0}\|D(t,0)\|_{L_2^0}^2}{(\Gamma(\alpha)^2)(2\alpha-1)}(t-t_0)^{2H-1}\left(t^{2\alpha-1} - t_0^{2\alpha-1}\right) \\
&\quad + \frac{4HM^2L_1\|u\|_\infty^2 t_0^{2\alpha-2}}{(\Gamma(\alpha)^2(3-2H-2\alpha))}(t-t_0)^{2-2\alpha}.
\end{aligned} \tag{58}$$

This means $\lim_{t\longrightarrow t_0} N_1(t) = 0$.

Next, we find that

$$\begin{aligned} N_2(t) &\le 2HM^2 t_0^{2H-1} \int_0^{t_0} (t_0 - s)^{2\alpha-2} \left(E_{\alpha,\alpha}\left(-\delta(t_0 - s)^{\alpha}\right)\right)^2 \left\| D(s + t - t_0, u(s + t - t_0)) - D(s, u(s)) \right\|_{L_2^0}^2 \mathrm{d}s \\ &\le 4HM^2 t_0^{2H-1} \int_0^{t_0} (t_0 - s)^{2\alpha-2} \left(E_{\alpha,\alpha}\left(-\delta(t_0 - s)^{\alpha}\right)\right)^2 \Big(\left\| D(s + t - t_0, u(s + t - t_0)) - D(s, u(s + t - t_0)) \right\|_{L_2^0}^2 \\ &\quad + L_1 s^{2(1-\alpha-H)} \left\| u(s + t - t_0) - u(s) \right\|^2 \Big) \mathrm{d}s. \end{aligned} \tag{59}$$

Note that $N_2(t) = 0$ for $t_0 = 0$, and for $t_0 > 0$, we have

$$\begin{aligned} \int_0^{t_0} (t_0 - s)^{2\alpha-2} \left(E_{\alpha,\alpha}\left(-\delta(t_0 - s)^{\alpha}\right)\right)^2 \mathrm{d}s &= \int_0^{t_0} \tau^{2\alpha-2} \left(E_{\alpha,\alpha}\left(-\delta\tau^{\alpha}\right)\right)^2 \mathrm{d}\tau \\ &\le \int_0^{t_0/2} \tau^{2\alpha-2} \left(E_{\alpha,\alpha}\left(-\delta\tau^{\alpha}\right)\right)^2 \mathrm{d}\tau + \int_{t_0/2}^{t_0} \tau^{2\alpha-2} \left(E_{\alpha,\alpha}\left(-\delta\tau^{\alpha}\right)\right)^2 \mathrm{d}\tau \\ &\le \frac{1}{(\Gamma(\alpha))^2} \int_0^{t_0/2} s^{2\alpha-2} \mathrm{d}s + \left(\frac{t_0}{2}\right)^{2\alpha-2} \int_{t_0/2}^{t_0} \frac{1}{(\delta\tau^{\alpha})^2} \mathrm{d}\tau \\ &\le \frac{1}{(\Gamma(\alpha))^2} \cdot \frac{1}{2\alpha - 1} \left(\frac{t_0}{2}\right)^{2\alpha-1} + \frac{1}{\delta^2} \left(\frac{t_0}{2}\right)^{2\alpha-2} \int_{t_0/2}^{t_0} \tau^{-2\alpha} \mathrm{d}\tau \\ &= \frac{1}{(\Gamma(\alpha))^2} \cdot \frac{1}{2\alpha - 1} \left(\frac{t_0}{2}\right)^{2\alpha-1} + \frac{1}{\delta^2 (2\alpha - 1)} \left(\left(\frac{t_0}{2}\right)^{-1} - \left(\frac{t_0}{2}\right)^{2\alpha-2} \cdot t_0^{1-2\alpha} \right) \\ &= \frac{t_0^{2\alpha-1}}{(\Gamma(\alpha))^2 (2\alpha - 1) 2^{2\alpha-1}} + \frac{2 - 2^{2-2\alpha}}{\delta^2 (2\alpha - 1) t_0} < \infty, \end{aligned} \tag{60}$$

$$t_0^{2H-1} \int_0^{t_0} (t_0 - s)^{2\alpha-2} s^{2(1-\alpha-H)} \mathrm{d}s = B(2\alpha - 1, 3 - 2\alpha - 2H).$$

Then, by a similar argument to that used in (46), we deduce $\lim_{t\longrightarrow t_0} N_2(t) = 0$.

Moreover, we apply $(H_3) - (H_4)$ and Lemma 2 and use Lemma 4 to conclude that

$$\begin{aligned} E\left\| (\Gamma_2 u)(t) \right\| &\le 2Ht^{2H-1} \int_0^t (t - s)^{2\alpha-2} \left\| V(t - s) \right\|^2 \left\| D(s, u(s)) \right\|_{L_2^0}^2 \mathrm{d}s \\ &\le \frac{4M^2 H t^{2H-1}}{(\Gamma(\alpha))^2} \int_0^t (t - s)^{2\alpha-2} \left(\left\| D(s, 0) \right\|_{L_2^0}^2 + L_1 s^{2(1-H-\alpha)} E\left\| u(s) \right\|^2 \mathrm{d}s \right) \\ &\le \frac{4L_1 M^2 H \left\| u \right\|_{\infty}^2}{(\Gamma(\alpha))^2} \int_0^1 (1 - \mu)^{2\alpha-2} \mu^{2(1-H-\alpha)} \mathrm{d}s \\ &= \frac{4L_1 M^2 H \left\| u \right\|_{\infty}^2 B(2\alpha - 1, 3 - 2H - 2\alpha)}{(\Gamma(\alpha))^2}. \end{aligned} \tag{61}$$

Applying this and the above arguments, we conclude that $(\Gamma_2 u)(t)$ is stochastically bounded and continuous.

By combining Steps 1 and 2, we obtain $\Gamma_2 u \in \mathrm{SAP}_{\omega}([0, \infty), L^2(\Omega; \mathbb{H}))$.

Now we state our main results. □

Theorem 1. *If (H_1)–(H_4) are satisfied, then equation (14) has a unique mild solution $u \in SAP_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$ provided with*

$$2M^2\left(\frac{2L_1HB(2\alpha-1,3-2(\alpha+H))}{(\Gamma(\alpha))^2}+\frac{L}{\delta^2}\right)<1. \tag{62}$$

Proof. We define Γ by

$$(\Gamma u)(t) = U(t)u_0 + \int_0^t (t-s)^{\alpha-1}V(t-s)f(s,u(s))\mathrm{d}s + \int_0^t (t-s)^{\alpha-1}V(t-s)D(s,u(s))\mathrm{d}B_Q^H(s). \tag{63}$$

It follows from Lemmas 5 and 6 that $\Gamma u \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$. For all $t \geq 0$, $u, v \in \mathrm{SAP}_\omega([0,\infty), L^2(\Omega;\mathbb{H}))$, we get

$$\begin{aligned}
E\|(\Gamma u)(t)-(\Gamma v)(t)\|^2 &\leq E\left\|\int_0^t (t-s)^{\alpha-1}V(t-s)(f(s,u(s))-f(s,v(s)))\mathrm{d}s + \int_0^t (t-s)^{\alpha-1}V(t-s)(D(s,u(s))-D(s,v(s)))\mathrm{d}B_Q^H(s)\right\|^2\\
&\leq 2E\left\|\int_0^t (t-s)^{\alpha-1}V(t-s)(f(s,u(s))-f(s,v(s)))\mathrm{d}s\right\|^2\\
&\quad + 2E\left\|\int_0^t (t-s)^{\alpha-1}V(t-s)(D(s,u(s))-D(s,v(s)))\mathrm{d}B_Q^H(s)\right\|^2\\
&=: I(t)+J(t).
\end{aligned} \tag{64}$$

On the one hand, we can see that

$$\begin{aligned}
I(t) &\leq 2\int_0^t (t-s)^{\alpha-1}\|V(t-s)\|\mathrm{d}s\int_0^t (t-s)^{\alpha-1}\|V(t-s)\|E\|(f(s,u(s))-f(s,v(s)))\|^2\mathrm{d}s\\
&\leq 2LM^2\sup_{t\geq 0}E\|u(t)-v(t)\|^2\left(\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t-s)^\alpha)\mathrm{d}s\right)^2\\
&= \frac{2M^2L}{\delta^2}\|u-v\|_\infty^2\left(E_\alpha(-\delta(t-s)^\alpha)\Big|_0^t\right)^2\\
&\leq \frac{2M^2L}{\delta^2}\|u-v\|_\infty^2.
\end{aligned} \tag{65}$$

On the other hand, we get

$$\begin{aligned}
J(t) &\leq 4M^2Ht^{2H-1}\int_0^t (t-s)^{2\alpha-2}\left(E_{\alpha,\alpha}(-\delta(t-s)^\alpha)\right)^2\cdot\|D(s,u(s))-D(s,v(s))\|_{L_2^0}^2\mathrm{d}s\\
&\leq \frac{4M^2L_1Ht^{2H-1}}{(\Gamma(\alpha))^2}\int_0^t (t-s)^{2\alpha-2}s^{2(1-\alpha-H)}\mathrm{d}s\|u-v\|_\infty^2\\
&= \frac{4M^2L_1H}{(\Gamma(\alpha))^2}\int_0^1 (1-\tau)^{2\alpha-2}\tau^{2(1-\alpha-H)}\mathrm{d}\tau\|u-v\|_\infty^2\\
&= \frac{4M^2L_1HB(2\alpha-1,3-2(\alpha+H))}{(\Gamma(\alpha))^2}\|u-v\|_\infty^2.
\end{aligned} \tag{66}$$

Thus,

$$\|(\Gamma u)-(\Gamma v)\|_{\infty}^{2} \leq 2M^{2}\left(\frac{2L_{1}HB(2\alpha-1,3-2(\alpha+H))}{(\Gamma(\alpha))^{2}}+\frac{L}{\delta^{2}}\right)\|u-v\|_{\infty}. \tag{67}$$

If (62) holds, Γ is a contraction mapping. Using the Banach fixed-point theorem, we have that Γ has a unique fixed point in $\mathrm{SAP}_{\omega}([0,\infty),L^{2}(\Omega;\mathbb{H}))$. This completes the proof. □

4. Asymptotic Stability of Square-Mean S-Asymptotically ω-Periodic Solutions

Theorem 2. *Assume that (H_1) and (H_2') D: $[0,\infty)\longrightarrow L_2^0(\mathbb{H})$ is stochastically and continuous. Furthermore, for $\forall\varepsilon>0$, there exists $T_\varepsilon>0$ such that $t^{2(\alpha+H-1)}\|D(t+\omega)-D(t)\|_{L_2^0}^2<\varepsilon$ and $\|D(t)\|_{L_2^0}^2\leq t^{2(1-\alpha-H)}$ for $t>T$ hold. If $\delta^2>2M^2L$, there exists a unique asymptotically stable S-asymptotically ω-periodic solution u^* in square-mean sense to equation (14) with $D(t,\cdot)=D(t)$.*

Proof. From the proof process of Theorem 1, we get the existence and uniqueness of the S-asymptotically ω-periodic solution $u^*(t)$ similarly. In addition, for $\forall u_1\in L^2(\Omega;\mathbb{H})$, equation (14) has a unique mild solution $u(t)$ with the new initial value $u(0)=u_1$. And then from (17) and Lemmas 2 and 4, we have

$$\begin{aligned}
E\|u(t)-u^*(t)\|^2 &\leq 2E\|U(t)u_1-U(t)u_0\|^2\\
&\quad+2E\left\|\int_0^t(t-s)^{\alpha-1}V(t-s)\left(f(s,u(s))-f(s,u^*(s))\right)\mathrm{d}s\right\|^2\\
&\leq 2M^2E_\alpha(-\delta t^\alpha)E\|u_1-u_0\|^2+2\int_0^t(t-s)^{\alpha-1}\|V(t-s)\|\mathrm{d}s\\
&\quad\cdot\int_0^t(t-s)^{\alpha-1}\|V(t-s)\|E\|f(s,u(s))-f(s,u^*(s))\|^2\mathrm{d}s\\
&\leq 2M^2E_\alpha(-\delta t^\alpha)E\|u_1-u_0\|^2+2M^2L\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t-s)^\alpha)\mathrm{d}s\\
&\quad\cdot\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t-s)^\alpha)E\|u(s)-u^*(s)\|^2\mathrm{d}s\\
&=E_\alpha(-\delta t^\alpha)2M^2E\|u_1-u_0\|^2+\frac{2M^2L}{\delta}\int_0^t(t-s)^{\alpha-1}E_{\alpha,\alpha}(-\delta(t-s)^\alpha)\cdot E\|u(s)-u^*(s)\|^2\mathrm{d}s\\
&=E_\alpha\left(\left(\frac{2M^2L}{\delta}-\delta\right)t^\alpha\right)2M^2E\|u_1-u_0\|^2.
\end{aligned}\tag{68}$$

If $\delta^2>2M^2L$, by Lemma 1, we obtain that the square-mean S-asymptotically ω-periodic solution u^* to equation (14) with $D(t,\cdot)=D(t)$ is asymptotically stable in square-mean sense.

Now the proof could be finished. □

5. Numerical Simulation

Example 1. We consider the following fractional stochastic equation with FBM:

$$\begin{cases}D_0^{0.85}u(t)+4u(t)=\sin u(t)+\sin\sqrt{t}+\sin\dfrac{1}{t}\dfrac{\mathrm{d}B_Q^{0.6}(t)}{\mathrm{d}t}, & t\in(0,+\infty),\\ u(0)=0.22,\end{cases}\tag{69}$$

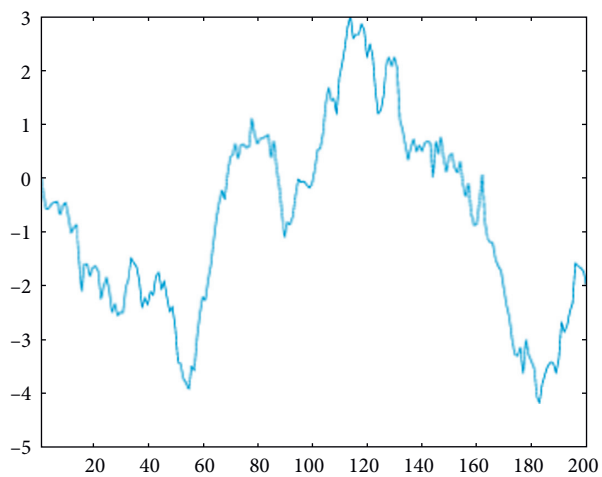


Figure 1: Fractional Brownian motion (FBM) with $H = 0.6$.

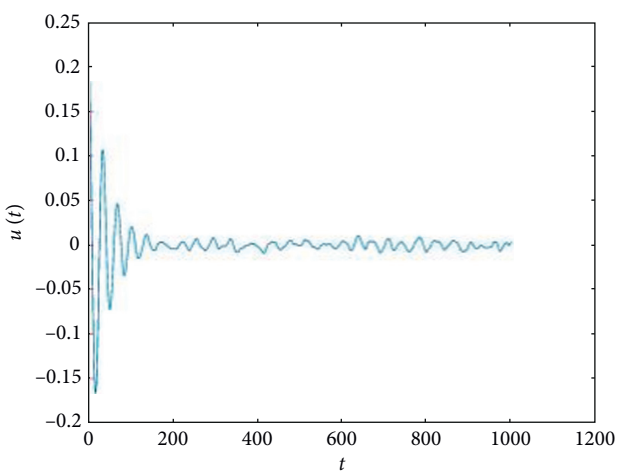


Figure 2: Time-series graph of asymptotically stable S-asymptotically ω-periodic solution of (69).

where $D_0^{0.85}$ is the Caputo fractional derivative and $B_Q^{0.6}(t)$ is one-dimensional FBM with $H = 0.6$ (Figure 1).

We notice that -4 generates an exponentially stable semigroup $\{e^{-4t}\}_{t\geq 0}$, $\delta = 4$, and $M = 1$. Set $f(t, u(t)) = \sin u(t) + \sin\sqrt{t}$ and $D(t) = \sin(1/t)$. Then, by the Lagrange differential mean value theorems, we obtain that (H_1) and (H_2') are satisfied with $L = 1$. It is easy to see that $\delta^2 > 2M^2L$. From Theorem 2, we conclude that equation (69) has a unique square-mean S-asymptotically ω-periodic solution (Figure 2), which is asymptotically stable in square-mean sense.

From the above example, we find that although there is no periodic solution for the fractional-order differential equation with finite lower limit [2], the S-asymptotically periodic solution can be found and is stable even for fractional stochastic differential equation.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Scientific Research Foundation of the Higher Education Institutions of Gansu Province (2019B-022), National Natural Science Foundation of China (11671339), Fundamental Research Funds for the Central Universities (31920180047), and Innovation Team of Intelligent Computing and Dynamical System Analysis and Application of Northwest Minzu University.

References

[1] J. Mu, B. Ahmad, and S. Huang, "Existence and regularity of solutions to time-fractional diffusion equations," *Computers & Mathematics with Applications*, vol. 73, no. 6, pp. 985–996, 2017.

[2] J. Mu, Y. Zhou, and L. Peng, "Periodic solutions and S-asymptotically periodic solutions to fractional evolution equations," *Discrete Dynamics in Nature and Society*, vol. 2017, Article ID 1364532, 12 pages, 2017.

[3] Y. A. Rossikhin and M. V. Shitikova, "Application of fractional calculus for dynamic problems of solid mechanics: novel trends and recent results," *Applied Mechanics Reviews*, vol. 63, no. 1, p. 10801, 2010.

[4] H. Sun, Y. Zhang, D. Baleanu, W. Chen, and Y. Chen, "A new collection of real world applications of fractional calculus in science and engineering," *Communications in Nonlinear Science and Numerical Simulation*, vol. 64, pp. 213–231, 2018.

[5] Q. Yang, D. Chen, T. Zhao, and Y. Chen, "Fractional calculus in image processing: a review," *Fractional Calculus and Applied Analysis*, vol. 19, no. 5, pp. 1222–1249, 2016.

[6] Y. Zhou, "Attractivity for fractional differential equations in Banach space," *Applied Mathematics Letters*, vol. 75, pp. 1–6, 2018.

[7] Y. Zhou, "Attractivity for fractional evolution equations with almost sectorial operators," *Fractional Calculus and Applied Analysis*, vol. 21, no. 3, pp. 786–800, 2018.

[8] Y. Zhou, *Basic Theory of Fractional Differential Equations*, World Scientific Publishing, Singapore, 2014.

[9] Y. Zhou, *Fractional Evolution Equations and Inclusions: Analysis and Control*, Academic Press, Cambridge, MA, USA, 2016.

[10] A. Gu, D. Li, B. Wang, and H. Yang, "Regularity of random attractors for fractional stochastic reaction-diffusion equations on $\mathbb{R}^n$," *Journal of Differential Equations*, vol. 264, no. 12, pp. 7094–7137, 2018.

[11] B. Wang, "Asymptotic behavior of non-autonomous fractional stochastic reaction-diffusion equations," *Nonlinear Analysis*, vol. 158, pp. 60–82, 2017.

[12] L. Chen, "Nonlinear stochastic time-fractional diffusion equations on $\mathbb{R}$: moments, Hölder regularity and intermittency," *Transactions of the American Mathematical Society*, vol. 369, no. 12, pp. 8497–8535, 2017.

[13] L. Peng and Y. Huang, "On nonlocal backward problems for fractional stochastic diffusion equations," *Computers & Mathematics with Applications*, vol. 78, no. 5, pp. 1450–1462, 2019.

[14] Y. Li and Y. Wang, "The existence and asymptotic behavior of solutions to fractional stochastic evolution equations with

infinite delay," *Journal of Differential Equations*, vol. 266, no. 6, pp. 3514–3558, 2019.

[15] K. Mathiyalagan and K. Balachandran, "Finite-time stability of fractional-order stochastic singular systems with time delay and white noise," *Complexity*, vol. 21, no. S2, pp. 370–379, 2016.

[16] K. Liu, J. Wang, Y. Zhou, and D. O'Regan, "Hyers-Ulam stability and existence of solutions for fractional differential equations with Mittag-Leffler kernel," *Chaos, Solitons & Fractals*, vol. 132, p. 109534, 2020.

[17] J. Liu, W. Xu, and Q. Guo, "Global attractiveness and exponential stability for impulsive fractional neutral stochastic evolution equations driven by FBm," *Advances in Difference Equations*, vol. 2020, no. 1, pp. 1–17, 2020.

[18] L. Zhang, Y. Ding, K. Hao, L. Hu, and T. Wang, "Moment stability of fractional stochastic evolution equations with Poisson jumps," *International Journal of Systems Science*, vol. 45, no. 7, pp. 1539–1547, 2014.

[19] H. Ye, J. Gao, and Y. Ding, "A generalized Gronwall inequality and its application to a fractional differential equation," *Journal of Mathematical Analysis and Applications*, vol. 328, no. 2, pp. 1075–1081, 2007.

[20] M. A. Diop and M. J. Garrido-Atienza, "Retarded evolution systems driven by fractional Brownian motion with Hurst parameter $H > (1/2)$," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 97, pp. 15–29, 2014.

[21] K. Claudia and K. Christoph, "Fractional Brownian motion as a weak limit of Poisson shot noise processes with applications to finance," *Stochastic Processes and Their Applications*, vol. 113, no. 2, pp. 333–351, 2004.

[22] G. Arthi, J. H. Park, and H. Y. Jung, "Existence and exponential stability for neutral stochastic integrodifferential equations with impulses driven by a fractional Brownian motion," *Communications in Nonlinear Science and Numerical Simulation*, vol. 32, pp. 145–157, 2016.

[23] T. Caraballo, M. J. Garrido-Atienza, and T. Taniguchi, "The existence and exponential behavior of solutions to stochastic delay evolution equations with a fractional Brownian motion," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 74, no. 11, pp. 3671–3684, 2011.

[24] L. H. Duc, M. J. Garrido-Atienza, A. Neuenkirch, and B. Schmalfuß, "Exponential stability of stochastic evolution equations driven by small fractional Brownian motion with Hurst parameter in (1/2, 1)," *Journal of Differential Equations*, vol. 264, no. 2, pp. 1119–1145, 2018.

[25] M. J. Garrido-Atienza, K. Lu, and B. Schmalfuss, "Random dynamical systems for stochastic partial differential equations driven by a fractional Brownian motion," *Discrete and Continuous Dynamical Systems-Series B*, vol. 14, no. 2, pp. 473–493, 2010.

[26] P. Tamilalagan and P. Balasubramaniam, "Moment stability via resolvent operators of fractional stochastic differential inclusions driven by fractional Brownian motion," *Applied Mathematics and Computation*, vol. 305, pp. 299–307, 2017.

[27] K. Li, "Stochastic delay fractional evolution equations driven by fractional Brownian motion," *Mathematical Methods in The Applied Sciences*, vol. 38, no. 8, pp. 1582–1591, 2015.

[28] Z. Ji, H. Lin, S. Cao, Q. Qi, and H. Ma, "The complexity in complete graphic characterizations of multiagent controllability," *IEEE Transactions on Cybernetics*, 2020.

[29] L. Mo and S. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[30] S. Liu, Z. Ji, and H. Ma, "Jordan form-based algebraic conditions for controllability of multiagent systems under directed graphs," *Complexity*, vol. 2020, Article ID 7685460, 18 pages, 2020.

[31] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, *Theory and Applications of Fractional Differential Equations*, Elsevier, Amsterdam, Netherlands, 2006.

[32] Z. Wei, W. Dong, and J. Che, "Periodic boundary value problems for fractional differential equations involving a Riemann-Liouville fractional derivative," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 73, no. 10, pp. 3232–3238, 2010.

[33] Z. Wei, Q. Li, and J. Che, "Initial value problems for fractional differential equations involving Riemann-Liouville sequential fractional derivative," *Journal of Mathematical Analysis and Applications*, vol. 367, no. 1, pp. 260–272, 2010.

[34] X.-B. Shu, Y. Lai, and Y. Chen, "The existence of mild solutions for impulsive fractional partial differential equations," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 74, no. 5, pp. 2003–2011, 2011.

[35] A. M. Krägeloh, "Two families of functions related to the fractional powers of generators of strongly continuous contraction semigroups," *Journal of Mathematical Analysis and Applications*, vol. 283, no. 2, pp. 459–467, 2003.

[36] R.-N. Wang, D.-H. Chen, and T.-J. Xiao, "Abstract fractional Cauchy problems with almost sectorial operators," *Journal of Differential Equations*, vol. 252, no. 1, pp. 202–235, 2012.

[37] M. M. El-Borai, "Some probability densities and fundamental solutions of fractional evolution equations," *Chaos, Solitons & Fractals*, vol. 14, no. 3, pp. 433–440, 2002.

[38] Y. Zhou and F. Jiao, "Existence of mild solutions for fractional neutral evolution equations," *Computers & Mathematics with Applications*, vol. 59, no. 3, pp. 1063–1077, 2010.

[39] B. Boufoussi and S. Hajji, "Neutral stochastic functional differential equations driven by a fractional Brownian motion in a Hilbert space," *Statistics & Probability Letters*, vol. 82, no. 8, pp. 1549–1558, 2012.

Hindawi
Complexity
Volume 2020, Article ID 3680518, 19 pages
https://doi.org/10.1155/2020/3680518

Research Article

Attack-Defense Game between Malicious Programs and Energy-Harvesting Wireless Sensor Networks Based on Epidemic Modeling

Guiyun Liu, Baihao Peng, Xiaojing Zhong, Lefeng Cheng, and Zhifu Li

School of Mechanical and Electric Engineering, Guangzhou University, Guangzhou 510006, China

Correspondence should be addressed to Baihao Peng; 2111807063@e.gzhu.edu.cn

Received 15 July 2020; Revised 19 August 2020; Accepted 27 August 2020; Published 14 September 2020

Academic Editor: Ning Cai

As energy-harvesting wireless sensor networks (EHWSNs) are increasingly integrated with all walks of life, their security problems have gradually become hot issues. As an attack means, malicious programs often attack sensor nodes in critical locations in the networks to cause paralysis and information leakage of the networks, resulting in security risks. Based on the previous works and the introduction of solar charging, we proposed a novel model, namely, Susceptible-Infected-Low (energy)-Recovered-Dead (SILRD) with solar energy harvesters. Meanwhile, this paper takes Logistic Growth as the drop rate of sensor nodes and the infection rate of multitype malicious programs under nonlinear condition into consideration. Finally, an Λ-Susceptible-Infected-Low (energy)-Recovered-Dead (ΛSILRD) model is proposed. Based on the Pontryagin Maximum Principle, this paper proposes the optimal strategies based on the SILRD with solar energy harvesters and the ΛSILRD. The effectiveness of SILRD with solar energy harvesters was demonstrated by comparison with the general epidemic model. At the same time, by analyzing different charging strategies, we conclude that solar charging is highly efficient. Moreover, we further analyze the influence of controllable and uncontrollable input and various node degrees on ΛSILRD model.

1. Introduction

With the rapid development of wireless sensor networks (WSNs) in the past few years, the unique characteristics of WSNs have enabled them to play a key role in many fields, such as military strike, agricultural production, intelligent transportation, medical and health systems, and industrial fields. Typical WSNs consist of a series of sensor nodes with different functions of gathering environment information and transmitting processed information, which is either fixed or randomly distributed. Generally, the coverage area of the networks is much larger than the maximum transmission distance of each sensor node. Therefore, transmission between sensor nodes and terminal computers or control centers is normally conducted in multihop. However, the limited energy vastly confines the lifetime of the networks. Renewable natural resources, such as wind, solar, and tidal energy, can be transferred to electricity by certain energy harvesters, which can greatly mitigate the impact of energy shortage on the lifetime of energy-harvesting wireless sensor networks (EHWSNs) which is equipped with energy harvesters on each sensor node.

However, the structure of WSNs provides a hotbed for the propagation of malicious programs. Furthermore, low defense capabilities render sensor nodes more vulnerable to malicious programs. With the booming of WSNs in all walks of life, paralysis and information leakage owing to malicious programs will cause unpredictable economic losses. Along with the ceaseless invasion of malicious programs, many scholars have studied the behavior characteristics of malicious programs and develop corresponding countermeasures. Due to the similarity of propagation mechanisms between malicious programs and infectious disease, epidemic models which are suitable for WSNs have been further developed after decades of study based on the initial Susceptible-Infected-Recovered (SIR) model proposed by

Kermack and McKendrick in 1927 [1]. Unlike computer viruses, the spread of malicious programs in WSNs is affected by node distribution density and communication radius [2]. Also, geospatial limitation [3], spatial correlation [4], coupling degree [5], and transmission delay [6] have effects on it. In addition to a series of removal methods, predicting the risk of infection of sensor nodes is also one of the current research hotspots [7]. Furthermore, a two-level bidirectional data prediction model proposed by Wang et al. can effectively reduce the data collection cost of the underwater acoustic sensor network and improve the utilization rate of bandwidth [8]. Li et al. applied machine learning methods to improve the efficiency of malware detection [9]. Han et al. proposed a DMPPR scheme to protect users' privacy of WSNs [10]. Zhao et al. proposed a method to detect an undetectable false data-injection attack in cyber-physical systems [11]. In the prevention of malicious programs, Li et al. [12] and Liu et al. [13], respectively, proposed a LightLRFMS algorithm and TPE-FTED algorithm to screen false data caused by malicious attacks and system failures. Dynamic defenses are also valid methods [14, 15] compared with static approaches. Wu et al. proposed a novel model and defense mechanism to effectively protect big data in the networks owing to its vulnerability to virus attacks [16].

Although few scholars discuss the issue of energy in the sensor network attacked by malicious programs, effectively, the increase in the energy of WSNs is always a hot topic. For example, Mo et al. used multiple mobile chargers to supplement the energy of WSNs [17, 18]. It is worth mentioning that the control of multiagent is one of the research hotspots in recent years [19, 20]. In this paper, sensor nodes have been divided into five states based on the remaining energy, and the energy-harvesting technology is introduced to supplement the energy of sensor nodes to extend the lifespan of the EHWSNs. On the basis of [21], this paper changes the charging method to solar charging and constructs a new model, namely, Susceptible-Infected-Low (energy)-Recovered-Dead (SILRD) with solar energy harvesters. At the same time, considering networks input and multitypes of malicious programs attacks with nonlinear infection rates, a novel model named Λ-Susceptible-Infected-Low (energy)-Recovered-Dead (ΛSILRD) is proposed.

Similar to conflict of interest in a game, game theory can be applied to get optimal solutions, like optimal Dissemination of Security Patches [22], optimal power control [23], optimal detection rate [24], optimal data transmission strategy [25], optimal hardware deployment cost in EHWSNs [26], optimal delay and transmission times in the networks [27], suitable game strategy and price adjustment principle in cyber-physical-social systems (CPSS) [28], maximization of energy efficiency [29], and optimal multipath routing [30]. As an essential part of game theory, the differential game can describe the dynamic process with differential equations. Mylvaganam et al. find the optimal control in multiagent collision avoidance [31]. Miao and Li [32] derive the optimal strategies for the attackers and the intrusion prevention systems. Miao et al. [33] find an optimal solution based on tradeoff between network throughput and energy efficiency. In this paper, the differential game will be applied to solve the confrontation problem between malicious programs and EHWSNs, and the optimal attack-defense strategies for both parties have been proposed.

Our contributions are summarized in the following paragraphs.

The low-energy state is introduced into the basic epidemic model considering the limited energy of sensor nodes. To suppress the spread of malicious programs, the method of charging by solar energy harvesters is put forward which is helpful to alleviate the security problems and maintenance of the networks. Thus, a model named SILRD with solar energy harvesters which better fits EHWSNs has been proposed. The effectiveness of SILRD with solar energy harvesters is obtained by comparison with existing epidemic models. The efficiency of solar charging is demonstrated by comparing with different charging strategies.

Three nonlinear factors will be considered in this paper, including Logistic Growth, nonlinear infection rate, and charging power provided by solar energy harvesters. At the same time, this paper takes the impact of multitypes of malicious programs into consideration. Finally, a novel attack-defense game model named ΛSILRD is proposed in this paper. Meanwhile, the influence of controllable input, uncontrollable input, and various node degree on ΛSILRD model and EHWSNs has been discussed.

Based on the ΛSILRD model, the optimal dynamic control strategies for EHWSNs and malicious programs under various node degrees are proposed by applying Pontryagin Maximum Principle.

The rest of the paper is organized as follows. In Section 2, the nonlinear factors involved in ΛSILRD model will be proposed first, and then ΛSILRD model will be introduced in detail. In Section 3, the Pontryagin Maximum Principle will be used to find the optimal dynamic control strategies and the optimality will be proved briefly after the introduction of the expressions of control variables and game cost. In Section 4, the effectiveness of SILRD with solar energy harvesters and solar charging, the influence of controllable and uncontrollable input, and node degree on ΛSILRD will be demonstrated through simulations. Section 5 is the conclusion and prospect of this paper.

2. ΛSILRD Model in EHWSNs

This section explains the nonlinear factors at first, including Logistic Growth, nonlinear infection rate, and solar charging power. Then, on the premise of considering multiple types of malicious programs' attacks, the ΛSILRD model is proposed.

2.1. Nonlinear Factors in ΛSILRD Model. In this paper, EHWSNs consist of identical sensor nodes equipped with solar energy harvesters, which are distributed statically. The solar energy harvesters energize sensor nodes according to the duration of sunlight. In the case of the diurnal period, the charging power generally goes through a process of increasing firstly and then decreasing. Specifically, the

charging power increases slowly at first since the sunlight intensity is weak at dawn. With the arrival of noon, sunlight intensity increasingly reaches the maximum value of the day, while charging power at this time is also at a maximum. However, the charging power will show a downward trend when night falls. In this paper, 24 hours is assumed as a period and the case of fine days is only considered. In particular, (1) is used to describe the trend of solar charging power over a day [34]:

$$P(t) = \frac{A}{\sqrt{2\pi}n} e^{-\left((t-m)^2/2n^2\right)}, \tag{1}$$

where A is the intensity of solar charging power and m and n are the mean value and variance value of the power distribution, respectively.

To maintain the functioning of EHWSNs, it is indispensable to deploy new nodes when conditions permit. This paper considers Logistic Growth as input rate of new nodes. Logistic Growth is formulated by

$$\Lambda(t) = rS(t)\left[1 - \frac{S(t)}{k}\right], \tag{2}$$

where r represents the node degree, k represents the capacity of EHWSNs, and $S(t)$ represents the quantity of susceptible nodes at time t.

Compared with the linear infection rate, the nonlinear infection rate can better describe the ability of malicious programs to propagate in a limited area. Models with linear infection rates, where the quantity of infected nodes grows linearly, are impractical. Actually, the quantity of infections is bound to increase exponentially at first. As time progresses, the quantity grows steadily until it eventually infects the entire networks. According to the above description of infection process, (3) is applied to express it:

$$Y(t) = \left[1 - \left(1 - P_{SI}\right)^{nI(t)}\right], \tag{3}$$

where P_{SI} is the probability of infection, $I(t)$ is the quantity of infected nodes at time t, and n represents the connectivity of nodes.

2.2. Model with Solar Energy Harvester. There exist two-time intervals without sunlight in one day. Specifically, one is from 0 am to 5 am and the other is from 8 pm to 12 pm [34]. In these two intervals, solar energy harvesters knock off and sensor nodes may be dysfunctional since electricity drains out. According to the energy levels and the infection status, sensor nodes have been divided into five states.

Susceptible (S) State. Sensor nodes in the susceptible state are with high-energy level and can complete assignment normally. Without defense measures, susceptible sensor nodes are vulnerable to malicious programs.

Infected (I) State. Sensor nodes in the infected state are transformed from susceptible, recovered, or low-energy sensor nodes by running malicious programs. In the early stage of infection or after charging, infected sensor nodes are still at a high-energy level because the extent of damage has not yet been reached.

Low-Energy (L) State. Sensor nodes in the low-energy state are with energy which are too insufficient to function properly, including information transmission. Therefore, malicious programs attached to low-energy sensor nodes do not have the ability to continue infecting. Similarly, low-energy sensor nodes will not be patched.

Recovered (R) State. Sensor nodes in the recovered state have installed the patches successfully. Also, recovered sensor nodes are all in high-energy level. The patches are only applicable to relevant malicious programs. In the face of attacks by inhomogeneous malicious programs, these sensor nodes will also be helpless and transform to infected state.

Dead (D) State. Sensor nodes in the dead state are absolute dysfunction compared with low-energy sensor nodes. Dead sensor nodes no longer own the ability to collect, process, and transmit information.

At time t, the proportion of the number of sensor nodes in susceptible, infected, recovered, low-energy, and dead states is $S(t)$, $I(t)$, $L(t)$, $R(t)$, and $D(t)$, respectively. And the following equation must be met:

$$S(t) + I(t) + L(t) + R(t) + D(t) = 1. \tag{4}$$

In the absence of sunlight, the networks rely on the residual energy to maintain functioning. New sensor nodes are cast randomly to keep the connectivity of EHWSNs. Susceptible nodes still consume electricity at night to continue data acquisition, processing, and transmission. Under the attack of malicious programs, susceptible sensor nodes are transformed into infected sensor nodes with probability P_{SI}. Some susceptible sensor nodes are fortunately enough to be patched to possess immunity with probability P_{SR}. The rest stick at their daily tasks normally with probability P_{SL}.

Infected sensor nodes transmit data to neighbors at higher frequencies to spread malicious programs rapidly and disrupt the transmission mechanism. Therefore, infected sensor nodes will consume the remaining electricity at a faster rate and transform to low-energy or dead state with probability P_{IL} and P_{ID} according to the attack power of malicious programs. While malicious programs are spreading arbitrarily, patches carried by unmanned aerial vehicles (UAVs) transmitted to the infected sensor nodes located at the corresponding district with probability P_{IR}.

The existence of multiple types of malicious programs is considered and the common feature of these malicious programs is that their attack mechanisms are embodied in the accelerated consumption of sensor nodes' energy. For this reason, even recovered sensor nodes will be infected again with probability P_{RI}. Similarly, few recovered sensor nodes work normally until low-energy level with probability P_{RL}.

Sensor nodes at high-energy levels include sensor nodes in susceptible, infected, and recovered state. Energy consumption owing to damage or normal operation will eventually convert sensor nodes in high-energy state to low-energy state. Sensor nodes at low-energy levels suspend some functions, including data transmission, for their own subsistence. Therefore, low-energy sensor nodes will not receive and transmit malicious programs. Even if the consumption is lower, the energy will eventually run out with probability P_{LD}. With the use of solar energy harvesters, the probabilities of sensor nodes in low-energy state converting into susceptible, infected, and recovered states are $P_{LS}P(t)$, $P_{LI}P(t)$, and $P_{LR}P(t)$, respectively. Among them, P_{LS} is related to the number of sensor nodes that transformed from susceptible state to low-energy state at the previous moment, P_{LI} is related to the number of sensor nodes that transformed from infected state to low-energy state at the previous moment, and P_{LR} is related to the number of nodes that switched from an infected state to a low-energy state at the previous moment. In particular, Figure 1 is used to visualize the evolution of sensor nodes. Figure 1 shows a part of sensor nodes in EHWSNs. Among them, the letter in the circle represents the node state. Specifically, Figure 1(a) shows the initial node state when the patch-carrying UAV has not yet passed the sensor nodes and the solar energy harvesters have started working. Figure 1(b) shows the evolution of sensor nodes after the UAV drives over a part of sensor nodes and the solar energy harvesters charge sensor nodes.

Considering the Logistic Growth (2) and the nonlinear incidence rate (3), the above dynamic processes are formulated in (5)–(9), and the flow diagram of propagation is shown in Figure 2:

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = \Lambda(t) - Y(t)S(t) - P_{SR}S(t) - P_{SL}S(t) + P_{LS}P(t)L(t), \tag{5}$$

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = Y(t)S(t) - P_{IR}I(t) - P_{IL}I(t) + P_{RI}R(t) - P_{ID}I(t) + P_{LI}P(t)L(t), \tag{6}$$

$$\frac{\mathrm{d}L(t)}{\mathrm{d}t} = P_{IL}I(t) + P_{SL}S(t) - P_{LD}L(t) + P_{RL}R(t) - P_{LS}P(t)L(t) - P_{LI}P(t)L(t) - P_{LR}P(t)L(t), \tag{7}$$

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = P_{IR}I(t) + P_{SR}S(t) - P_{RI}R(t) - P_{RL}R(t) + P_{R}(t)L(t), \tag{8}$$

$$\frac{\mathrm{d}D(t)}{\mathrm{d}t} = P_{LD}L(t) + P_{ID}I(t). \tag{9}$$

3. Optimal Controls in Attack-Defense Game

In this section, control variables between malicious programs and EHWSNs are introduced at first. Then, the process of attack-defense game has been analyzed and the overall cost has been formulated. Finally, the Hamiltonian function has been built and constructed and the optimal strategies of both sides are obtained on the basis of proving the existence and the uniqueness.

3.1. Control Variables in the ΛSILRD Model. The attacks of malicious programs are mainly reflected in the propagation performance and the damage capacity. The more contagious malicious programs are, the more sensor nodes they can infect. Infected sensor nodes can spread malicious programs by increasing communication frequency. The damage capacity is incarnated in the consumption of energy and the destruction of hardware. Some malicious programs can overload sensor nodes so that they can become dysfunctional quickly, and other malicious programs cannot directly destroy sensor nodes because damage capacities are not powerful enough [35].

The defense measures applied by EHWSNs are the deployment of patch-carrying UAVs and the installing and running of solar energy harvesters. Because of the periodicity of sunlight, patching is the only countermeasure at night. By identifying and analyzing multitype malicious programs, UAVs will download corresponding patches from the base station. Energy supplements do not cure infected sensor nodes but only alleviate their severe consumption.

It is not hard to find that the propagation performance of malicious programs is the process of transformation from susceptible or recovered state to infected state. The attacks on the above transformations are defined as $A_{SI}(t)$ and $A_{RI}(t)$. At the same time, the damage capacities are reflected in the process of transformation from infected state to low-energy or dead state. Thus, the attacks are defined as $A_{IL}(t)$ and $A_{ID}(t)$. The defenses of EHWSNs are embodied in all sensor nodes that are transformed into recovered state. Therefore, the defense measures are defined as $D_{SR}(t)$ and $D_{IR}(t)$.

According to the above statement, the corresponding probability can be replaced by the equations containing the control variables. Specifically, P_{IL} can be replaced by $(A_{IL}(t)S_{IL}/(A_{IL\max} + A_{IL\min}))$, P_{ID} can be replaced by

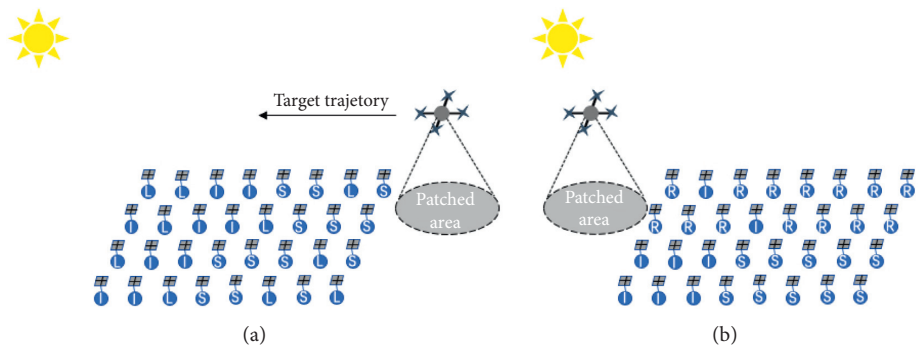


Figure 1: Schematic of ΛSILRD model. (a) The initial state of sensor nodes; (b) the current state of sensor nodes after solar charging and UAVs' patching.

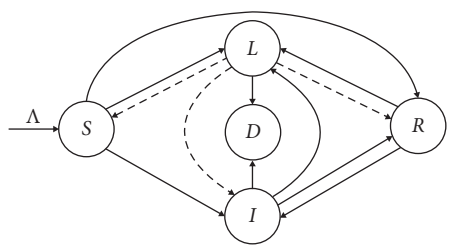


Figure 2: The flow diagram of ΛSILRD model.

$(A_{ID}(t)S_{ID}/(A_{ID\max}+A_{ID\min}))$, P_{RI} can be replaced by $(A_{RI}(t)S_{RI}/(A_{RI\max}+A_{RI\min}))$, P_{SR} can be replaced by $(D_{SR}(t)S_{SR}/(D_{SR\max}+D_{SR\min}))$, and P_{IR} can be replaced by $(D_{IR}(t)S_{IR}/(D_{IR\max}+D_{IR\min}))$. In particular, for the convenience of calculation, $(A_{SI}(t)S_{SI}/(A_{SI\max}+A_{SI\min}))$ will be directly multiplied by the term corresponding to the nonlinear incidence rate. The subscripts max and min, respectively, represent the maximum and the minimum values of the attacks or defenses, and the letter S represents the probability of the successful attacks or defenses and its subscript represents relevant state transition relationship.

3.2. Cost Function in ΛSILRD Model. The continuous confrontation between malicious programs and EHWSNs constitutes the attack-defense game. The purposes of malicious programs as attacker are to infect and destroy as many nodes as possible, while EHWSNs as defender aim to keep as many nodes immune and alive as possible. Both sides achieve their goals through the control means mentioned in the previous part. The cost function is applied to indicate the consequence of attack-defense game.

Deployment costs include production costs and human costs. Cost factor C_N times the drop rate formulated in (2) is used to represent the deployment costs at time t, where C_N is greater than 0. By completing daily assignments, sensor nodes in susceptible state which send involved information to clients to meet their requirements will generate positive benefits.

Although the unstoppable spread of malicious programs causes more sensor nodes to be infected, infected sensor nodes do not initially incur additional costs until malicious programs begin to run. $C_I I(t)$ is used to describe the costs incurred after malicious programs run at time t, where C_I is greater than 0.

As a defense means for EHWSNs, the transmission of patches to susceptible and infected sensor nodes will incur costs. $C_{SR}P_{SR}S(t)$ is used to describe the cost of transmitting patches to susceptible sensor nodes at time t, and $C_{IR}P_{IR}I(t)$ is used to describe the cost of transmitting patches to infected sensor nodes at time t, where C_{SR} and C_{IR} are greater than 0.

Recovered sensor nodes are similar to susceptible nodes. Because they have immunity to certain types of malicious programs, the revenue generated by recovered nodes at time t will be higher than that of the susceptible sensor nodes.

Low-energy sensor nodes cannot operate normally compared with the high-energy sensor nodes, so that they can incur the cost $C_L L(t)$ at time t, where C_L is greater than 0. The generation of dead sensor nodes will lead to the interruption of the connection of sensor nodes, or even the paralysis of the networks, and $C_D D(t)$ is used to describe the cost at time t, where C_D is greater than 0. It is worth noting that since solar energy harvesters capture natural resources, the cost of energy harvesting and transformation is not considered.

At the end of the game, each type of sensor nodes will incur a series of termination payoff. Among them, susceptible sensor nodes and recovered sensor nodes will still generate revenue in the future, so their terminal costs should be less than 0; that is, C_{S_f} and C_{R_f} are both less than 0. Conversely, sensor nodes at infected, low-energy, and dead states will still cause loss in the future; that is, C_{I_f}, C_{L_f}, and C_{D_f} are greater than 0.

Based on the above statement, the cost function (10) is constructed as follows:

$$J(t) = \int_{t_0}^{t_f} \left\{ C_I I(t) + rC_N S(t)\left[1 - \frac{S(t)}{k}\right] + C_L L(t) + C_D D(t) + C_{SR} P_{SR} S(t) + C_{IR} P_{IR} I(t) \right\} \mathrm{d}t + C_{S_f} S(t_f) + C_{I_f} I(t_f) + C_{L_f} L(t_f) + C_{R_f} R(t_f) + C_{D_f} D(t_f). \tag{10}$$

3.3. Optimal Strategies in the ΛSILRD Model. Suppose the duration of the game is T. Within T, according to (5)–(9), it is not difficult to find that the state variables are continuous and uninterrupted. Meanwhile, the state variables are continuous in the cost function (10).

For the control variables, they are not only continuous in state functions (5)–(9) and cost function (10), but also linear. Specifically, according to the assumptions in Section 3.1, each control variable has a maximum value and a minimum value; that is, each control variable is bounded. Furthermore, we define $\nu(t)$ as a set of control strategies of attacker (malicious programs), that is, $\nu(t) = \{A_{SI}(t), A_{IL}(t), A_{ID}(t), A_{RI}(t)\}$, and $\mu(t)$ as a set of control strategies of defender (EHWSNs), that is, $\mu = \{D_{SR}(t), D_{IR}(t)\}$.

Thus, according to the conditions put forward by [36], there must exist a saddle point in (10), which satisfies

$$J(\mu^*(t), \nu(t)) \le J(\mu^*(t), \nu^*(t)) \le J(\mu(t), \nu^*(t)), \tag{11}$$

where $J(\mu^*(t), \nu(t))$ represents the cost incurred when only the optimal strategy is selected by EHWSNs, $J(\mu(t), \nu^*(t))$ denotes that only malicious programs choose the optimal strategy and $J(\mu^*(t), \nu^*(t))$ indicates that both EHWSNs and malicious programs choose the optimal strategies.

Specifically, the inequality on the left indicates that when strategy chosen by EHWSNs is unchanged, the cost will be maximized when the malicious programs choose the optimal strategy; the inequality on the right indicates the cost will be minimized when EHWSNs select the optimal strategy, while the strategy chosen by malicious programs remains unchanged. When both parties choose the optimal strategies, an equilibrium point, the saddle point, will be formed between the maximum cost generated by the malicious programs and the minimum cost generated by EHWSNs.

According to [37] and the characteristics of this model, we can further know that there must be V satisfying the following equation:

$$\begin{aligned} V &= \max_{\nu(t)} \min_{\mu(t)} J(\mu(t), \nu(t)) \\ &= \min_{\mu(t)} \max_{\nu(t)} J(\mu(t), \nu(t)) \\ &= J(\mu^*(t), \nu^*(t)), \end{aligned} \tag{12}$$

where $\max_{\nu(t)}\min_{\mu(t)} J(\mu(t), \nu(t))$ represents the cost incurred by EHWSNs in selecting the optimal strategy after the malicious programs make an optimal decision, while $\min_{\mu(t)}\max_{\nu(t)} J(\mu(t), \nu(t))$ denotes the cost incurred when the order of two sides is switched.

Theorem 1. *In the attack-defense game based on the ΛSILRD model, the optimal dynamic strategies of EHWSNs and malicious programs are*

$$A_{SI}^*(t) = \begin{cases} A_{SI\max}; & \beta_{A_{SI}} > 0, \\ \text{unknown}; & \beta_{A_{SI}} = 0, \\ A_{SI\min}; & \beta_{A_{SI}} < 0, \end{cases} \tag{13}$$

$$A_{IL}^*(t) = \begin{cases} A_{IL\max}; & \beta_{A_{IL}} > 0, \\ \text{unknown}; & \beta_{A_{IL}} = 0, \\ A_{IL\min}; & \beta_{A_{IL}} < 0, \end{cases} \tag{14}$$

$$A_{ID}^*(t) = \begin{cases} A_{ID\max}; & \beta_{A_{ID}} > 0, \\ \text{unknown}; & \beta_{A_{ID}} = 0, \\ A_{ID\min}; & \beta_{A_{ID}} < 0, \end{cases} \tag{15}$$

$$A_{RI}^*(t) = \begin{cases} A_{RI\max}; & \beta_{A_{RI}} > 0, \\ \text{unknown}; & \beta_{A_{RI}} = 0, \\ A_{RI\min}; & \beta_{A_{RI}} < 0, \end{cases} \tag{16}$$

$$D_{SR}^*(t) = \begin{cases} D_{SR\max}; & \beta_{D_{SR}} > 0, \\ \text{unknown}; & \beta_{D_{SR}} = 0, \\ D_{SR\min}; & \beta_{D_{SR}} < 0, \end{cases} \tag{17}$$

$$D_{IR}^*(t) = \begin{cases} D_{IR\max}; & \beta_{D_{IR}} > 0, \\ \text{unknown}; & \beta_{D_{IR}} = 0, \\ D_{IR\min}; & \beta_{D_{IR}} < 0, \end{cases} \tag{18}$$

where discriminant parameters are shown in Table 1.

Proof. Define $x(t) = \{S(t), I(t), L(t), R(t), D(t)\}$ in the ΛSILRD model. For all t which belongs to T, if $H(x(t), \mu^*(t), \nu(t), t) \le H(x(t), \mu^*(t), \nu^*(t), t) \le H(x(t), \mu(t), \nu^*(t), t)$ is satisfied, there must be an optimal set of strategies $(\mu^*(t), \nu^*(t))$ according to [38].

First, the generalization of the cost function in the game will be described as follows:

$$J(\mu(t), \nu(t)) = \varphi(x(t_1), t_1) + \int_{t_0}^{t_f} L(x(t), \mu(t), \nu(t), t)\mathrm{d}t + \upsilon\psi(x(t_1), t_1). \tag{19}$$

Define a new function ϕ as follows:

Table 1: Table of parameters in optimal strategies.

Letter	Counterpart
$\beta_{A_{SI}}$	$(\lambda_{I(t)} - \lambda_{S(t)})[1 - (1 - P_{SI})^{nI^*(t)}]S^*(t)$
$\beta_{A_{IL}}$	$(\lambda_{L(t)} - \lambda_{I(t)})P_{IL}I^*(t)$
$\beta_{A_{ID}}$	$(\lambda_{D(t)} - \lambda_{I(t)})P_{ID}I^*(t)$
$\beta_{A_{RI}}$	$(\lambda_{I(t)} - \lambda_{R(t)})P_{RI}R^*(t)$
$\beta_{D_{SR}}$	$(\lambda_{R(t)} - \lambda_{S(t)})P_{SR}S^*(t) + C_{SR}P_{SR}S^*(t)$
$\beta_{D_{IR}}$	$(\lambda_{R(t)} - \lambda_{I(t)})P_{IR}I^*(t) + C_{IR}P_{IR}I^*(t)$

$$\phi = \varphi(x(t_1), t_1) + v\psi(x(t_1), t_1). \tag{20}$$

Then, the above general cost function will be simplified as follows:

$$J(\mu(t), \nu(t)) = \phi(x(t_1), t_1) + \int_{t_0}^{t_f} L(x(t), \mu(t), \nu(t), t)\mathrm{d}t, \tag{21}$$

where $L(x(t), \mu(t), \nu(t), t)$ corresponds to the integral term in (10) and $\phi(x(t_1), t_1)$ corresponds to the nonintegral term in (10).

According to the definition of Hamiltonian function in differential game, we have the following formula:

$$\begin{aligned} H(\lambda(t), x(t), \mu(t), \nu(t), t) &\triangleq \sum_{i=0}^{5} \lambda_i(t) f_i(x(t), \mu(t), \nu(t), t) \\ &= \sum_{i=1}^{5} \lambda_i(t) f_i(x(t), \mu(t), \nu(t), t) + L(x(t), \mu(t), \nu(t)), \end{aligned} \tag{22}$$

where $\lambda(t)$ is the set of costate variables; that is, $\lambda(t) = \{\lambda_S(t), \lambda_I(t), \lambda_L(t), \lambda_R(t), \lambda_D(t)\}$, and $f_i(x(t), \mu(t), \nu(t), t)$ is the differential equation of node state corresponding to (5)–(9).

Among them, when the costate functions satisfy the following equations, there exists an optimal strategy (μ^*, ν^*):

$$\begin{aligned} \frac{\partial \lambda_i}{\partial t} &= -\frac{\partial H}{\partial x_i}, \\ \lambda_i(t_1) &= -\frac{\partial \phi}{\partial x_i(t_1)}, \end{aligned} \tag{23}$$

where t_1 represents the terminal moment when the game ends.

Therefore, the Hamiltonian function, the differential equations, and the end-value constraints of costate variables in this paper can be formulated from (24) to (30):

$$\begin{aligned} H(\lambda(t), x(t), \mu(t), \nu(t), t) = {} & \lambda_S(t)\frac{\mathrm{d}S(t)}{\mathrm{d}t} + \lambda_I(t)\frac{\mathrm{d}I(t)}{\mathrm{d}t} + \lambda_R(t)\frac{\mathrm{d}R(t)}{\mathrm{d}t} + \lambda_L(t)\frac{\mathrm{d}L(t)}{\mathrm{d}t} + \lambda_D(t)\frac{\mathrm{d}D(t)}{\mathrm{d}t} \\ & + rC_N S(t)\left[1 - \frac{S(t)}{k}\right] + C_I I(t) + C_{SR}S(t)\frac{D_{SR}S_{SR}}{D_{SR\max} + D_{SR\min}} + C_L L(t) \\ & + C_{IR}I(t)\frac{D_{IR}S_{IR}}{D_{IR\max} + D_{IR\min}} + C_D D(t), \end{aligned} \tag{24}$$

$$\begin{aligned} \frac{\mathrm{d}\lambda_S(t)}{\mathrm{d}t} = {} & \frac{2\lambda_S(t)S(t)}{k} + (\lambda_S(t) - \lambda_I(t))A_{SI}(t)(1 - (1 - P_{SI}))^{nI(t)} - \lambda_S(t)r \\ & + (\lambda_S(t) - \lambda_R(t))\frac{D_{SR}(t)S_{SR}}{D_{SR\max} + D_{SR\min}} + (\lambda_S(t) - \lambda_L(t))P_{SL} - rC_{N(t)} \\ & + \frac{2rC_N S(t)}{k} - C_{SR}\frac{D_{SR}(t)S_{SR}}{D_{SR\max} + D_{SR\min}}, \end{aligned} \tag{25}$$

$$\begin{aligned} \frac{\mathrm{d}\lambda_I(t)}{\mathrm{d}t} = {} & nA_{SI}(\lambda_I(t) - \lambda_S(t))S(t)(1 - P_{SI})^{nI(t)}\ln(1 - P_{SI}) - C_I - C_{IR}\frac{D_{IR}(t)S_{IR}}{D_{IR\max} + D_{IR\min}} \\ & + (\lambda_I(t) - \lambda_R(t))\frac{D_{IR}(t)S_{IR}}{D_{IR\max} + D_{IR\min}} + (\lambda_I(t) - \lambda_D(t))\frac{A_{ID}(t)S_{ID}}{A_{ID\max} + A_{ID\min}} \\ & + (\lambda_I(t) - \lambda_L(t))\frac{A_{IL}(t)S_{IL}}{A_{IL\max} + A_{IL\min}}, \end{aligned} \tag{26}$$

$$\begin{aligned} \frac{\mathrm{d}\lambda_L(t)}{\mathrm{d}t} = {} & (\lambda_L(t) - \lambda_D(t))P_{LD} + (\lambda_L(t) - \lambda_R(t))P_{LR}P(t) \\ & + (\lambda_L(t) - \lambda_S(t))P_{LS}P(t) + (\lambda_L(t) - \lambda_I(t))P_{LI}P(t) - C_L, \end{aligned} \tag{27}$$

$$\frac{d\lambda_R(t)}{dt} = (\lambda_R(t) - \lambda_I(t))\frac{A_{RI}(t)S_{RI}}{A_{RI\max} + A_{RI\min}} + (\lambda_R(t) - \lambda_L(t))P_{RL}, \tag{28}$$

$$\frac{d\lambda_D(t)}{dt} = -C_D, \tag{29}$$

$$\begin{cases} \lambda_S(t_f) = \dfrac{d\phi}{dS(t)} = C_{S_f}, \\ \lambda_I(t_f) = \dfrac{d\phi}{dI(t)} = C_{I_f}, \\ \lambda_L(t_f) = \dfrac{d\phi}{dL(t)} = C_{L_f}, \\ \lambda_R(t_f) = \dfrac{d\phi}{dR(t)} = C_{R_f}, \\ \lambda_D(t_f) = \dfrac{d\phi}{dD(t)} = C_{D_f}. \end{cases} \tag{30}$$

When $H(x(t), \mu^*(t), \nu(t), t) \le H(x(t), \mu^*(t), \nu^*(t), t)$ is satisfied, if $(\lambda_I(t) - \lambda_S(t))[1 - (1 - P_{SI})^{nI^*(t)}]S^*(t)$ is greater than 0, $A_{SI}(t)$ takes the maximum value, and if $(\lambda_I(t) - \lambda_S(t))[1 - (1 - P_{SI})^{nI^*(t)}]S^*(t)$ is less than 0, $A_{SI}(t)$ takes the minimum value; if $(S_{IL}(\lambda_L(t) - \lambda_I(t))I^*(t)/(A_{IL\max} + A_{IL\min}))$ is greater than 0, $A_{IL}(t)$ takes the maximum value, and if $(S_{IL}(\lambda_L(t) - \lambda_I(t))I^*(t)/(A_{IL\max} + A_{IL\min}))$ is less than 0, $A_{IL}(t)$ takes the minimum value; if $(S_{ID}(\lambda_D(t) - \lambda_I(t))I^*(t)/(A_{ID\max} + A_{ID\min}))$ is greater than 0, $A_{ID}(t)$ takes the maximum value, and if $(S_{ID}(\lambda_D(t) - \lambda_I(t))I^*(t)/(A_{ID\max} + A_{ID\min}))$ is less than 0, $A_{ID}(t)$ takes the minimum value; if $(S_{RI}(\lambda_I(t) - \lambda_R(t))R^*(t)/(A_{RI\max} + A_{RI\min}))$ is greater than 0, $A_{RI}(t)$ takes the maximum value, and if $(S_{RI}(\lambda_I(t) - \lambda_R(t))R^*(t)/(A_{RI\max} + A_{RI\min}))$ is less than 0, $A_{RI}(t)$ takes the minimum value. On the contrary, when $H(x(t), \mu^*(t), \nu^*(t), t) \le H(x(t), \mu, \nu^*(t), t)$ is to be satisfied, if $(S_{SR}(\lambda_R(t) - \lambda_S(t) + C_{SR})S^*(t)/(D_{SR\max} + D_{SR\min}))$ is greater than 0, $D_{SR}(t)$ chooses the minimum value, and if $(S_{SR}(\lambda_R(t) - \lambda_S(t) + C_{SR})S^*(t)/(D_{SR\max} + D_{SR\min}))$ is less than 0, $D_{SR}(t)$ chooses the maximum value; if $(S_{IR}(\lambda_R(t) - \lambda_I(t) + C_{IR})I^*(t)/(D_{IR\max} + D_{IR\min}))$ is greater than 0, $D_{IR}(t)$ chooses the minimum value, and if $(S_{IR}(\lambda_R(t) - \lambda_I(t) + C_{IR})I^*(t)/(D_{IR\max} + D_{IR\min}))$ is less than 0, $D_{IR}(t)$ chooses the maximum value.

4. Simulation

In this section, we will expand into three parts. The first part is to compare with the existing general epidemic models in turn. The second part is to analyze the impact of charging on the SILRD model. The third part is to discuss the impact of controllable and uncontrollable system input and node degree on the ΛSILRD model. In all three parts, the simulations are implemented in MATLAB R2017b. The abbreviations are applied in the section list in Table 2.

4.1. Comparison with General Epidemic Model. In this part, three general epidemic models will be compared, namely, Susceptible-Infected-Recovered (SIR) model [39], Susceptible-Exposed-Infected-Recovered (SEIR) model [40], and EiSIRS model [41]. Among the three models, SIR model is the basic, SEIR model extends the E state on the basis of the SIR model, and EiSIRS adds the corresponding sleeping state on the basis of the SIR model.

For the unification and reasonability of the analysis, we did not consider the multiple rounds of infection in the EiSIRS model, and EiSIRS model would be renamed as Susceptible-Susceptible & sleep-Infected-Infected & sleep-Recovered-Recovered & sleep-Dead (SsIiRrD) model to facilitate understanding, where lowercase letters represent the sleep state of the corresponding state.

Experimental parameters are set as follows: $P_{SI} = 0.1$, $P_{SR} = 0.4$, $P_{SD} = 0.0008$, $P_{ID} = 0.005$, $P_{IR} = 0.21$, $P_{RD} = 0.008$, $P_{EI} = 0.005$, $P_{ER} = 0.21$, $P_{ED} = 0.005$, $P_{Ss} = 0.006$, $P_{sS} = 0.006$, $P_{Ii} = 0.006$, $P_{iI} = 0.009$, $P_{Rr} = 0.006$, $P_{rR} = 0.006$, $P_{SL} = 0.0008$, $P_{IL} = 0.001$, $P_{RL} = 0.0008$, $P_{LD} = 0.3$, $P_{LR} = 0.6$.

Three general epidemic models have the same parameter settings except for their own defensive measures. Similarly, the SILRD with UAVs and the SILRD with solar energy harvesters have the same parameter settings except for the introduction of the L state and the corresponding defensive measures. In particular, the difference between the two SILRD models lies in the different charging methods. The first method is to use energy harvesters to capture solar energy and convert light energy into electrical energy to

Table 2: Table of abbreviations.

Abbreviation	Full name
EHWSNs	Energy-harvesting wireless sensor networks
WSNs	Wireless sensor networks
SILRD	Susceptible-Infected-Low (energy)-Recovered-Dead
ΛSILRD	Λ-Susceptible-Infected-Low (energy)-Recovered-Dead
UAVs	Unmanned aerial vehicles
SIR	Susceptible-Infected-Recovered
SEIR	Susceptible-Exposed-Infected-Recovered
SsIiRrD	Susceptible-Susceptible & sleep-Infected-Infected & sleep-Recovered-Recovered & sleep-Dead

supplement the energy of sensor nodes. The second method is to deploy UAVs to charge sensor nodes [21].

It is worth noting that the purpose of this part is to highlight the characteristics of the two SILRD model by comparing with other general epidemic models, so the system input, multiple types of malicious programs, and the nonlinear infection rate will be ignored. Figure 3 shows the evolution of sensor node under five epidemic models.

It can be seen from Figure 3(a) that the changes in the quantity of susceptible sensor nodes in the five models are very close. Except for SEIR, the other models had a high infection rate in the first few days, as depicted in Figure 3(b). Because the SEIR model exists an exposed state between the susceptibility and infection state, some infected sensor nodes were cleared during the exposure period. For recovered sensor nodes, the decline was more pronounced in SsIiRrD, followed by SEIR and SIR, as depicted in Figure 3(c). Among them, the quantity of recovered sensor nodes in SIRLD model decreased the most slowly and stay around 96% after 20 days. As shown in Figure 3(d), the order of increasing quantity of dead sensor nodes from fast to slow is SIR, SEIR, SsIiRrD, SILRD with UAVs, and SILRD with solar energy harvesters.

It can be seen from the comparison with other general epidemic models that the SILRD model can more effectively increase the quantity of recovered sensor nodes and reduce the quantity of dead sensor nodes. The phenomenon is more obvious in SILRD with solar energy harvesters. At the same time, the two SILRD model directly charges low-energy sensor nodes, which will effectively reduce the energy depletion of sensor nodes due to infections or daily work.

4.2. Effect of Charging on SILRD Model. Charging factors as one of the features of SILRD model will be discussed here. In this part, variations in the quantity of five node states, control variables and the quantity of high- and low-energy nodes, and the overall costs will be applied as indicators to explain the impact of charging.

Three scenarios will be discussed here, namely, SILRD model with solar energy harvesters, SILRD model with UAVs [21], and SILRD model without charging capability. In order to facilitate the analysis of the impact of charging, this part will ignore the impact of system input but will consider multiple types of malicious programs' attacks and nonlinear infection rates.

Unlike the previous section, the simulation here only considers one day, so the relevant simulation parameters have also been modified. Experimental parameters are set as follows: $S_{SI} = 0.005, S_{SR} = 0.05$, $S_{IR} = 0.05, S_{IL} = 0.001$, $S_{ID} = 0.005, S_{RI} = 0.005$, $P_{SL} = 0.0008, P_{LD} = 0.0016$, $P_{RL} = 0.0008, C_{SR} = 5$, $C_{IR} = 7, C_I = 10$, $C_L = 12, C_D = 20$, $A = 8, a = 0.5$, $b = 0.5, m = 12$, $n = 3$, and $C_N = 50$.

4.2.1. Evolution of Sensor Node under Various Charging Strategies. The solar charging power is formulated in (1). It is worth noting that SILRD with UAVs considers the situation of patching and charging at the same time, so low-energy sensor nodes will be transformed to recovered state directly. Figure 4 shows the evolution of sensor nodes under three charging strategies.

As can be seen from Figure 4(a), SILRD with solar energy harvesters can effectively increase the quantity of susceptible sensor nodes. However, under the attack of multiple types of malicious programs, the SILRD model with charging strategy cannot effectively restrain the growth of malicious programs, among which the case with solar charging is the most serious, as shown in Figure 4(b). Nevertheless, charging can effectively reduce the quantity of low-energy sensor nodes, as shown in Figure 4(c). Similarly, the quantity of recovered sensor nodes also increased due to the charging strategies, as depicted in Figure 4(d). The situation of energy depletion is very close as shown in Figure 4(e). Among them, the more accurate sorting from high to low should be SILRD with UAVs, followed by SILRD without charging and SILRD with solar energy harvesters.

Under the attack of multitype of malicious programs, the strategies with charging cannot inhibit the increase of the quantity of infected sensor nodes effectively, but it can greatly reduce the quantity of low-energy sensor nodes so as to increase the quantity of recovered sensor nodes. The strategy with solar charging is more widely distributed, so it can increase the quantity of recovered sensor nodes in highly efficient to keep the networks running well.

4.2.2. Variation on Dynamic Control Level. The variation of dynamic control will further reveal the cause of the evolution of node state, as depicted in Figure 5. Specifically, Figure 5(a) shows the changes in control variables in SILRD with solar energy harvesters, Figure 5(b) shows the changes in control

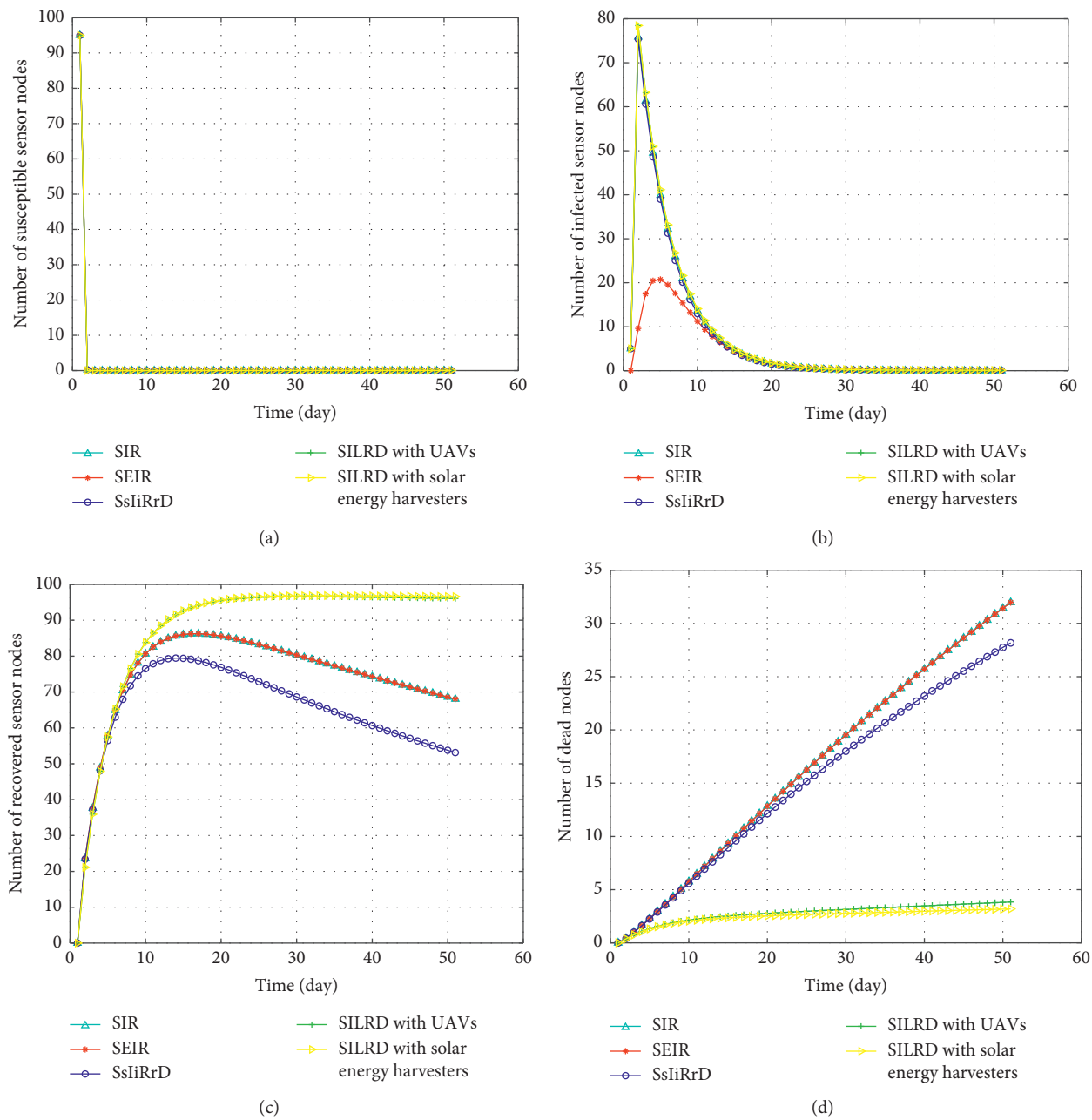


Figure 3: Evolution of node states under 5 epidemic models. (a) Susceptible state; (b) infected state; (c) recovered state; (d) dead state.

variables in SILRD with UAVs, and Figure 5(c) shows the changes in control variables in SILRD without charging.

In all three cases, the malicious programs stopped spreading at the beginning and peaked when $t = 2$. After propagation stops, the malicious programs still exist in the infected sensor nodes. After patching with maximum effort, due to the accumulation of costs, the networks stop patching after weighing. If the networks were patched again, the cost of patching would be higher than the cost of damage caused by malicious programs, so the networks stopped using the UAVs.

After UAVs stop patching, there are still exist malicious programs with strong and weak ability to destroy in the networks. Among them, the strategy with UAVs can quickly eliminate the malicious programs with weak damage ability ($t = 2$), followed by the solar charging strategy ($t = 4$) and finally the noncharging strategy ($t = 15$). However, malicious programs with strong destructive ability still present in the networks without completely clearing away.

4.2.3. Variation on the Quantity of High- and Low-Energy Nodes. In order to directly express the quantity of high- and low-energy sensor nodes in the networks, the form of histogram has been applied to show the variation on the quantity of high- and low-energy sensor nodes over time under different charging strategies, as depicted in Figure 6.

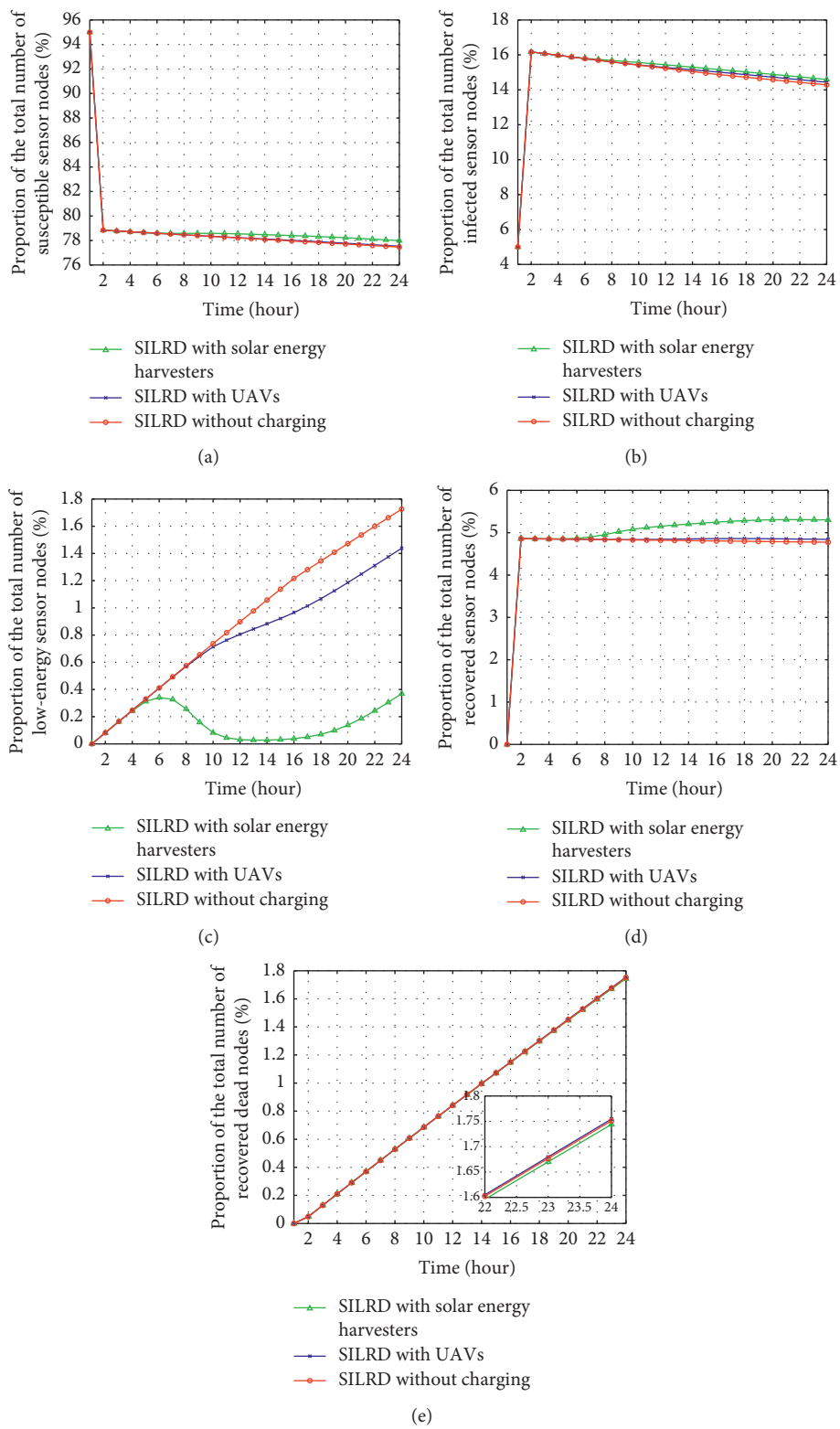


Figure 4: Evolution of sensor nodes under 3 charging strategies. (a) Susceptible state; (b) infected state; (c) low-energy state; (d) recovered state; (e) dead state.

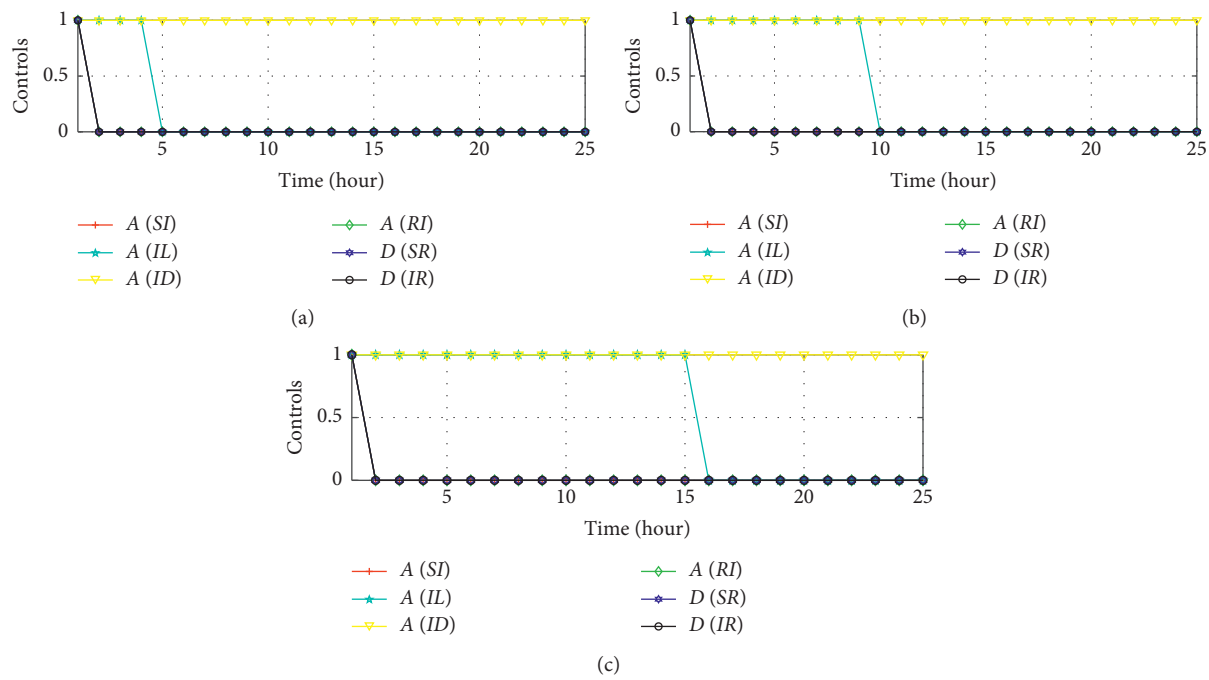


Figure 5: Dynamic control under 3 charging strategies. (a) Strategy with solar energy harvesters; (b) strategy with UAVs; (c) strategies without charging.

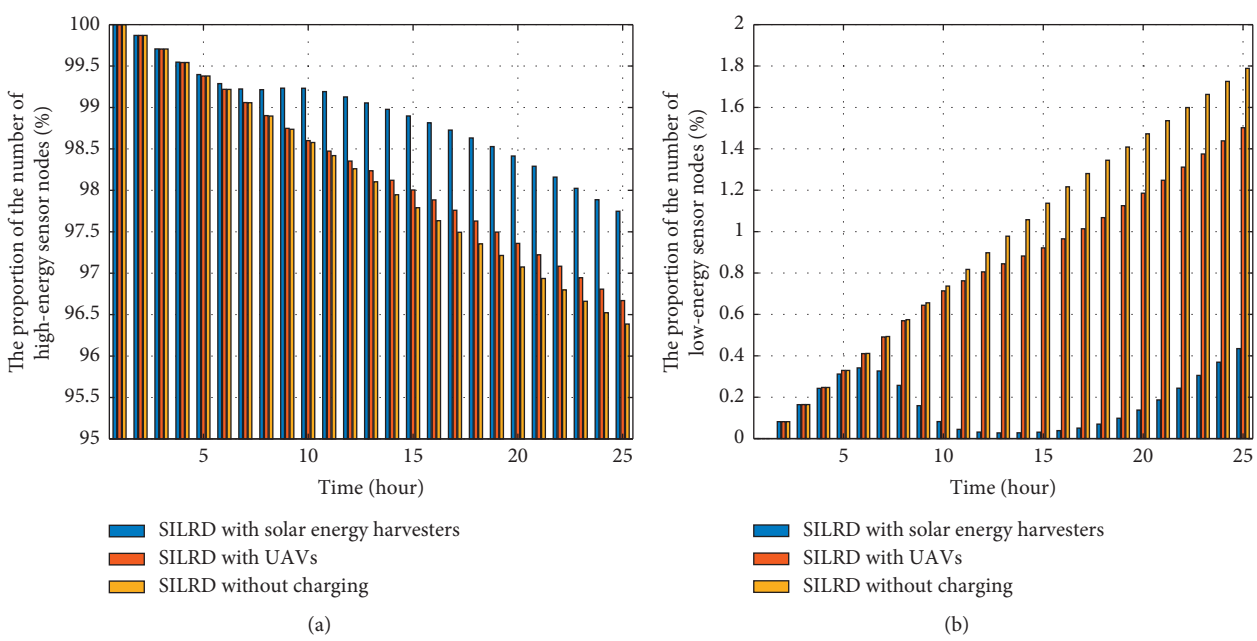


Figure 6: Variations on the quantity of high- and low-energy sensor nodes. (a) High-energy sensor nodes; (b) low-energy sensor nodes.

As can be seen from Figure 6(a), sensor nodes with high energy show a downward trend under the three strategies. Among them, the strategy without charging fell by the most quickly. The degree of elevation of the strategy with UAVs is determined by the quantity of UAVs. This paper assumes a small quantity of UAVs deployed because too many deployments would be costly. The strategy with solar energy harvesters assumes that each sensor node is equipped with the energy harvester, which is close to reality. Based on the analysis of Figures 6(a) and 6(b), it can be seen that, due to the comprehensive deployment of sensor nodes equipped with solar energy harvesters, the quantity of high-energy

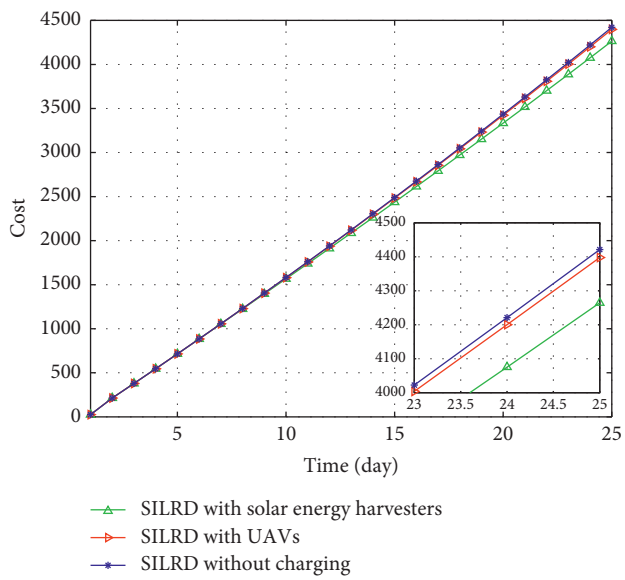


Figure 7: Overall costs under three charging strategies.

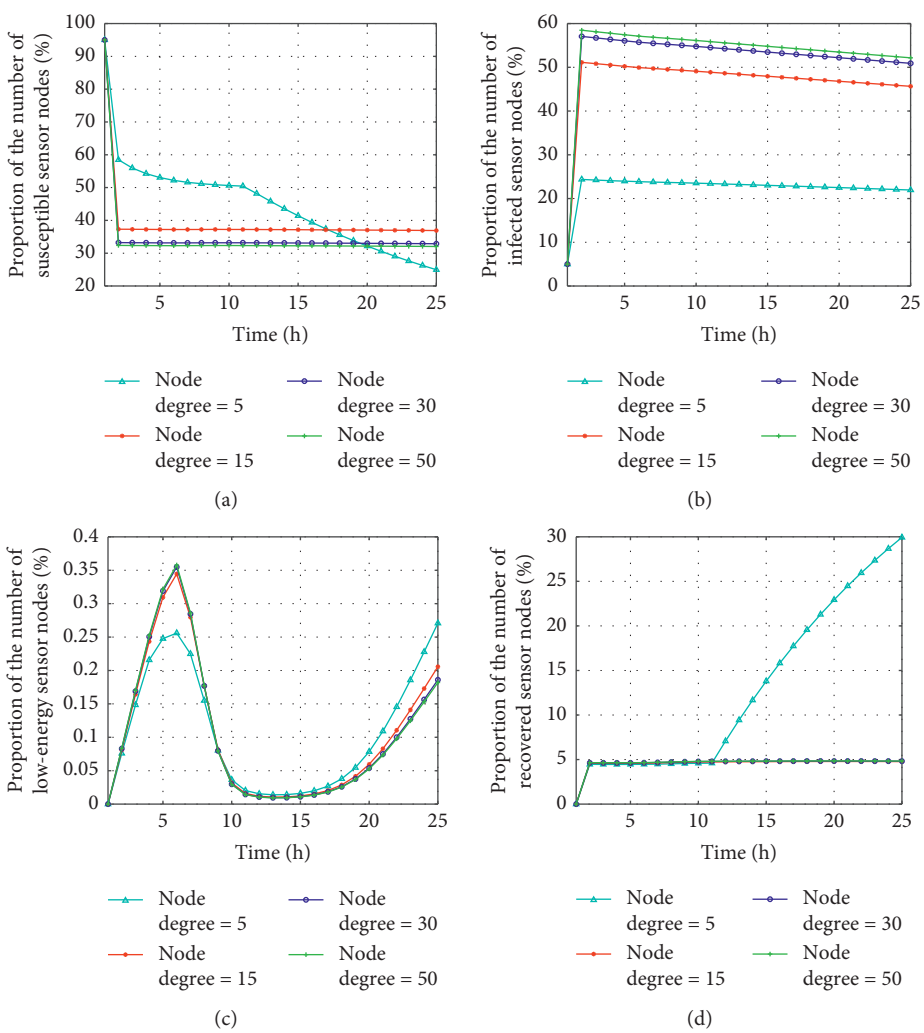


Figure 8: Continued.

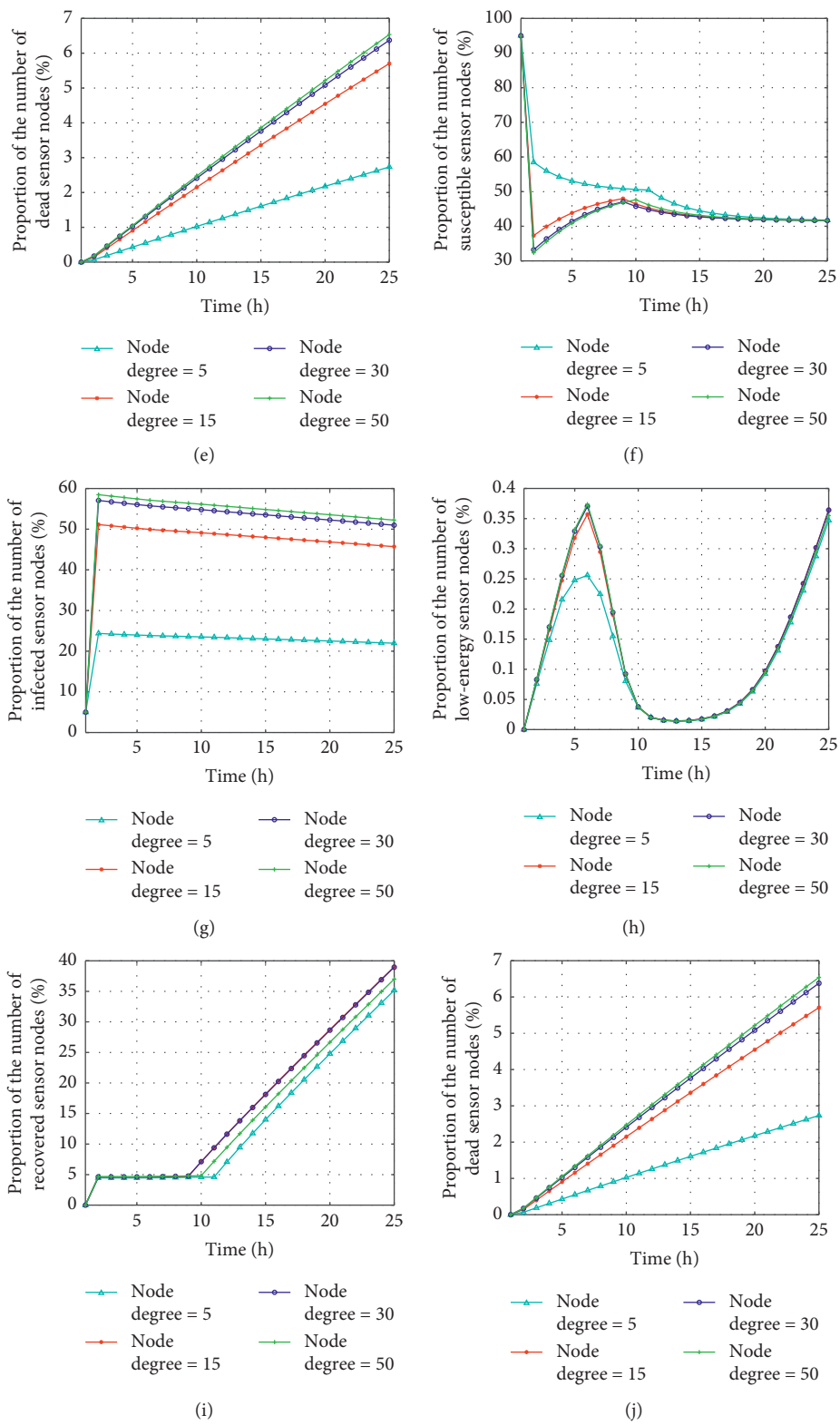


Figure 8: Evolution of sensor nodes with Logistic Growth in susceptible sensor nodes. (a) The quantity of susceptible sensor nodes with controllable input; (b) the quantity of infected sensor nodes with controllable input; (c) the quantity of low-energy sensor nodes with controllable input; (d) the quantity of recovered sensor nodes with controllable input; (e) the quantity of dead sensor nodes with controllable input; (f) the quantity of susceptible sensor nodes with uncontrollable input; (g) the quantity of infected sensor nodes with uncontrollable input; (h) the quantity of low-energy sensor nodes with uncontrollable input; (i) the quantity of recovered sensor nodes with uncontrollable input; (j) the quantity of dead sensor nodes with uncontrollable input.

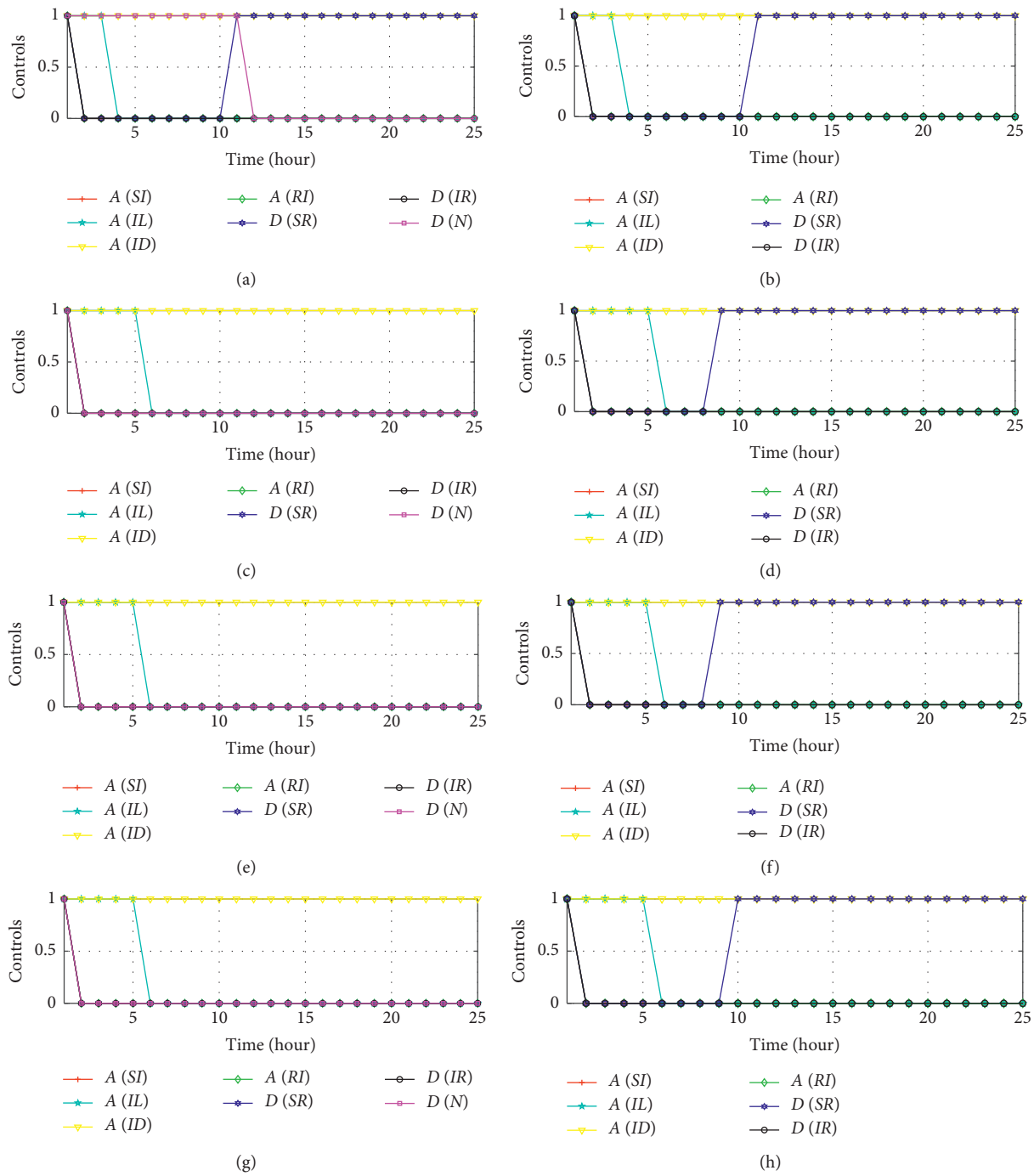


Figure 9: Variation of dynamic control variables. (a) Controllable input with node degree = 5; (b) controllable input with node degree = 15; (c) controllable input with node degree = 30; (d) controllable input with node degree = 50; (e) uncontrollable input with node degree = 5; (f) uncontrollable input with node degree = 15; (g) uncontrollable input with node degree = 30; (h) uncontrollable input with node degree = 50.

sensor nodes declines slowly. Moreover, the low-energy sensor nodes not only did not rise but also fell.

4.2.4. Overall Cost under Various Charging Strategies. Figure 7 shows the cost trends of the three strategies. The cost of solar charging is the lowest, followed by the strategy of deploying UAVs and finally the strategy of noncharging. For solar charging, since each sensor node is equipped with a solar energy harvester, there is no additional cost during capturing solar energy compared to the deployment of UAVs. In the conclusion of the previous section, charging can effectively improve the immunity of EHWSNs and reduce the quantity of dead sensor nodes, so as to achieve the purpose of reducing costs.

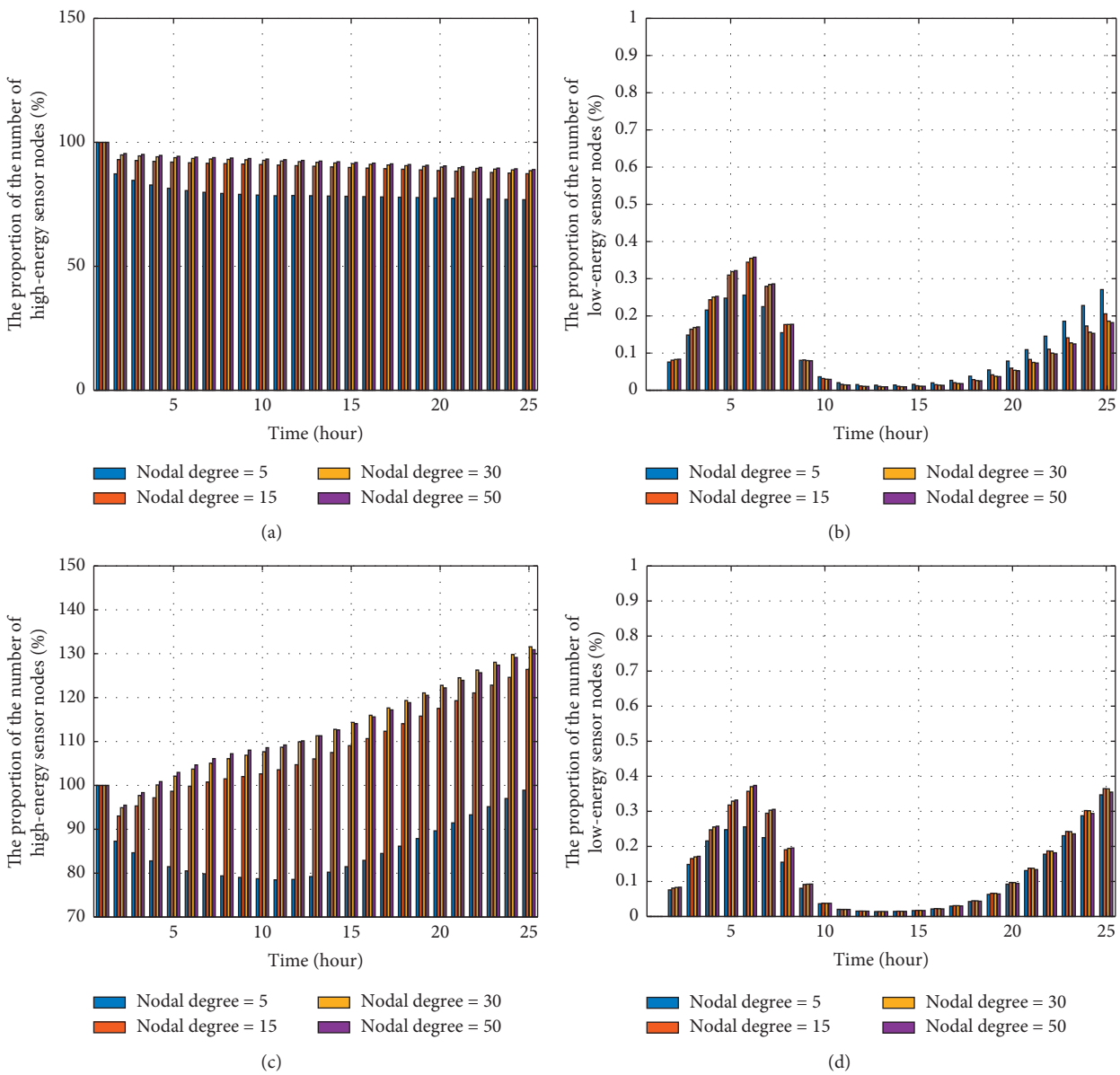


Figure 10: Variations in the quantity of high- and low-energy sensor nodes. (a) Variation of the quantity of high-energy sensor nodes with controllable input; (b) variation of the quantity of low-energy sensor nodes with controllable input; (c) variation of the quantity of high-energy sensor nodes with uncontrollable input; (d) variation of the quantity of low- energy sensor nodes with uncontrollable input.

4.3. Influence of Networks Input and Node Degree on ΛSILRD Model. This section mainly discusses the influence of controllable and uncontrollable input and node degree on ΛSILRD model. The formulation of networks input adopts (17). Node degrees are set to 5, 15, 30, and 50, respectively.

Similar to the previous section, this section also discusses the variation trend of four aspects, which are sensor nodes in various states, control variables, quantity of high- and low-energy sensor nodes, and overall costs.

The experimental parameters are the same as those in Section 4.2.

4.3.1. Evolution of Sensor Nodes in ΛSILRD Model. As a control group, uncontrollable input into the system will be considered. Figures 8(a)–8(e) show the evolution of node state when networks input contains control variables, and Figures 8(f)–8(j) show the evolution with uncontrollable input.

In both cases, evolutions of sensor nodes change uniformly except when node degree equals 5, as depicted in Figure 8. The quantity of susceptible sensor nodes fell rapidly in the first hour before reaching equilibrium, as depicted in Figures 8(a) and 8(f). In the case of input with control variables, the final stable value of the quantity of susceptible

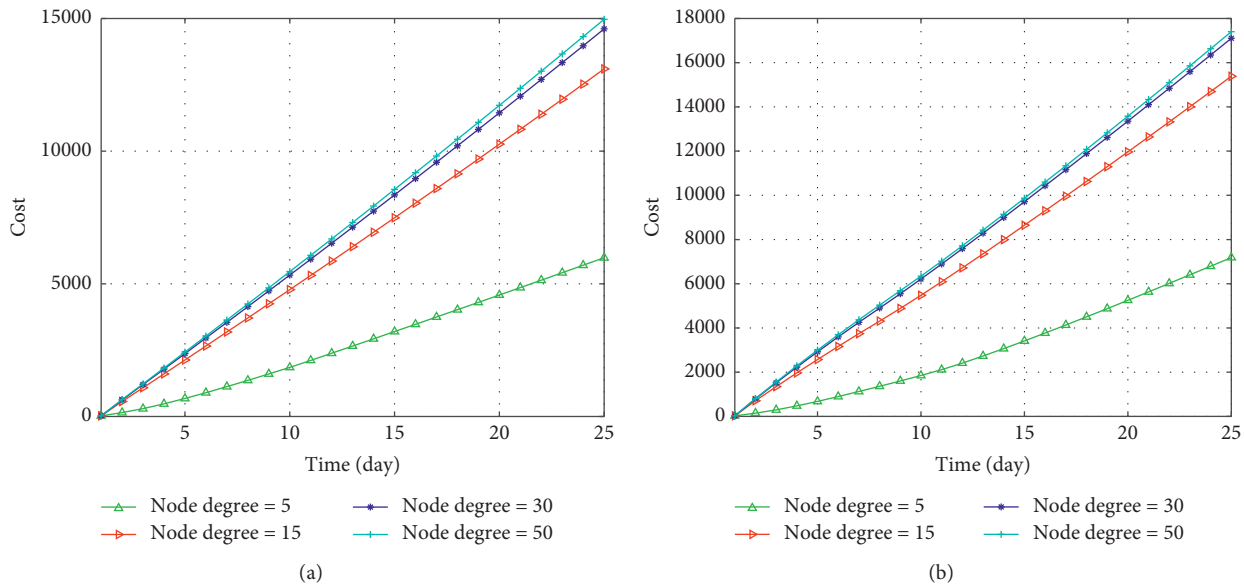


Figure 11: Overall cost with Logistic Growth in susceptible sensor nodes under different node degrees. (a) Strategy with controllable input; (b) strategy with uncontrollable input.

sensor nodes is related to node degree, as shown in Figure 8(a). With the increase of node degree, the stable value of the quantity of susceptible sensor nodes becomes lower. On the contrary, in the case of input without control variables, the quantity of susceptible nodes experienced a process of rapid decrease and then a slow rise and finally reached an equilibrium, as shown in Figure 8(f).

The changes of the quantity of infected sensor nodes, low-energy sensor nodes, and dead sensor nodes are very similar, as shown in Figures 8(b), 8(c), 8(e), 8(g), 8(h), and 8(j). In the case of input with control variables, the quantity of recovered sensor nodes remained stable except when the node degree equals 5, as shown in Figure 8(d). In the case of input without control variables, the quantity of recovered sensor nodes increased rapidly after a period of stabilization, as shown in Figure 8(i). In a word, as the node degree increases, the infection will become more severe, and the quantity of low-energy and death sensor nodes will increase.

4.3.2. Variation of Control Variables. In both cases, the quantity of infected sensor nodes reaches its peak when $t = 1$, so the networks stop patching, as shown in Figure 9. When $t = 10$, as more vulnerable sensor nodes exist in the networks, there is a risk of reinfection by malicious programs, so the networks start to patch the vulnerable sensor nodes again, as depicted in Figures 9(a), 9(b), 9(d), 9(f), and 9(h). When $t = 11$, the existing quantity of vulnerable sensor nodes is enough to maintain the normal operation of the networks, and the deployment of new sensor nodes will only bring more extra burden, so the networks will stop casting new sensor nodes, as shown in Figure 9(a). With the increase of node degree, the control strategy will not change greatly, as shown in Figures 9(c)–9(h).

4.3.3. Variation on the Quantity of High- and Low-Energy Sensor Nodes. Figure 10 shows the changing trend of the quantity of high- and low-energy sensor nodes in the networks. The variation trend of low-energy sensor nodes is the same in both cases, but the influence of node degree on both cases is different when T is greater than 16, as shown in Figures 10(b) and 10(d). As can be seen from Figure 10(a), when network input contains control variables, the quantity of high-energy sensor nodes basically remains at a very high level, about 90%. In the case of uncontrollable input, the quantity of high-energy sensor nodes will increase continuously, as shown in Figure 10(c). It is worth noting that, with the increase of node degree, the quantity of high-energy sensor nodes will increase. As the infection rate continues to rise, sensor nodes in susceptible state will quickly convert to infected state, so that the quantity of susceptible sensor nodes will rapidly decline. Therefore, it is necessary to quickly cast new sensor nodes to maintain the operation of the networks.

4.3.4. Overall Cost. Figure 11 shows the cumulative cost in both cases. With the increase of node degree, the costs do not show the same growth trend but tend to be saturated. In Figure 11(a), the numerical difference between the cost of node degree equal to 5 and the cost of node degree equal to 15 is about 7000, but the difference between 15 and 30 is about 1000, and the difference between 30 and 50 is about 200. Figure 11(b) shows the same phenomenon. The costs are lower in the case of controllable input than uncontrollable input under the same node degree owing to the networks with controllable input which can reduce the quantity of new sensor nodes and cut the cost of patching new sensor nodes, as shown in Figures 11(a) and 11(b).

5. Conclusion

By introducing Logistic Growth, nonlinear incidence, and charging by solar energy, this paper builds an ΛSILRD model suitable for EHWSNs. At the same time, the introduction of multiple types of malicious programs refines the model. By comparing with the existing epidemic models, we found that the SILRD has obvious advantages in increasing the quantity of recovered sensor nodes and reducing the quantity of death sensor nodes, especially SILRD with solar charging. Meanwhile, compared with the three charging strategies, we found that the SILRD with solar charging has the lowest cost. Finally, the influence of controllable input, uncontrollable input, and node degree on ΛSILRD model is revealed through the simulations. When the node degree is higher, the quantity of infected sensor nodes and dead sensor nodes will increase rapidly under the attack of multitype malicious programs with nonlinear infection rate but will tend to be saturated. At the same time, input that contains control variables can timely stop the delivery of new nodes and affect the subsequent network patching behavior, thereby reducing costs.

Although this paper proposes a malicious programs' propagation model that is close to reality, there are still many deficiencies. In the establishment of the solar charging model, this paper uses a simplified model and does not consider some random factors, like weather factors, human factors, and so on. At the same time, the topology of the EHWSNs and various delay phenomena are not further analyzed in this paper. However, in spite of this, the model and analytical methods proposed in this paper are believed to provide scholars in related fields some inspiration in the future.

Data Availability

The data used to support the findings of this study are included within the article, such as the coverage area of the WSNs, the maximum transmission radius of nodes, and transition probabilities among five nodal states.

Disclosure

All authors declare that (1) no support, financial or otherwise, has been received from any organization that may have an interest in the submitted work and (2) there are no other relationships or activities that could appear to have influenced the submitted work.

Conflicts of Interest

The authors have no conflicts of interest, financial or otherwise.

References

[1] W. O. Kermack and A. G. McKendrick, "Contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London Series A*, vol. 115, pp. 700–721, 1927.

[2] A. Singh, A. K. Awasthi, K. Singh, and P. K. Srivastava, "Modeling and analysis of worm propagation in wireless sensor networks," *Wireless Personal Communications*, vol. 98, no. 3, pp. 2535–2551, 2018.

[3] M. S. Haghighi, S. Wen, Y. Xiang, B. Quinn, and W. L. Zhou, "On the race of worms and patches: modeling the spread of information in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2854–2865, 2016.

[4] R. K. Shakya, "Modified SI epidemic model for combating virus spread in spatially correlated wireless sensor networks," 2018, https://arxiv.org/pdf/1801.04744.pdf.

[5] T. Wang, Y. Z. Liang, Y. Yang et al., "An intelligent edge computing-based method to counter coupling problems in cyber-physical systems," *IEEE Network*, vol. 34, no. 3, pp. 16–22, 2020.

[6] N. Keshri and B. K. Mishra, "Two time-delay dynamic model on the transmission of malicious signals in wireless sensor network," *Chaos, Solitons & Fractals*, vol. 68, pp. 151–158, 2014.

[7] Y. Wu, H. Huang, Q. Wu, A. Liu, and T. Wang, "A risk defense method based on microscopic state prediction with partial information observations in social networks," *Journal of Parallel and Distributed Computing*, vol. 131, pp. 189–199, 2019.

[8] T. Wang, D. Zhao, S. B. Cai, W. J. Jia, and A. F. Liu, "Bi-directional prediction-based underwater data collection protocol for end-edge-cloud orchestrated system," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4761–4799, 2020.

[9] Z. Li, W. Li, F. Lin et al., "Hybrid malware detection approach with feedback-directed machine learning," *Science China Information Sciences*, vol. 63, no. 3, Article ID 139103, 2020.

[10] G. Han, H. Wang, X. Miao, L. Liu, J. Jiang, and Y. Peng, "A dynamic multipath scheme for protecting source-location privacy using multiple sinks in WSNs intended for IIoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5527–5538, 2020.

[11] Z. G. Zhao, Y. M. Huang, Z. Y. Zhen, and Y. Z. Li, "Data-driven false data-injection attack design and detection in cyber-physical systems," *IEEE Transactions on Cybernetics*, 2020.

[12] X. C. Li, K. Xie, X. Wang et al., "Quick and accurate false data detection in mobile crowd sensing," in *Proceedings of the 2020 IEEE Conference on Computer Communications*, pp. 2215–2223, Paris, France, April 2020.

[13] L. Liu, G. Han, Y. He, and J. Jiang, "Fault-tolerant event region detection on trajectory pattern extraction for industrial wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2072–2080, 2020.

[14] T. Wang, Q. Wu, S. Wen et al., "Propagation modeling and defending of a mobile sensor worm in wireless sensor and actuator networks," *Sensors*, vol. 17, no. 12, pp. 139–156, 2017.

[15] R. Z. Nicola, N. Enrico, R. Giuseppe, P. Michael, and M. Antonella, "MeDrone: on the use of a medical drone to heal a sensor network infected by a malicious epidemic," *Ad Hoc Networks*, vol. 50, pp. 115–127, 2016.

[16] Y. Wu, H. Huang, N. Wu, Y. Wang, M. Z. Alam Bhuiyan, and T. Wang, "An incentive-based protection and recovery strategy for secure big data in social networks," *Information Sciences*, vol. 508, pp. 79–91, 2020.

[17] L. Mo, A. Kritikakou, and S. He, "Energy-aware multiple mobile chargers coordination for wireless rechargeable sensor networks," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8202–8214, 2019.

[18] L. Mo, P. C. You, X. H. Cao, Y. Q. Song, and J. M. Chen, "Decentralized multi-charger coordination for wireless

rechargeable sensor networks," in *Proceedings of the 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC)*, pp. 1–8, Nanjing, China, December 2015.

[19] Z. J. Ji, H. Lin, S. B. Cao, Q. Y. Qi, and H. Z. Ma, "The complexity in complete graphic characterizations of multi-agent controllability," *IEEE Transactions on Cybernetics*, 2020.

[20] L. P. Mo and S. Y. Guo, "Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback," *Journal of Systems Science and Complexity*, vol. 32, no. 3, pp. 835–845, 2019.

[21] G. Y. Liu, B. H. Peng, X. J. Zhong, and X. J. Lan, "Differential games of rechargeable wireless sensor networks against malicious programs based on SILRD propagation model," *Complexity*, vol. 2020, Article ID 5686413, 13 pages, 2020.

[22] M. H. R. Khouzani, S. Sarkar, and E. Altman, "Optimal dissemination of security patches in mobile wireless networks," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4714–4732, 2012.

[23] H. Xu and X. Zhou, "Optimal power control in cooperative relay networks based on a differential game," *ETRI Journal*, vol. 36, no. 2, pp. 280–285, 2014.

[24] M. Hosseinzadeh, B. Sinopoli, and E. Garone, "Feasibility and detection of replay attack in networked constrained cyber-physical systems," in *Proceedings of the 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 712–717, Monticello, IL, USA, September 2019.

[25] J. Hu, Q. Qian, A. Fang, S. Wu, and Y. Xie, "Optimal data transmission strategy for healthcare-based wireless sensor networks: a stochastic differential game approach," *Wireless Personal Communications*, vol. 89, no. 4, pp. 1295–1313, 2016.

[26] X. Ju, W. Liu, C. Zhang et al., "An energy conserving and transmission radius adaptive scheme to optimize performance of energy harvesting sensor networks," *Sensors*, vol. 18, no. 9, pp. 2885–2926, 2018.

[27] W. Qi, W. Liu, X. Liu et al., "Minimizing delay and transmission times with long lifetime in code dissemination scheme for high loss ratio and low duty cycle wireless sensor networks," *Sensors*, vol. 18, no. 10, pp. 3516–3560, 2018.

[28] M. Huang, W. Liu, T. Wang et al., "A game-based economic model for price decision making in cyber-physical-social systems," *IEEE Access*, vol. 7, pp. 111559–111579, 2019.

[29] J. A. Ansere, G. J. Han, L. Liu, Y. Peng, and M. Kamal, "Optimal resource allocation in energy efficient internet of things networks with imperfect CSI," *IEEE Internet Things Journal*, vol. 7, no. 6, pp. 5401–5411, 2020.

[30] J. Hu and Y. Xie, "A stochastic differential game theoretic study of multipath routing in heterogeneous wireless networks," *Wireless Personal Communications*, vol. 80, no. 3, pp. 971–991, 2015.

[31] T. Mylvaganam, M. Sassano, and A. Astolfi, "A differential game approach to multi-agent collision avoidance," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 4229–4235, 2017.

[32] L. Miao and S. Li, "A differential game-theoretic approach for the intrusion prevention systems and attackers in wireless networks," *Wireless Personal Communications*, vol. 103, no. 3, pp. 1993–2003, 2018.

[33] X.-N. Miao, X.-W. Zhou, and H.-Y. Wu, "A cooperative differential game model based on network throughput and energy efficiency in wireless networks," *Optimization Letters*, vol. 6, no. 1, pp. 55–68, 2012.

[34] S. Climent, A. Sanchez, S. Blanc, J. V. Capella, and R. Ors, "Wireless sensor network with energy harvesting: modeling and simulation based on a practical architecture using real radiation levels," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 6, pp. 1812–1830, 2016.

[35] M. H. R. Khouzani, S. Sarkar, and E. Altman, "Maximum damage battery depletion attack in mobile sensor networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2358–2368, 2011.

[36] F. Avner, "Differential games," in *Handbook of Game Theory*, pp. 781–799, Elsevier, Amsterdam, Netherlands, 1994.

[37] A. Bressan, "Noncooperative differential games," *Milan Journal of Mathematics*, vol. 79, no. 2, pp. 357–427, 2011.

[38] R. Isaacs, *Differential Game*, John Wiley and Sons, New York, NY, USA, 1965.

[39] J. C. Frauenthal, *Mathematical Modeling in Epidemiology*, Springer, New York, NY, USA, 1981.

[40] J. E. Cohen and E. Joel, "Infectious diseases of humans: dynamics and control," *JAMA: The Journal of the American Medical Association*, vol. 268, no. 23, p. 3381, 1992.

[41] X. Wang, Q. Li, and Y. Li, "EiSIRs: a formal model to analyze the dynamics of worm propagation in wireless sensor networks," *Journal of Combinatorial Optimization*, vol. 20, no. 1, pp. 47–62, 2010.