Cooperative Differential Game-Based Optimal Control and Its Application to Power Systems

Chaoxu Mu, *Senior Member, IEEE*, Ke Wang, Zhen Ni, *Member, IEEE*, and Changyin Sun

***Abstract*—Differential games have been extensively applied to optimal control problems. Nash equilibrium captures the tradeoff among players' policies when every player independently tries to minimize a predefined index. When considering potential cooperation, Pareto equilibrium plays an important role in cooperative differential games. This article studies the cooperative control of multiplayer systems on the quadratic infinite horizon. First, by defining a joint cost function using a parameter set, a cooperative differential game is reformulated as a general optimal control problem, where all players form a grand coalition. Then, the joint cost function is approximated by a critic neural network, and for the first time, a novel adaptive dynamic programming algorithm with two learning stages is proposed to determine the parameter selection and then obtain Pareto optimal solutions. A numerical example demonstrates that this algorithm can achieve optimal policies and Pareto frontier. As for its application, the cooperative control of a two-area interconnected power system is investigated, where the primary frequency control and secondary frequency control are regarded as two players. Simulation results indicate that the proposed scheme can obtain binding cooperation agreements, such that cooperative control scheme can get better overall performance compared to Nash control method and another three control methods.**

***Index Terms*—Adaptive dynamic programming, cooperative differential game, neural network, Pareto equilibrium, two-area interconnected power system.**

I. Introduction

DIFFERENTIAL game (DG) theory is concerned with the dynamic decision-making in multiplayer interactive systems [1]. From the optimal control perspective, a complete differential game consists of three elements: Players (controllers or agents), control policies, and the cost function (cost) [2]. The interest of this article lies in nonzero-sum DGs. In a noncooperative differential game (NCDG), it is desirable that all players simultaneously take policies to obtain the Nash equilibrium [3], where the outcome of one player cannot be improved through a unilateral changing. While cooperative differential games (CDGs) arise when multiple players, participating in a dynamic system, collaborate their actions with the intent to optimize their objectives [4]. At this point, all players need to minimize their own costs while considering a joint cost to achieve the overall optimality, i.e., the Pareto equilibrium. Therefore, according to [5], the CDG problem is equivalent to an optimization problem where every player has multiple objectives. In practice, CDGs can be employed to solve optimization problems of the multicontroller systems, such as power system [6] and vehicle stability system [7]. Pareto optimality plays a pivotal role in analyzing and solving these problems [5]. This article mainly focuses on handling CDG problems by the aid of the adaptive dynamic programming (ADP) methodology [8]–[11]; besides, its application in the load frequency control (LFC) problem of power systems is another interest. Before elaborating our work, the relevant results are primarily introduced from three perspectives.

First of all, some pioneering theoretical studies of cooperative games are instructive. The essence of cooperative games was early elaborated by Starr [12] and Schmitendorf [13], and they proposed that cooperative games were actually problems where one player has multiple vector-valued performance indices. Engwerda [14] profoundly studied cooperative game problems and gave sufficient and necessary conditions for the existence of a Pareto optimum. Similarly, existence conditions of Pareto optimal solutions were derived for infinite-horizon linear quadratic CDGs in [15] and [16]. Based on these studies, it can be learned that, in cooperative games, an optimal control policy namely Pareto optimal policy is related to the joint cost function. In other words, a player's control policy is not only depending on his own cost but also relying on the costs of other individuals. It can also be further known that obtaining Pareto optimal solutions can be translated into solving a parameterized optimal problem with all weight coefficients being strictly positive, which means every player minimizes a weighted sum of all cost functions [16]. The set of Pareto optimal solutions is called Pareto frontier. As for those solutions, superior to the Nash outcome (threatpoint) which appears in a NCDG, are called Pareto improvement set [6]. Thus, the notion of Pareto optimality provides an appropriate solution for players who are willing to cooperate. Another

Manuscript received August 1, 2019; revised October 13, 2019; accepted October 30, 2019. Date of publication November 26, 2019; date of current version April 13, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61773284 and Grant 61921004. This work of Z. Ni was partially supported by National Science Foundation under Grant 1949921. Paper no. TII-19-3564. *(Corresponding author: Chaoxu Mu.)*

C. Mu and K. Wang are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: cxmu@tju.edu.cn; walker_wang@tju.edu.cn).

Z. Ni is with the Department of Computer, Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431 USA (e-mail: zhenni@fau.edu).

C. Sun is with the School of Automation, Southeast University, Nanjing 210096, China (e-mail: cysun@seu.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2019.2955966

noteworthy point is that these studies are mainly focusing on problem formulation and lack algorithm design. It is of some theoretical significance to develop a general algorithm.

Second, some ADP researches have provided us with inspiration to tackle the cooperative differential game. ADP usually takes advantage of neural networks (NNs) to perform function approximation and adaptive control. In theoretical aspect, many efforts have been made to solve NCDG games. For example, [17] elaborated how to solve multiplayer games online by adaptive learning using the data measured from players. Johnson *et al.* [18] designed an actor–critic-identifier structured algorithm for the multiplayer nonzero-sum NCDG, where each player only minimized its own cost function without mutual cooperation. In application aspect, ADP control schemes are also extensively applied to industrial systems [19]–[21]. Such as [19] proposed an adaptive critic controller for the heating-ventilation–air-conditioning system, and [20] developed an event-based dual heuristic programming method for a class of networked systems. In this article, an ADP algorithm is designed to solve the LFC problem.

Third, frequency regulation of power systems has been a hot issue in the control community, and LFC is extensively applied to balance tieline power and frequency when the load fluctuates. Considerable control algorithms have been proposed to solve the LFC for interconnected power systems. Among them a conventional approach is proportional-integral (PI) control strategy, and some other typical representatives such as sliding mode control (SMC) [22]–[24], robust H_∞ control [25], [26], adaptive control [27], [28], and multiagent reinforcement learning (MARL) control [29]–[31]. Specifically, in [22], an SMC-based frequency controller was designed for multiarea interconnected power systems, where the switching surface was constructed for each area to improve dynamic performance. By adopting an adversarial idea, Peng *et al.* [25] proposed an event-driven robust control algorithm with a given attenuation level γ. In addition, [27] and [28] used a novel hierarchical structure to construct the composite controller, which consists of the primary PI controller and supplementary adaptive controller. This hierarchical design improved the dynamic process. MARL is another typical learning control. It includes two agents in each area: The estimator agent provides the area control error signal while the controller agent uses reinforcement learning to control the power system by trial-and-error method [30].

Comparing these studies, it can find that PI is not adaptive and SMC improves the dynamic effect but is still not optimal. In H_∞ control, the performance is related to the level γ. Moreover, most existing intelligent learning algorithms rely on trial-and-error scheme, that is, the success rate must be considered. So how to devise an adaptive controller in the optimal sense is still meaningful. To the best of our knowledge, there is no result to solve the CDG using ADP technique, and an effective algorithm for obtaining Pareto optimal solutions is also a well-recognized conundrum. On the other hand, by modeling a power system as a differential game, it is desirable to provide a novel idea to tackle LFC problem in a cooperative manner. These motivate our work.

In this article, we are only concentrating in obtaining Pareto solutions of a CDG where all players form a grand coalition and adopt a joint control policy. It is assumed that players make a binding cooperation agreement at the start and resolutely enforce later; besides, all players share the closed-loop state-feedback information and know mutual cost functions. The whole work is composed of theoretical algorithm design and practical verification. The main contributions are summarized as follows: 1) The cooperative problem of N-player nonzero-sum differential game is transformed into a general optimal control problem, which is embodied by a parametric Hamilton–Jacobian–Bellman (HJB) equation. This scheme is applicable to both nonlinear systems and linear systems. 2) For the first time, a novel ADP algorithm is proposed to approximately obtain Pareto solutions. This algorithm consists of two adaptive learning stages, and can be easily implemented with only one critic NN. 3) In terms of application, the proposed algorithm is used successfully to solve the LFC for a two-area benchmark power system. This problem is tackled within the framework of cooperative game, which brings a novel idea to solve LFC. Comparative simulation results indicate that the proposed method has better control performance.

The remainder of this article is organized as follows. First, Section II formulates the optimal problem for N-player nonzero-sum differential game and completes the problem transformation. Section III implants the critic NN to approximate the joint cost function and afterwards develops an adaptive algorithm consisting of two learning stages. Furthermore, in Section IV, a numerical system and a two-area power system are provided to verify the cooperative performance, which is achieved by comparison with the Nash equilibrium. Finally, Section V concludes this article.

II. Problem Formulation and Mathematical Description

Notation 1: $\mathcal{N} = \{1, 2, \ldots, N\}$ indicates a grand coalition composed of all players. $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$ denote the set of real numbers, the n-dimensional Euclidean space and the set of real $n \times m$ matrices. "I" signifies the identity matrix of the appropriate dimension. "$\top$" means the transpose operation and "∇" means the gradient operation.

Consider the N-player nonzero-sum game which is described by the input-affine nonlinear differential equation

$$\dot{x} = f(x(t)) + \sum_{i=1}^{N} g_i(x(t))u_i(t), \quad x(0) \triangleq x_0 \in \mathbb{R}^n \tag{1}$$

where $x(t) \in \mathbb{R}^n$ represents the state variable that is influenced by all players. $f(x(t)) \in \mathbb{R}^n$ represents the system drift dynamic, $g_i(x(t)) \in \mathbb{R}^{n\times m_i}$ represents the input matrix. $u_i(t) \in \mathbb{R}^{m_i}$ is the control policy manipulated by player i, and belongs to an admissible control space $\mathcal{U}_i, i \in \mathcal{N}$ [2]. It is normally assumed that $f(x), g_i(x)$ are locally Lipschitz with $f(0) = 0$. For concision, the time variable t is omitted in the sequel.

In this nonzero-sum game, every player wants to minimize the following quadratic cost function on a infinite horizon:

$$J_i(x_0,u_1,\cdots,u_N)\triangleq J_i:=\int_t^\infty (x^\top Q_i x+u_i^\top R_i u_i)d\tau \quad (2)$$

where two penalty matrices $Q_i\in\mathbb{R}^{n\times n}$ and $R_i\in\mathbb{R}^{m_i\times m_i}$ are symmetric and positive definite.

The differential game is essentially an optimal control problem where the control goal is to minimize the predefined cost function, with as little as possible control action. In practice, the state and control input are usually defined to describe deviations of certain variables from their target values or their long-run equilibrium levels [16]. Therefore, the cost function usually consists of different deviation terms, where every deviation is quadratically penalized, i.e., the state penalty $x^\top Q_i x$ and control energy cost $u_i^\top R_i u_i$. Also note that, by properly selecting the matrix Q_i, the preference of every player can be determined. In the LFC problem, the system state is mainly composed of the frequency deviation and power deviation. Therefore, the frequency regulation can be achieved by minimizing the defined cost function.

Next, we analyze two kinds of desired gaming results for NCDG and CDG, namely Nash equilibrium and Pareto equilibrium.

Case A: Noncooperative Optimal Control

In a NCDG, each player unilaterally pursues the minimization of (2) to get the optimal policy. Hence, this game is equivalent to solving N single-objective optimization problems

$$\begin{cases}\min\limits_{u_i} J_i\\ \text{s.t. } \dot{x}=f(x)+\sum_{i=1}^N g_i(x)u_i(t),\ x(0)\triangleq x_0\end{cases}. \quad (3)$$

With no cooperation, the ideal gaming results for (3) can be described by the following definition.

Definition 1 (cf. [2]): A set of polices $\{u_1^*,\cdots,u_N^*\}$ will constitute a Nash equilibrium for a nonzero-sum NCDG, if the following inequalities are simultaneously held:

$$J_i^{Nash}\triangleq J_i(u_1^*,\cdots,u_i^*,\cdots,u_N^*)\le J_i(u_1^*,\cdots,u_i,\cdots,u_N^*).$$

According to the existing studies [3], [17], [18], one can obtain the Nash optimal control policy as

$$u_i^*\triangleq u_i^*(x)=-\frac{1}{2}R_i^{-1}g_i^\top(x)\nabla J_i^{Nash}. \quad (4)$$

Case B: Cooperative Optimal Control

In a CDG, the cooperation goal of the coalition $\mathcal{N}$ is to ensure the minimization of overall cost under a formulated agreement. So we consider this overall performance situation in which all players cannot be improved simultaneously with at least one player being improved, i.e., so-called Pareto equilibrium. This allows us to reformulate this problem as a weighted sum optimal control problem [4].

Before proceeding further, Definition 2, given below, states such a control policy which can minimize the joint cost function combined by all players' cost is Pareto optimal.

Definition 2 (cf. [16]): Let the weight $\alpha_i\in(0,1)$, if a parameter set $\Phi=\{(\alpha_1,\alpha_2,\cdots,\alpha_N)\mid \alpha_i\ge 0,\sum_{i=1}^N\alpha_i=1\}$ exists such that it makes $U^*\in\mathcal{U}$ satisfy

$$U^*\in\arg\min_{u\in\mathcal{U}}\left\{\sum_{i=1}^N\alpha_i J_i(x,U)\right\}$$

then U^* is Pareto optimal, where U is the joint control $U\triangleq[u_1,u_2,\cdots,u_N]^\top\in\mathcal{U}_1\times\mathcal{U}_2\times\cdots\times\mathcal{U}_N\triangleq\mathcal{U}$.

Formally, control policy U^* is Pareto optimal if at least one of the following inequalities is strict:

$$J_i(x,U)\ge J_i(x,U^*)\triangleq J_i^*, i=\mathcal{N}. \quad (5)$$

The correlative cost point $\left(J_1^*,J_2^*,\ldots,J_N^*\right)$ is thus called a Pareto optimal solution.

As a consequence, for system (1), the joint cost function of CDG is then defined as follows:

$$J_\alpha\triangleq J_\alpha(x,U)=\sum_{i=1}^N\alpha_i J_i=\int_t^\infty\left(x^\top\mathcal{Q}_\alpha x+U^\top MU\right)d\tau, \quad (6)$$

where the definite matrix $\mathcal{Q}_\alpha=\sum_{i=1}^N\alpha_i Q_i$ and the diagonal matrix $M=\text{blockdiag}(\alpha_1R_1,\alpha_2R_2,\cdots,\alpha_NR_N)$.

When considering cooperation, each player is responsible for the minimization of all cost functions, and hence it is a multiobjective optimization. By defining the argument input-matrix $G(x):=[g_1,g_2,\cdots,g_N]\in\mathbb{R}^{n\times m}$ and considering the joint cost (6). Hence, this CDG game is equivalent to solving such a single-objective optimization problem

$$\begin{cases}\min\limits_{U} J_\alpha\\ \text{s.t. } \dot{x}=f(x)+G(x)U(t),\ x(0)\triangleq x_0\end{cases}. \quad (7)$$

Next, with feedback information, we present how to derive the optimal joint control policy $U^*(x)$.

1) Taking the differential of (6) along the state trajectory, one can derive the following Lyapunov equation:

$$0=x^\top\mathcal{Q}_\alpha x+U^\top MU+(\nabla J_\alpha)^\top\left(f(x)+G(x)U\right),J_\alpha(0)=0. \quad (8)$$

2) Define the Hamiltonian of this problem as

$$H(x,U,\nabla J_\alpha)=x^\top\mathcal{Q}_\alpha x+U^\top MU+(\nabla J_\alpha)^\top\left(f+GU\right). \quad (9)$$

3) Employing Bellman's optimality principle [32], the optimal joint cost $J_\alpha^*(x)$ can guarantee the HJB equation

$$0=\min_U H(x,U,\nabla J_\alpha^*). \quad (10)$$

4) By adopting the stationary conditions, it yields the optimal joint control policy

$$\frac{\partial H(x,U,\nabla J_\alpha^*)}{\partial U}=0\Rightarrow U^*(x)=-\frac{1}{2}M^{-1}G^\top(x)\nabla J_\alpha^*. \quad (11)$$

Obviously, every element of $U^*(x)$, i.e., the individual Pareto optimal control policy is given as

$$u_i^*\triangleq u_i^*(x)=-\frac{1}{2}(\alpha_iR_i)^{-1}g_i^\top(x)\nabla J_\alpha^*. \quad (12)$$

Remark 1: In principle, varying the weight α_i means different Pareto optimal control policies. A binding agreement is

built on a suitable selection of α_i, which can ensure the overall performance $J_{\text{sum}} = \sum_{i=1}^{N} J_i$ in the sense of Pareto.

Remark 2: One major difference between CDG and NCDG is within the determination of control policies, that is, the Nash policy is only related to its own cost function, seeing (4). While in a CDG, every player seeks its Pareto policy in accordance with the costs of the coalition $\mathcal{N}$, seeing (12).

Remark 3: A Pareto optimal control policy participates in gaming from two aspects: The individual control policy affects the joint cost (6); all costs of the coalition $\mathcal{N}$ determine the individual control policy with the form of weight α_i.

Remark 4: According to Definition 1, the associated Nash equilibrium point (or threatpoint) is $(J_1^{Nash}, \cdots, J_N^{Nash})$. Then a Pareto optimal solution $(J_1^*, \cdots, J_N^*)$ can be called a Pareto improvement solution if it satisfies $J_i^* \leq J_i^{Nash}$ for all $i \in \mathcal{N}$. It is desired that mutual cooperation can achieve a smaller overall cost, i.e., $J_{\text{sum}}^* \leq J_{\text{sum}}^{Nash}$.

5) Applying (11) into (10), the optimal joint cost can be obtained by solving the parametric HJB equation

$$0 = x^\top \mathcal{Q}_\alpha x + \nabla J_\alpha^{*\top} f(x) - \frac{1}{4} \nabla J_\alpha^{*\top} G(x) M^{-1} G^\top(x) \nabla J_\alpha^*. \tag{13}$$

It is obvious that if (13) can be deduced to obtain J_α^*, then using the conditions $J_{\text{sum}}^* \leq J_{\text{sum}}^{Nash}$ and $J_i^* \leq J_i^{Nash}$, the required Pareto frontier and Pareto improvement set can be obtained. However, this equation is intractable or impossible to analytically solve due to its nonlinear nature and partial derivative. Till now, this CDG has been translated to a general optimal problem with the joint control $U(x)$, and this formulation contributes to approximate the solutions of (13) using the developed ADP-based algorithm.

III. Adaptive Critic Learning-Based Controller Design Using Neural Network

In this section, we devise the ADP-based approximate scheme. Different from the noncooperative differential game in [3] and [18], which takes advantage of N critic NNs to approximate all players' cost functions, only one critic NN is adopted here to approximate the joint cost function J_α^*.

First, a three-layer feedforward NN is utilized to reconstruct J_α^* under the precondition that ω_c is a bounded constant ideal weight vector [33], then the approximations of the joint cost and its gradient can be expressed as

$$J_\alpha^*(x) = \omega_c^\top \varphi(x) + \varepsilon(x) \tag{14a}$$

$$\nabla J_\alpha^*(x) = \nabla \varphi^\top(x) \omega_c + \nabla \varepsilon(x). \tag{14b}$$

This NN consists of three layers: Input-layer, hidden-layer, and output-layer. For simplicity, the input-to-hidden weight is fixed to be 1, and thus the input information of hidden-layer is the state $x(t)$. The weight $\omega_c(t) \in \mathbb{R}^{n_h}$ connects the hidden-layer and output-layer, and the corresponding activation function is $\varphi(x) \in \mathbb{R}^{n_h}$, where n_h signifies the number of hidden-layer neurons. The output of critic NN is $\nabla J_\alpha^*(x)$ and thus the structure is '$n - n_h - 1$'. Besides, $\varepsilon(x)$ is the reconstruction error and is also assumed to be bounded.

Substituting (14b) into (11), the optimal joint control policy can be rewritten as

$$U^*(x) = -\frac{1}{2} M^{-1} G^\top(x) \big(\nabla \varphi^\top(x) \omega_c + \nabla \varepsilon(x) \big). \tag{15}$$

Let the estimated weight $\hat{\omega}_c$ approximate the ideal weight ω_c. Then, the actual outputs of the critic NN are

$$\hat{J}_\alpha(x) = \hat{\omega}_c^\top \varphi(x) \tag{16a}$$

$$\nabla \hat{J}_\alpha(x) = \nabla \varphi^\top(x) \hat{\omega}_c. \tag{16b}$$

Correspondingly, the approximation of optimal joint control policy can be derived as

$$\hat{U} \triangleq \hat{U}(x) = -\frac{1}{2} M^{-1} G^\top(x) \nabla \varphi^\top(x) \hat{\omega}_c \tag{17}$$

and the approximated individual Pareto control policy is

$$\hat{u}_i \triangleq \hat{u}_i(x) = -\frac{1}{2} (\alpha_i R_i)^{-1} g_i^\top(x) \nabla \varphi^\top(x) \hat{\omega}_c \tag{18}$$

Furthermore, using the approximate expressions (16b) and (17), the approximation of Hamiltonian is thus, as follows:

$$\begin{aligned} &\hat{H}(x, \hat{U}, \hat{\omega}_c) \\ &= x^\top \mathcal{Q}_\alpha x + \hat{U}^\top M \hat{U} + \hat{\omega}_c^\top \nabla \varphi(x) \big(f(x) + G(x) \hat{U} \big) \triangleq \mathcal{E}_c. \end{aligned} \tag{19}$$

Remark 5: It is worth emphasizing that the policy (18) will be adopted during the learning phase to achieve the estimation of $\hat{\omega}_c \to \omega_c$, and the optimal policy is then derived by (15).

For deducing the weight tuning law, one first establishes the following relation:

$$\frac{\partial \mathcal{E}_c}{\partial \hat{\omega}_c} = \nabla \varphi(x) \big(f(x) + G(x) \hat{U} \big) \triangleq \eta. \tag{20}$$

Our purpose is to derive $\mathcal{E}_c \to 0$, for convenient operations, which can be transformed to minimize the squared residual error $E_c = \mathcal{E}_c^\top \mathcal{E}_c / 2$. Based on the normalized gradient-descent algorithm, a parameterized weight tuning law is given by

$$\dot{\hat{\omega}}_c = -l_c \frac{\partial E_c}{\partial \hat{\omega}_c} = -l_c \frac{\eta}{(1 + \eta^\top \eta)^2} \underbrace{\big(\eta^\top \hat{\omega}_c + x^\top \mathcal{Q}_\alpha x + \hat{U}^\top M \hat{U} \big)}_{T_1} \tag{21}$$

where $l_c > 0$ is the learning rate and $(1 + \eta^\top \eta)^2$ is a normalized processing term [17]. In the learning implementation, the term T_1 is actually

$$\begin{aligned} T_1 = &\left[\nabla \varphi(x) \big(f(x) + \sum_{i=1}^{N} g_i(x) \hat{u}_i \big) \right]^\top \hat{\omega}_c \\ &+ \sum_{i=1}^{N} \alpha_i x^\top Q_i x + \sum_{i=1}^{N} \alpha_i \hat{u}_i^\top R_i \hat{u}_i. \end{aligned} \tag{22}$$

Remark 6: Note that the tuning law (21) is related to the weight coefficient α_i, which requires network learning to be conducted multiple times. Therefore, how to determine the weight α_i should also be taken into account in the algorithm design. Moreover, the well-known persistency-of-excitation (PE)

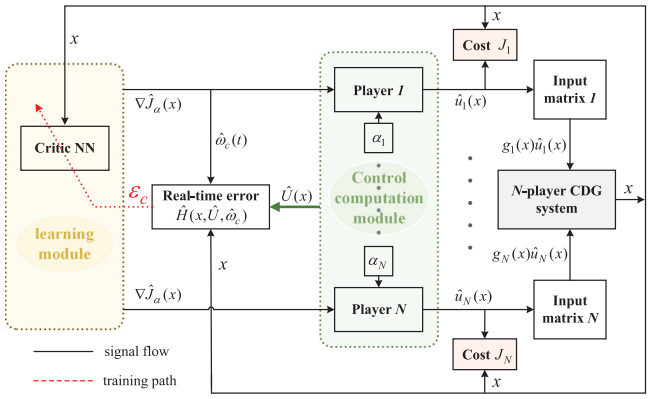


Fig. 1. Implementation of the proposed cooperative scheme.

condition is necessary during the learning phase. This proposed scheme can ensure closed-loop stability and weight convergence, this proof is omitted here due to page limitation. Some similar proofs can be referred to [3] and [9].

In order to better describe this control scheme and to present the training process, a structural diagram Fig. 1 is interpolated to highlight crucial components of this learning system and their mutual relationships. The critic NN transmits its real-time approximation value to every player. By means of the coefficient α_i, every player can thus individually compute its Pareto control policy and cost.

For the sake of clarity, a two-player CDG system is employed to illustrate the algorithm flow, seeing Algorithm 1. Note that, in this situation, there exists such a relation: $\alpha_2 = 1 - \alpha_1$. Therefore, our algorithm only needs to determine α_1. This algorithm *will determine the scope of* α_1 *(the first learning stage), and eventually obtain the Pareto frontier (the second learning stage).*

Remark 7: For implementing this algorithm, it is necessary to solve the Nash equilibrium, i.e., step 2. In this design, in order to obtain the Nash solution, we adopt the following weight tuning law for every critic NN

$$\dot{\hat{\omega}}_{ci} = -l_{c,i}\frac{\sigma_i}{(1+\sigma_i^\top\sigma_i)^2}\left(\sigma_i^\top\hat{\omega}_{ci} + x^\top Q_i x + (\hat{u}_i^n)^\top R_i \hat{u}_i^n\right) \tag{23}$$

where $\hat{\omega}_{ci}$ is ith critic weight and $l_{c,i} > 0$ is the associated learning rate. The Nash policy and variable σ_i are

$$\hat{u}_i^n = -\frac{1}{2}R_i^{-1}g_i^\top(x)\nabla\varphi_i^\top(x)\hat{\omega}_{ci}$$

$$\sigma_i = \nabla\varphi_i(x)\left(f(x) + \sum_{i=1}^{N} g_i(x)\hat{u}_i^n\right).$$

In addition, other methods can be found in [3] and [33].

Remark 8: Some annotations are helpful

1) For N-player CDG, Algorithm 1 also can determine the values of α_i when fixing other $N-2$ weight coefficients.

2) The initial coefficient α_0 is generally given to a smaller value, such as 0.1. If this value does not match the condition $J_{\text{sum}}^* \le J_{\text{sum}}^{Nash}$, then one can choose a larger value.

Algorithm 1: Neural Learning for Finding Pareto Optimal Solutions of a Two-player CDG.

Input: α^0, $\Delta\alpha^b$, $\Delta\alpha^s$ and T (the single running time).

1: **Initialization:**
   System settings: x_0 and $R_i, Q_i;\ i \in \mathcal{N}$;
   Learning parameters: initial weight $\hat{\omega}_c(0)$ and l_c.
2: **Calculate the Nash outcome for (3):**
   Obtain Nash optimal costs J_i^{Nash} and J_{sum}^{Nash}.
3: **First learning with a large step:**
   Set the single excitation time T_m;
   give α_1 an initial guess α^0 and a large step $\Delta\alpha^b$.
4: **for** $\alpha_1 = \alpha^0; \alpha_1 = \alpha_1 + \Delta\alpha^s; \alpha_1 < 1$ **do**
5: **while** $t \le T_m$ **do**
6: approximate the joint cost function using (16);
7: update the control policy of player i by (18);
8: train the critic weight according to the law (21);
9: **end while**
10: acquire the converged weight $\omega_c \approx \hat{\omega}_c(T)$;
11: applying this weight to the system yields J_α^* and J_{sum}^*.
12: **if** $J_{\text{sum}}^* > J_{\text{sum}}^{Nash}$ **then**
13: the current J_α^* is not persuasive solution; break.
14: **end if**
15: **end for**
16: determine the satisfied interval of α_1: $[\alpha_{\min}, \alpha_{\max}]$.
17: **Second learning with a small step:**
   Give a small step $\Delta\alpha^s$.
18: **for** $\alpha_1 = \alpha_{\min}; \alpha_1 = \alpha_1 + \Delta\alpha^b; \alpha_1 = \alpha_{\max}$ **do**
19: **while** $t \le T_m$ **do**
20: perform the learning phase as steps 6–8.
21: **end while**
22: use $\hat{\omega}_c(T)$ to compute J_i^*, J_{sum}^* and record them.
23: **end for**
24: obtain all the cost combinations and get Pareto frontier.

Output: $\alpha_{\min}$, $\alpha_{\max}$ and Pareto solutions J_i^*, J_{sum}^*.

3) Note that the joint cost function with $\alpha_i \in \mathbb{R}$ is a linear combination of $J_1, J_2, \ldots, J_N$, so the interval $[\alpha_{\min}, \alpha_{\max}]$ is continuous and the resulting Pareto frontier is also smooth.

4) Algorithm 1 can obtain Pareto solutions under different coefficients, which can provide some analytical insights and guidelines. An appropriate selection means a binding agreement that each player voluntarily adheres.

IV. Numerical Results and Analysis

In this section, using the proposed cooperation scheme, a nonlinear CDG system is simulated to verify the effectiveness of Algorithm 1. Subsequently, we investigate the cooperation between primary frequency control and secondary frequency control of a two-area interconnected power system.

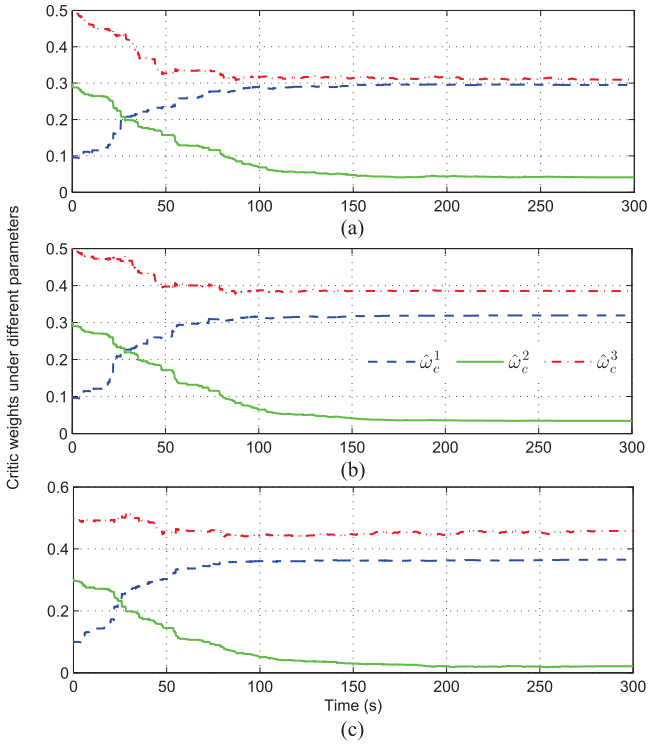


Fig. 2. Weight convergences under different parameters. (a) $\alpha_1 = 0.2$. (b) $\alpha_1 = 0.3$. (c) $\alpha_1 = 0.5$.

A. Two-Player Nonlinear CDG System

Consider the following two-player nonlinear DG system [3]:

$$\dot{x} = f(x) + g_1(x)u_1(x) + g_2(x)u_2(x) \tag{24}$$

$$\text{with } f(x) = \begin{bmatrix} x_2 - 2x_1 \\ -x_2 - 0.5x_1 + 0.25x_2\left(\cos\left(2x_1 + 2\right)\right)^2 \\ + 0.25x_2(\sin\left(4x_1^2\right) + 2)^2 \end{bmatrix}$$

$$g_1(x) = \begin{bmatrix} 0 \\ \cos\left(2x_1 + 2\right) \end{bmatrix}, g_2(x) = \begin{bmatrix} 0 \\ \sin\left(4x_1^2\right) + 2 \end{bmatrix}. \tag{25}$$

In (24), $x = [x_1, x_2]^\top$ denotes the state vector. The related parameters are configured as $x_0 = [0.5, 0.2]^\top$; $R_1 = 2I, R_2 = I$, $Q_1 = 2I, Q_2 = I$; $\hat{\omega}_c(0) = [1, 0.3, 0.5]^\top$; $l_c = 5, T_m = 300$ s; $\alpha^0 = 0.1, \Delta\alpha^b = 0.1, \Delta\alpha^s = 0.005$.

The critic NN is structured as "$2 - 3 - 1$". The activation function is $\varphi(x) = [x_1^2, x_1x_2, x_2^2]^\top$, and thus the estimated critic weight is $\hat{\omega}_c(t) = [\hat{\omega}_c^1, \hat{\omega}_c^2, \hat{\omega}_c^3]^\top$. The probing noises are injected in the first 280 s of the learning.

According to the running of the first learning stage in Algorithm 1, feasible values of α_1 locate in [0.2, 0.6]. For simplicity and without loss of generality, here, three sets of representative results are given to illustrate different cooperation consequents. Fig. 2 shows that three sets of weights. The cooperative control policies can be obtained using the converged weights and exhibited as Fig. 3. As can be seen, in three schemes,

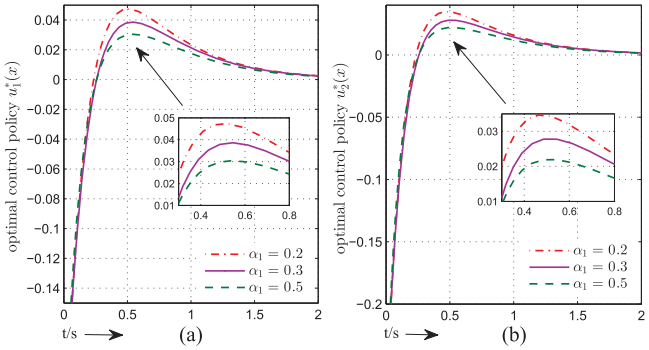


Fig. 3. Cooperative control comparison among three schemes.

TABLE I
TWO-PLAYER COOPERATIVE PERFORMANCE COMPARISON

Schemes	Nash	$\alpha_1 = 0.2$	$\alpha_1 = 0.3$	$\alpha_1 = 0.5$
J_1	0.1610	0.1647	0.1539	0.1421
J_2	0.0771	0.0714	0.0749	0.0827
J_{sum}	0.2381	0.2361	0.2288	0.2248

these two controls can be coordinated to complete control mission at about 2.5 s.

Then we continue to analyze the cooperative performance and a detailed cost comparison is given in Table I. As previously stated, cooperative control scheme seeks to optimize overall cost in the premise of decreasing at least one player's cost. It can be observed that three cooperative schemes have optimized the overall control cost, that is, any one of Pareto optimal costs, i.e., 0.2361, 0.2288, 0.2248, is definitely less than the Nash cost. Evidently, scheme $\alpha_1 = 0.3$ is a "*win–win*" cooperation agreement, where two players' costs are all superior to the Nash equilibrium point.

Finally, a total of 81 sets of Pareto optimal solutions can be calculated by performing the second learning stage, and the Pareto frontier is thus presented as Fig. 4. The relation between Nash equilibrium point and Pareto improvement set has been clearly marked. Based on afore-analyzed results, it can be concluded that, one can coordinate two players' control actions by regulating α_1 to optimize either player's cost or together optimize two players' costs.

B. Two-Area Benchmark Interconnected Power System

Next, the LFC problem of a two-area interconnected power system is studied. According to [34]–[37], primary frequency (PF) control and secondary frequency (SF) control play important roles in maintaining the frequency stability. We will first formulate this system as a two-player CDG differential game, and then study the cooperation between PF control and SF control, which are regarded as two players.

The classic IEEE two-area power system frequency control model is shown in Fig. 5, and some critical notations have been annotated. Besides, ACE_i means the ith area control error. The simulation environment is: $T_{G1} = T_{G2} = 0.08$ s, $T_{T1} =$

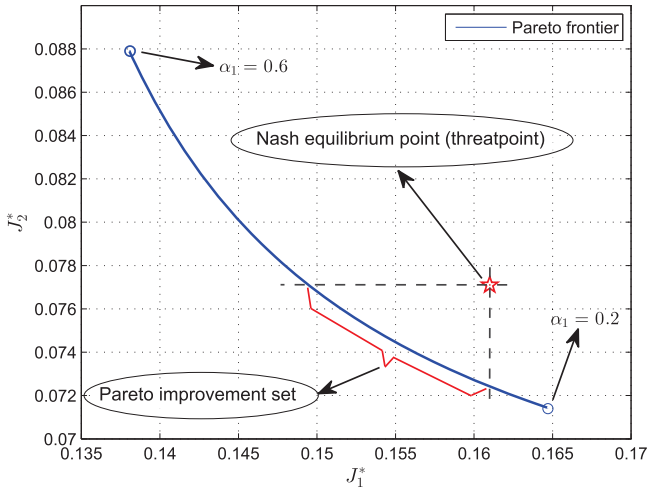


Fig. 4. Positional presentation between Pareto frontier and Nash equilibrium point.

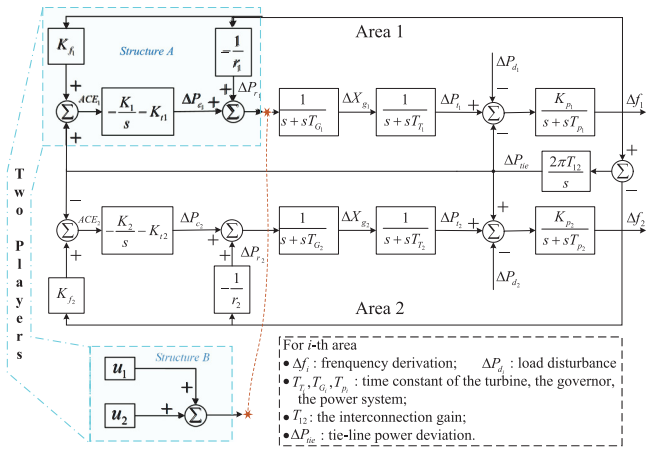


Fig. 5. Structure sketch of a two-area power system.

$T_{T2} = 0.08$ s, $T_{p1} = T_{p2} = 0.08$ s, $K_{p1} = K_{p2} = 0.08$ Hz, $T_{12} = 0.00545$ p.u. and $K_{f1} = K_{f2} = 0.45$ [37].

Next, we discuss five control schemes to comparatively illustrate the merits of CDG-based controller.

1) *Scheme 1 (PI control method):* In the light of the typical values in [35], the PF signal ΔP_{r_i} is provided by setting $r_1 = r_2 = 2.4$ Hz, and the SF signal ΔP_{c_i} is provided by adopting the gains $K_1 = K_2 = 0.15, K_{t1} = K_{t2} = 0.3$. For comparison, the SF signal is calculated as $\mathrm{d}\Delta P_{c_i}/\mathrm{d}t$.

2) *Scheme 2 (NCDG-based control method):* In the area 1, with the help of DG theory, we employ PF and SF control signals u_1 and u_2 to substitute the original control loop, i.e., *Struture A→Struture B*, as presented in blue part of Fig. 5. In order to simplify this application process, we do not consider control constraints and climbing rates, etc. Therefore, this system can be modeled by the linear equation

$$\dot{x} = Ax + B_1u_1 + B_2u_2 + D\Delta P_d, \tag{26}$$

where

$$x = [\Delta f_1, \Delta P_{t_1}, \Delta X_{g_1}, \Delta f_2, \Delta P_{t_2}, \Delta X_{g_2}, \Delta P_{c_2}, \Delta P_{tie}]^\top$$

$$A = \begin{bmatrix} A_{11} & A_{12} & 0 & 0 & 0 & 0 & 0 & A_{18} \\ 0 & A_{22} & A_{23} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_{33} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A_{44} & A_{45} & 0 & 0 & A_{48} \\ 0 & 0 & 0 & 0 & A_{55} & A_{56} & 0 & 0 \\ 0 & 0 & 0 & A_{64} & 0 & A_{66} & A_{67} & 0 \\ A_{71} & 0 & 0 & A_{74} & A_{75} & 0 & 0 & A_{78} \\ A_{81} & 0 & 0 & A_{84} & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_{11} = -\frac{1}{T_{p_1}}, A_{12} = \frac{K_{p_1}}{T_{p_1}}, A_{18} = -\frac{K_{p_1}}{T_{p_1}}, A_{22} = -\frac{1}{T_{T_1}}$$

$$A_{23} = \frac{1}{T_{T_1}}, A_{33} = -\frac{1}{T_{G_1}}, A_{44} = -\frac{1}{T_{p_2}}, A_{45} = \frac{K_{p_2}}{T_{p_2}}$$

$$A_{48} = \frac{K_{p_2}}{T_{p_2}}, A_{55} = -\frac{1}{T_{T_2}}, A_{56} = \frac{1}{T_{T_2}}, A_{64} = -\frac{1}{r_2T_{G_2}}$$

$$A_{66} = -\frac{1}{T_{G_2}}, A_{67} = -\frac{1}{T_{G_2}}, A_{71} = -2\pi T_{12}K_{t_2}$$

$$A_{74} = \frac{K_2}{r_2} - \frac{K_{t_2}}{r_2T_{p_2}} + 2\pi T_{12}K_{t_2}, A_{75} = \frac{K_{t_2}K_{p_2}}{r_2T_{p_2}}$$

$$A_{78} = \frac{K_{t_2}K_{p_2}}{r_2T_{p_2}} - K_2, A_{81} = 2\pi T_{12}, A_{84} = -2\pi T_{12}$$

$$B_1 = B_2 = \begin{bmatrix} 0 & 0 & \frac{1}{T_{G_1}} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^\top$$

$$D = \begin{bmatrix} -\frac{K_{p_1}}{T_{p_1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{K_{p_2}}{T_{p_2}} & 0 & 0 & 0 & 0 \end{bmatrix}^\top$$

$$\Delta P_d = \begin{bmatrix} \Delta P_{d_1} & \Delta P_{d_2} \end{bmatrix}^\top.$$

Note that (26) has the extra load disturbance term $D\Delta P_d$ when compared with the classical DG shown in (1). For simplicity, this article considers that ΔP_d substitutes the true random change with a known step load disturbance. Therefore, we adopt the idea of [6] to use the steady-state value after the disturbance as a reference point, and a deterministic NCDG problem is finally expressed as

$$\begin{cases} \min J_i = \int_t^\infty (x^\top Q_i x + u_i^\top R_i u_i) d\tau \\ \text{s.t. } \dot{x} = Ax + B_1u_1 + B_2u_2 \end{cases}. \tag{27}$$

Two players can determine the values of matrices Q_i and R_i according to their own preferences. In this article, they are

determined as follows:

$$Q_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, R_1 = 10$$

$$Q_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}, R_2 = 1.$$

As such, two associated costs can be derived as

$$J_1 = \int_t^\infty (x^\top Q_1 x + u_1^\top R_1 u_1) d\tau = \int_t^\infty (\Delta f_1^2(\tau) + 10 u_1^2(\tau)) d\tau$$

$$J_2 = \int_t^\infty (x^\top Q_2 x + u_2^\top R_2 u_2) d\tau$$

$$= \int_t^\infty (\Delta f_1^2(\tau) + \Delta f_2^2(\tau) + 2\Delta P_{tie}^2(\tau) + u_2^2(\tau)) d\tau.$$

These choices rely on two players' behaviors. The output of player 1 is only affected by the deviation Δf_1, but player 2 can influence area 2 through the tie line. Therefore, player 1 is committed to driving Δf_1 to 0; player 2 not only minimizes the frequency deviations Δf_1 and Δf_2 but also reduces ΔP_{tie} to support area 2. These two criteria express a balance between the cost of having nonzero deviations and the cost of the control required to make the deviations smaller. A similar selection can be found in [36].

3) *Scheme 3 (CDG-based control method):* In this case, two players cooperates to optimize the joint cost such that they can get a more happier overall cost. With the proposed scheme, the CDG-based formulation of (26) is given by

$$\begin{cases} \min J_\alpha = \alpha J_1 + (1-\alpha) J_2 \\ \text{s.t. } \dot{x} = Ax + B_1 u_1 + B_2 u_2 \end{cases}. \tag{28}$$

Since schemes 2 and 3 need optimal policies, adaptive learning is first implemented to approximate optimal control policies. In NCDG (27), it is desired to obtain every player's Nash policy, and hence, two critic NNs are structured as "$8-20-1$". The associated activation functions are $\varphi_1(x) = \varphi_2(x) = [x_1^2, x_1x_2, x_1x_8, x_2^2, x_2x_3, x_3^2, x_4^2, x_4x_5, x_4x_8, x_5^2, x_5x_6, x_6x_4, x_6^2, x_6x_7, x_7x_1, x_7x_4, x_7x_5, x_7x_8, x_1x_8, x_8x_4]$. Therefore, the corresponding estimated weight is $\hat{\omega}_{ci}(t) = [\hat{\omega}_{ci}^1, \hat{\omega}_{ci}^2, \ldots, \hat{\omega}_{ci}^{20}]^\top$ and will be randomly initialized. The running time is $T_m = 600$ s, and learning rates are $l_{c,1} = l_{c,2} = 2$.

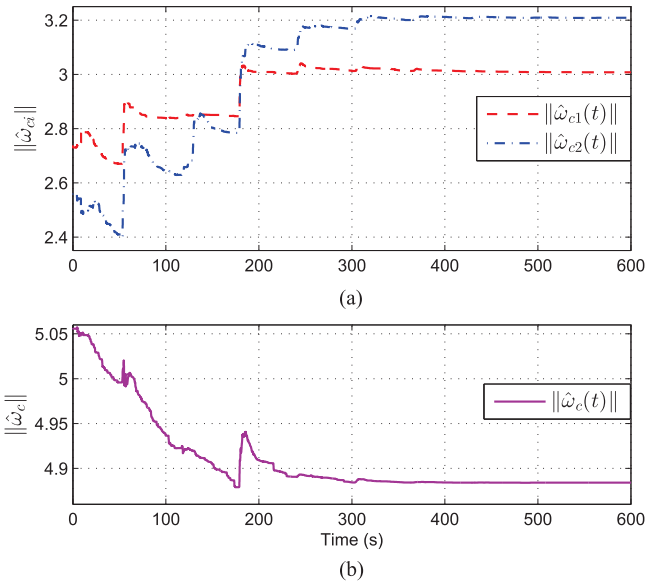


Fig. 6. Weight-norm trajectories.

After training, the weight-norm trajectories of two players are presented in Fig. 6(a). It can be seen that the final convergence roughly appears at 550 s.

For the CDG (28), only a critic NN is used to approximate the joint cost function, and the weight vector is similarly set as $\hat{\omega}_c(t) = [\hat{\omega}_c^1, \hat{\omega}_c^2, \ldots, \hat{\omega}_c^{20}]^\top$. By running Algorithm 1, it can be known that $\alpha = 0.6$ is a satisfactory value. Thus, the weight-norm convergence trajectory is revealed in Fig. 6(b).

4) *Scheme 4 (SMC method):* SMC is an effective control method with good dynamic response. Note that we are considering PF and SF control in area 1, so the SMC is only applied to area 1. Therefore, by defining $\Upsilon = [\Delta f_1, \Delta P_{t_1}, \Delta X_{g_1}]$ and using the methods in [23], [24], the sliding mode variable is defined as follows:

$$\rho(t) = C\Upsilon \tag{29}$$

where $C = [c_1, c_2, c_3]$ is the coefficient and the polynomial $c_3 p^3 + c_2 p^2 + c_1 p$ is Hurwitz. The reaching law is selected to be $\dot{\rho}(t) = -\xi sat(\rho)$ and the saturation function $sat(\rho)$ is to reduce chattering. In the simulation, the control gain is $\xi = 20$ and three coefficients are $c_1 = 0.9; c_2 = 0.05; c_3 = 0.05$.

5) *Scheme 5 (H_∞ method):* From the adversarial perspective, by giving an attenuation level, it is desired to do the utmost to attenuate the disturbance. This problem can be seen as a zero-sum game from the perspective of min–max optimization [38]. In order to implement this scheme, the attenuation level is $\gamma = 15$ according to [25].

Next, we apply PI control, Nash control, Pareto control, SMC, and H_∞ control into the power system, respectively. In this process, it is assumed that the prediction of ΔP_{d_i} is known, which indicates that ΔP_{d_i} stays zero throughout the control period while two step disturbances $\Delta P_{d_1} = -0.005$ p.u. and $\Delta P_{d_2} = +0.01$ p.u. occurs at $t = 0$. The evolution of three deviation signals $\Delta f_1, \Delta f_2, \Delta P_{tie}$ is, respectively, given in

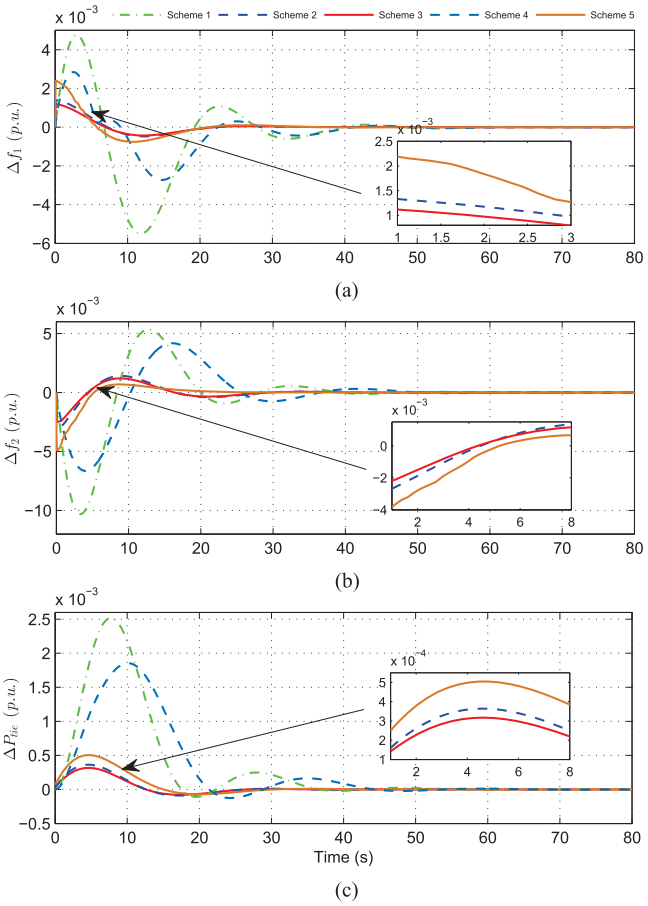


Fig. 7. LFC performance comparison of the two-area power system under three control schemes.

TABLE II
COST COMPARISON OF THREE CONTROL SCHEMES

Schemes	cost (10^{-4})		
	J_1	J_2	J_{sum}
Scheme 1	7.0882	9.4740	16.5622
Scheme 2	1.3763	2.5993	3.9756
Scheme 3	1.3223	2.4647	3.7870
Scheme 4	3.2299	7.1977	10.4276
Scheme 5	1.6624	2.9983	4.6607

Fig. 7(a)–(c). Table II also provides the specific cost (payoff) comparison during the entire frequency regulation process.

First, according to the above results, we conduct some analysis and illustrate the merits of the proposed scheme.

1) *Scheme 1* is only concerned with eliminating the deviations while *Schemes 2 and 3* achieve this goal with optimal manners, which indicates that the overshoot and oscillation of PI controller are relatively large, seeing Fig. 7. In *Scheme 2*, every player unilaterally optimizes, while in *Scheme 3*, two players cooperatively optimize. Therefore, Pareto policies are superior to Nash policies, seeing three close-ups.

2) Although *Scheme 4* ameliorates the dynamic response of area 1, it does not have the ability to coordinate area 2, and hence the overall cost is larger compared to *Schemes 2 and 3*. By attenuating the disturbance, *Scheme 5* can also obtain a relatively good control performance due to smaller fluctuation and less cost. But the nonzero-sum gaming results are more prominent and cooperation is a better choice.

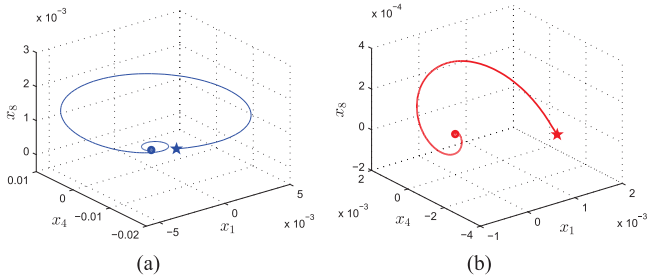


Fig. 8. Phase trajectories w.r.t. states x_1, x_4, x_8. (a) PI control. (b) Cooperative control. '$\circ$' marks the origin, '$\star$' marks the starting point after disturbances.

TABLE III
STATISTICAL RESULTS OF FIVE SCHEMES

Schemes		Statistical indices (10^{-3})				
		Overshoot	Undershoot	Mean	RMS	STD
Δf_1	*1*	4.764	-5.549	-0.229	1.793	1.778
	2	1.400	-0.047	0.033	0.313	0.311
	3	1.209	-0.043	0.021	0.265	0.263
	4	2.846	-2.725	0.255	0.550	0.488
	5	2.423	-0.076	0.083	0.588	0.583
Δf_2	*1*	5.333	-10.306	-0.232	2.715	2.705
	2	1.403	-2.901	-0.008	0.614	0.613
	3	1.211	-2.512	-0.002	0.513	0.513
	4	4.170	-6.646	-0.656	1.643	1.507
	5	0.691	-4.903	-0.332	1.286	1.243
ΔP_{tie}	*1*	2.516	-0.109	0.310	0.741	0.673
	2	0.364	-0.089	0.027	0.101	0.098
	3	0.317	-0.079	0.023	0.089	0.086
	4	1.855	-0.126	0.109	0.512	0.500
	5	0.505	-0.072	0.051	0.147	0.139

3) It can be known from Table II that cost J_2 is bigger because more deviation terms are considered for player 2. The overall cost of *Scheme 3* is the least, which means that the cooperation is necessary and the current agreement is persuasive. Fig. 8 also gives three dimensional (3-D) convergence curves of Δf_1, Δf_2, and ΔP_{tie} by using PI control and cooperative control.

Moreover, some dynamic analysis are presented by five statistical indices: Overshoot, undershoot, mean, roor-mean-square (RMS), and standard deviation (STD). They have been provided in Table III.

1) It is observed that five schemes all can drive three deviation terms to zero. These five indices can evaluate the dynamic process to some extent, such as STD reflects the dispersion of a collection of values. Besides, for RMS and STD, a smaller value means a better dynamic performance.

2) From Table III, one can see the obvious advantages of the proposed cooperation control in dynamic response. Especially, the overshoot and undershoot are the least, which indicates the CDG-based controller responds quickly.

To conclude, the first example illustrates the effectiveness of Algorithm 1 in obtaining cooperative Pareto solutions. This is a general algorithm and is beneficial for current studies. By comparing with different control methods, the second example shows that the proposed cooperation scheme has advantages in practical application. It provides a novel idea and an alternative method for solving LFC of multiarea power systems.

V. Conclusion

In this article, with the formulation of CDG, the optimal control problem of multiplayer systems was studied by defining a joint cost function, which was a weight coefficient combination of multiple costs. Then, a novel ADP-based learning algorithm was developed to find all Pareto optimal solutions, and it was performed by two learning stages based on a single-network structure. A two-player numerical example illustrated that the proposed algorithm obtained the feasible scope of the weight coefficient as well as Pareto optimal solutions. Finally, by designating the PF control and SF control as two players, the cooperative control of a two-area power system was studied by comparing with different control methods. The results demonstrated that the players' cooperation had some advantages and potential. The whole work theoretically presented a new algorithm and provided some references for practical engineering implementations. In future work, the random load changes, controller constraints, and climbing rates, etc. will be considered.

References

[1] R. Isaacs, *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization* (SIAM Series in Applied Mathematics). New York, NY, USA: Wiley, 1965.

[2] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia: PA, USA: SIAM, 1998.

[3] H. Zhang, L. Cui, and Y. Luo, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 206–216, Feb. 2013.

[4] H. L. Stalford, "Criteria for Pareto-optimality in cooperative differential games," *J. Optim. Theory Appl.*, vol. 9, no. 6, pp. 391–398, 1972.

[5] A. Haurie and B. Tolwinski, "Definition and properties of cooperative equilibria in a two-player game of infinite duration," *J. Optim. Theory Appl.*, vol. 46, no. 4, pp. 525–534, 1985.

[6] H. Chen, R. Ye, X. Wang, and R. Lu, "Cooperative control of power system load and frequency by using differential games," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 882–897, May 2015.

[7] S. H. Tamaddoni, S. Taheri, and M. Ahmadian, "Optimal preview game theory approach to vehicle stability controller design," *Vehicle Syst. Dyn.*, vol. 49, no. 12, pp. 1967–1979, 2011.

[8] D. V. Prokhorov, R. A. Santiago, and D. C. Wunsch, "Adaptive critic designs: A case study for neurocontrol," *Neural Netw.*, vol. 8, no. 9, pp. 1367–1372, 1995.

[9] C. Mu and K. Wang, "Approximate-optimal control algorithm for constrained zero-sum differential games through event-triggering mechanism," *Nonlinear Dyn.*, vol. 95, no. 4, pp. 2639–2657, 2019.

[10] X. Lu, B. Kiumarsi, T. Chai, Y. Jiang, and F. L. Lewis, "Operational control of mineral grinding processes using adaptive dynamic programming and reference governor," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2210–2221, Apr. 2019.

[11] Z. Ni, N. Malla, and X. Zhong, "Prioritizing useful experience replay for heuristic dynamic programming-based learning systems," *IEEE Trans. Cybern.*, vol. 49, no. 11, pp. 3911–3922, Nov. 2019.

[12] A. W. Starr and Y. C. Ho, "Nonzero-sum differential games," *J. Optim. Theory Appl.*, vol. 3, no. 3, pp. 184–206, 1969.

[13] W. Schmitendorf, "Cooperative games and vector-valued criteria problems," *IEEE Trans. Autom. Control*, vol. AC-18, no. 2, pp. 139–144, Apr. 1973.

[14] J. C. Engwerda, "Necessary and sufficient conditions for Pareto optimal solutions of cooperative differential games," *SIAM J. Control Optim.*, vol. 48, no. 6, pp. 3859–3881, 2010.

[15] P. V. Reddy and J. C. Engwerda, "Pareto optimality in infinite horizon linear quadratic differential games," *Automatica*, vol. 49, no. 6, pp. 1705–1714, 2013.

[16] J. Engwerda, *LQ Dynamic Optimization and Differential Games*. New York, NY, USA: Wiley, 2005.

[17] K. G. Vamvoudakis, H. Modares, B. Kiumarsi, and F. L. Lewis, "Game theory-based control system algorithms with real-time reinforcement learning: How to solve multiplayer games online," *IEEE Control Syst.*, vol. 37, no. 1, pp. 33–52, Feb. 2017.

[18] M. Johnson, R. Kamalapurkar, S. Bhasin, and W. E. Dixon, "Approximate N-player nonzero-sum game solution for an uncertain continuous nonlinear system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1645–1658, Aug. 2015.

[19] N. K. Dhar, N. K. Verma, and L. Behera, "Adaptive critic-based event-triggered control for HVAC system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 1, pp. 178–188, Jan. 2018.

[20] J. Yi, S. Chen, X. Zhong, W. Zhou, and H. He, "Event-triggered globalized dual heuristic programming and its application to networked control systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1383–1392, Mar. 2019.

[21] Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2684–2695, Sep. 2019.

[22] Y. Mi, Y. Fu, C. Wang, and P. Wang, "Decentralized sliding mode load frequency control for multi-area power systems," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4301–4309, Nov. 2013.

[23] C. Mu, Y. Tang, and H. He, "Improved sliding mode design for load frequency control of power system integrated an adaptive learning strategy," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6742–6751, Aug. 2017.

[24] H. Li, X. Wang, and J. Xiao, "Adaptive event-triggered load frequency control for interconnected microgrids by observer-based sliding mode control," *IEEE Access*, vol. 7, pp. 68 271–68 280, 2019.

[25] C. Peng, J. Zhang, and H. Yan, "Adaptive event-triggering H_∞ load frequency control for network-based power systems," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1685–1694, Feb. 2018.

[26] H. Zhang, J. Liu, and S. Xu, "H-Infinity load frequency control of networked power systems via an event-triggered scheme," *IEEE Trans. Ind. Electron.*, 2019, to be published, doi: 10.1109/TIE.2019.2939994.

[27] L. Dong, Y. Tang, H. He, and C. Sun, "An event-triggered approach for load frequency control with supplementary ADP," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 581–589, Jan. 2017.

[28] C. Mu, W. Liu, and W. Xu, "Hierarchically adaptive frequency control for an EV-integrated smart grid with renewable energy," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4254–4263, Sep. 2018.

[29] F. Daneshfar and H. Bevrani, "Load-frequency control: A GA-based multi-agent reinforcement learning," *IET Gener., Transmiss. Distribution*, vol. 4, no. 1, pp. 13–26, Jan. 2010.

[30] V. P. Singh, N. Kishor, and P. Samuel, "Distributed multi-agent system-based load frequency control for multi-area power system in smart grid," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5151–5160, 2017.

[31] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1653–1656, Mar. 2019.

[32] R. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[33] H. Jiang, H. Zhang, Y. Luo, and J. Han, "Neural-network-based robust control schemes for nonlinear multiplayer systems with uncertainties via adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 3, pp. 579–588, Mar. 2019.

[34] T. N. Pham, H. Trinh, and L. V. Hien, "Load frequency control of power systems with electric vehicles and diverse transmission links using distributed functional observers," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 238–252, Jan. 2016.

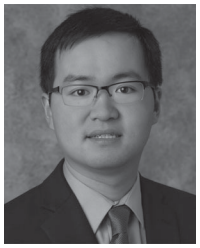
[35] L. Jiang, W. Yao, Q. H. Wu, J. Y. Wen, and S. J. Cheng, "Delay-dependent stability for load frequency control with constant and time-varying delays," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 932–941, May 2012.
[36] Sathans, and A. Swarup, "Intelligent load frequency control of two-area interconnected power system and comparative analysis," in *Proc. Int. Conf. Commun. Syst. Netw. Technol.*, 2011, pp. 360–365.
[37] T. P. I. Ahamed, P. S. Nagendra Rao, and P. S. Sastry, "A reinforcement learning approach to automatic generation control," *Electric Power Syst. Res.*, vol. 63, no. 1, pp. 9–26, 2002.
[38] Q. Zhang, D. Zhao, and Y. Zhu, "Event-triggered H_∞ control for continuous-time nonlinear system via concurrent learning," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1071–1081, Jul. 2017.

Chaoxu Mu (M'15–SM'18) received the Ph.D. degree in control science and engineering from the School of Automation, Southeast University, Nanjing, China, in 2012.

She was a Visiting Ph.D. Student with the Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia, from 2010 to 2011. She was a Postdoctoral Fellow with the Department of Electrical, Computer and Biomedical Engineering, The University of Rhode Island, Kingston, RI, USA, from 2014 to 2016. She is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. Her current research interests include nonlinear system control and optimization, adaptive and learning systems.

Ke Wang received the B.S. degree in control technology and instruments from the Shaanxi University of Science and Technology University, Xi'an, Shanxi, China, in 2017. He is currently working toward the Ph.D. degree in control science and engineering with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His current research interests include intelligent control, differential games and event-triggering.

Mr. Wang received the First Prize in the Mathematics competition of Chinese College Students.



Zhen Ni (M'15) received the B.S. degree in automation from the Huazhong University of Science and Technology, Wuhan, China, in 2010, and the Ph.D. degree in electrical, computer, and biomedical engineering from the University of Rhode Island, Kingston, RI, USA, in 2015.

He is currently an Assistant Professor with the Department of Computer, Electrical Engineering, and Computer Science, Florida Atlantic University, Boca Raton, FL, USA. He was with the Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, SD, USA, from 2015 to 2019. His current research interests include computational intelligence, reinforcement learning, and smart grid applications.

Prof. Ni has been actively involved in numerous conference and workshop organization committees in the society, including the General Co-Chair of the IEEE CIS Winter School, Washington, DC, USA, in 2016. He is an Associate Editor of the IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE since 2018, Associate Editor of the IEEE TRANSACTIONS OF NEURAL NETWORKS AND LEARNING SYSTEMS since 2019, and a Guest Editor for IET Cyber-Physical Systems: Theory and Applications (2017–2018). He is a recipient of the prestigious IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award (2020), International Neural Network Society Aharon Katzir Young Investigator Award (2019), URI Excellence in Doctoral Research Award (2016), and Chinese Government Award for Outstanding Students Abroad (2014).

Changyin Sun received the B.S. degree from the Department of Mathematics, Sichuan University, China, in 1996, the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2001 and 2003, respectively.

From March 2004 to September 2004, he was a Postdoctoral Fellow with the Department of Computer Science, The Chinese University of Hong Kong. Currently, he working as a Professor with the School of Automation, Southeast University. His research interests include intelligent control, neural networks, support vector machine, data-driven control, theory and design of intelligent control systems, optimization algorithms, pattern recognition, etc.

Dr. Sun is an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS (2010), *Neural Processing Letters* (2008), *International Journal of Swarm Intelligence Research* (2010), and *Recent Parents of Computer Science* (2008). He is also involved in organizing several international conferences.