

# Stable Reinforcement Learning for Optimal Frequency Control: A Distributed Averaging-Based Integral Approach

YAN JIANG <sup>1</sup> (Member, IEEE), WENQI CUI <sup>1</sup> (Graduate Student Member, IEEE),  
BAOSEN ZHANG <sup>1</sup> (Member, IEEE), AND JORGE CORTÉS <sup>2</sup> (Fellow, IEEE)

<sup>1</sup>Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195 USA

<sup>2</sup>Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093 USA

CORRESPONDING AUTHOR: YAN JIANG (e-mail: jiangyan@uw.edu).

This work was supported by the NSF under Grants ECCS-1807142, ECCS-1942326, ECCS-2153937, ECCS-2023531, and ECCS-1947050.

**ABSTRACT** Frequency control plays a pivotal role in reliable power system operations. It is conventionally performed in a hierarchical way that first rapidly stabilizes the frequency deviations and then slowly recovers the nominal frequency. However, as the generation mix shifts from synchronous generators to renewable resources, power systems experience larger and faster frequency fluctuations due to the loss of inertia, which adversely impacts the frequency stability. This has motivated active research in algorithms that jointly address frequency degradation and economic efficiency in a fast timescale, among which the distributed averaging-based integral (DAI) control is a notable one that sets controllable power injections directly proportional to the integrals of frequency deviation and economic inefficiency signals. Nevertheless, DAI does not typically consider the transient performance of the system following power disturbances and has been restricted to quadratic operational cost functions. This paper aims to leverage nonlinear optimal controllers to simultaneously achieve optimal transient frequency control and find the most economic power dispatch for frequency restoration. To this end, we integrate reinforcement learning (RL) to the classic DAI, which results in RL-DAI control. Specifically, we use RL to learn a neural network-based control policy mapping from the integral variables of DAI to the controllable power injections which provides optimal transient frequency control, while DAI inherently ensures the frequency restoration and optimal economic dispatch. Compared to existing methods, we provide provable guarantees on the stability of the learned controllers and extend the set of allowable cost functions to a much larger class. Simulations on the 39-bus New England system illustrate our results.

**INDEX TERMS** Frequency control, Lyapunov stability, reinforcement learning, steady-state and transient performance.

## I. INTRODUCTION

The key to the normal operation of a power system is the balance between electric power supply and demand over the network [1]. For instance, the main cause of the 2021 Texas power crisis is that the deficient supply of power due to frozen equipment could not meet the high demand for electricity in cold weather. A system frequency deviation from its nominal value is a reflection of a power imbalance [2], which makes frequency control a vital task of grid operators. Traditionally, this task is performed in a hierarchical

structure composed of three layers with timescale separation: primary—droop control (<20 s), secondary—frequency restoration (30 s–10 min), and tertiary—economic dispatch (>15 min) [2].

Nowadays, power systems are experiencing a change in the mix of generation, where conventional synchronous generators are gradually being replaced by renewable energy sources like solar and wind energy [3]. It is anticipated that the renewable share of the electricity generation mix in the United States will double from 21% in 2020 to

42% in 2050 [4]. Intermittent renewable sources are typically inverter-interfaced, which may adversely affect the robustness of the frequency dynamics due to the loss of inertia [5]. This exposes power systems to larger and faster frequency fluctuations than before, which has motivated active research on flexible distributed frequency control schemes that can break the hierarchy by addressing simultaneously frequency degradation and economic efficiency at a fast timescale.

A key challenge in frequency control is that the power imbalances across the network are not explicitly known. A number of studies have proposed distributed algorithms to overcome this challenge [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. The works in [6], [7], [8], [9], [10], [11], [12], [13], [14] focus on optimizing the steady-state frequency and economic performance using a principled design of fixed updating rules involving agent communication for control dynamics. The approaches mainly fall in two categories. The first category [7], [8], [12], [13], [14] rests on a primal-dual interpretation of power system dynamics under a properly designed optimization problem. This approach, however, always requires the estimation of certain system parameters. The second category [6], [9], [10], [11] builds upon various consensus algorithms to converge to an equilibrium with nominal frequency and economic efficiency. A notable example in this category is the distributed averaging-based integral (DAI) mechanism [6], [10], [11], [16], where the controllable power injections are directly proportional to the integrals of frequency deviation and economic inefficiency signals. A key caveat to this approach is that DAI control has been so far restricted to quadratic cost functions.

The works above focus on the optimization of the steady-state performance and do not typically consider the transient performance along the system trajectories following power disturbances. In fact, the optimization of transient performance is a challenging problem due to the nonlinearity of power dynamics and the uncertainty in power disturbances. Reinforcement learning (RL) [17] is a powerful tool for learning from interactions with uncertain environments and determining how to map situations to actions so that a desired performance is optimized. By virtue of the above feature, RL has emerged [18], [19] as an effective instrument to address the optimal transient frequency control problem in nonlinear power systems under unknown power disturbances. Nevertheless, the Achilles' heel of standard RL algorithms is their lack of provable stability guarantees, which presents a significant barrier to their practical implementation for the operation of power systems. In fact, many works [20], [21], [22], [23] optimize the transient performance by learning control policies that exhibit good performance against data but without any provable guarantees on steady-state performance. The recent paper [24] on RL for optimal primary frequency control proposes a way to address the stability issue by identifying a set of properties that make a control design stabilizing and then restricting the search space of neural network-based controllers. In this paper, we extend this idea to achieve provable

guarantees on frequency restoration and economic efficiency at the steady-state.

With the aim of filling the gap between the optimization of steady-state and transient frequency control performance, we propose to unify these two perspectives by integrating RL into DAI control. In our approach, we employ RL to seek the optimal control policy in terms of the transient performance as a map from the integral variables of DAI to the controllable power injections, while DAI inherently ensures the optimal steady-state performance. This results in nonlinear optimal frequency controllers which we term *RL-DAI control* that generalizes the standard DAI approach. Specifically, the generalization is manifold:

- Unlike the classic DAI control that only addresses quadratic operational cost functions, RL-DAI control admits any strictly convex cost functions whose gradients are identical up to heterogeneous scaling factors;
- Unlike the standard RL methods that do not provide stability guarantees, RL-DAI control encodes the stabilization requirement on the control policy as mild conditions on its continuity and monotonicity. Such conditions can be easily realized by an ingenious parameterization of the control policy as a monotonic neural network, which is trained by a RL algorithm based on a recurrent neural network (RNN);
- RL-DAI control jointly optimizes both steady-state and transient frequency control performance by leveraging the added degrees of freedom in tuning parameters that characterize the nonlinear control policy.

The rest of this paper is organized as follows. Section II describes the power system model and formalizes the optimal frequency control problem. Section III describes the proposed generalized DAI control and shows how it guarantees the asymptotic convergence of the closed-loop system to the equilibrium that achieves the steady-state performance objectives. Section IV illustrates how to integrate RL with DAI control such that the transient performance can be optimized without jeopardizing stability. Section V validates our results through detailed simulations. Section VI gathers our conclusions and outlook.

## II. MODELLING APPROACH AND PROBLEM STATEMENT

In this section, we describe the power system model used in this paper for analysis and the frequency control problem we aim to address.

### A. POWER SYSTEM MODEL

We consider<sup>1</sup> a  $n$ -bus power system whose topology can be characterized by a weighted undirected connected graph  $(\mathcal{V}, \mathcal{E})$ , where buses indexed by  $i, j \in \mathcal{N} := \{1, \dots, n\}$  are linked through transmission lines denoted by unordered pairs  $\{i, j\} \in \mathcal{E} \subset \{\{i, j\} \mid i, j \in \mathcal{N}, i \neq j\}$ .

<sup>1</sup>Throughout this paper, vectors are denoted in lower case bold and matrices are denoted in upper case bold, while scalars are unbolded, unless otherwise specified. Also,  $\mathbf{1}_n, \mathbf{0}_n \in \mathbb{R}^n$  denote the vectors of all ones and all zeros, respectively.

The set of buses  $\mathcal{N}$  is a disjoint union of two subsets: the buses with generators<sup>2</sup>  $\mathcal{G}$  and buses that are purely loads  $\mathcal{L}$ . Generator buses are described with differential equations, while load buses characterizing frequency-responsive devices such as frequency-sensitive loads and inverter-based resources performing droop control are governed by algebraic equations. More precisely, given the net power injection  $p_i$  (in p.u.) on each bus, the dynamics of the voltage angle  $\theta_i$  (in rad/s) and the frequency deviation  $\omega_i$  (in p.u.) from the nominal frequency  $f_0$  (50 Hz or 60 Hz depending on the particular system) evolve according to

$$\dot{\theta}_i = 2\pi f_0 \omega_i, \quad \forall i \in \mathcal{N}, \quad (1a)$$

$$m_i \dot{\omega}_i = -\alpha_i \omega_i - \sum_{j=1}^n v_i v_j B_{ij} \sin(\theta_i - \theta_j) + p_i + u_i, \quad \forall i \in \mathcal{G}, \quad (1b)$$

$$0 = -\alpha_i \omega_i - \sum_{j=1}^n v_i v_j B_{ij} \sin(\theta_i - \theta_j) + p_i + u_i, \quad \forall i \in \mathcal{L}, \quad (1c)$$

where the parameters are defined as:  $m_i := 2H_i > 0$  — generator inertia constant (in s),  $\alpha_i > 0$  — frequency sensitivity coefficient from generator or load (in p.u.) depending on whether  $i \in \mathcal{G}$  or  $i \in \mathcal{L}$ ,  $v_i > 0$  — voltage magnitude (in p.u.), and  $B_{ij}$  — susceptance (in p.u.) satisfying  $B_{ij} = B_{ji} > 0$  if  $\{i, j\} \in \mathcal{E}$  and  $B_{ij} = B_{ji} = 0$  if  $\{i, j\} \notin \mathcal{E}$ . Finally,  $u_i$  is a controllable power injection to be designed for frequency control.

*Remark 1 (Model assumptions):* The model in (1) implicitly makes the following assumptions that are well-justified for frequency control on transmission networks [8], [25].

- Lossless lines:  $\forall \{i, j\} \in \mathcal{E}$ , the line resistance is zero.
- Constant voltage profile:  $\forall i \in \mathcal{N}$ , the bus voltage magnitude  $v_i$  is constant.
- Decoupling: Reactive power flows do not affect bus voltage angles.
- $\forall \{i, j\} \in \mathcal{E}$ , the equilibrium bus voltage angle difference is within  $\pm\pi/2$ , i.e.,  $\forall \{i, j\} \in \mathcal{E}$ ,  $|\theta_i^* - \theta_j^*| \in [0, \pi/2)$ .

## B. FREQUENCY CONTROL

The basic goal of frequency control is to keep the frequency of the power system close to its nominal value. We now illustrate this in more detail by following the same line of argument as in [6], [9], [11], [26].

Since the frequency dynamics depend only on the phase angle differences, for easy of analysis, we express the dynamics in center-of-inertia coordinates [11], [26] by making the following change of variables:  $\delta_i := \theta_i - (\sum_{j=1}^n \theta_j)/n$ ,  $\forall i \in \mathcal{N}$ . Then, the dynamics in (1) can be rewritten and stacked into a vector form as:

$$\dot{\delta} = 2\pi f_0 \left( I_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \right) \omega, \quad (2a)$$

$$\mathbf{M} \dot{\omega}_{\mathcal{G}} = -\mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}} - \nabla_{\mathcal{G}} U(\delta) + \mathbf{p}_{\mathcal{G}} + \mathbf{u}_{\mathcal{G}}, \quad (2b)$$

$$\mathbb{0}_{|\mathcal{L}|} = -\mathbf{A}_{\mathcal{L}} \omega_{\mathcal{L}} - \nabla_{\mathcal{L}} U(\delta) + \mathbf{p}_{\mathcal{L}} + \mathbf{u}_{\mathcal{L}}, \quad (2c)$$

with

$$U(\delta) := -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n v_i v_j B_{ij} \cos(\delta_i - \delta_j),$$

where  $\delta := (\delta_i, i \in \mathcal{N}) \in \mathbb{R}^n$ ,  $\omega := (\omega_i, i \in \mathcal{N}) \in \mathbb{R}^n$ ,  $\mathbf{p} := (p_i, i \in \mathcal{N}) \in \mathbb{R}^n$ ,  $\mathbf{u} := (u_i, i \in \mathcal{N}) \in \mathbb{R}^n$ ,  $\mathbf{M} := \text{diag}(m_i, i \in \mathcal{N}) \in \mathbb{R}^{n \times n}$ ,  $\mathbf{A}_{\mathcal{G}} := \text{diag}(\alpha_i, i \in \mathcal{G}) \in \mathbb{R}^{|\mathcal{G}| \times |\mathcal{G}|}$ , and  $\mathbf{A}_{\mathcal{L}} := \text{diag}(\alpha_i, i \in \mathcal{L}) \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{L}|}$ . Here, a vector with a set as subscript denotes the subvector composed only of the elements from that set, e.g.,  $\omega_{\mathcal{G}} := (\omega_i, i \in \mathcal{G}) \in \mathbb{R}^{|\mathcal{G}|}$ .<sup>3</sup>

The equilibria of (2) satisfy

$$\omega^* = \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \omega^*, \quad (3a)$$

$$\mathbb{0}_{|\mathcal{G}|} = -\mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}}^* - \nabla_{\mathcal{G}} U(\delta^*) + \mathbf{p}_{\mathcal{G}} + \mathbf{u}_{\mathcal{G}}^*, \quad (3b)$$

$$\mathbb{0}_{|\mathcal{L}|} = -\mathbf{A}_{\mathcal{L}} \omega_{\mathcal{L}}^* - \nabla_{\mathcal{L}} U(\delta^*) + \mathbf{p}_{\mathcal{L}} + \mathbf{u}_{\mathcal{L}}^*. \quad (3c)$$

Clearly, (3a) implies that there exists a scalar  $\omega^*$  such that  $\omega^* = \mathbb{1}_n \omega^*$ . Then, (3b) and (3c) can be combined into

$$\mathbb{0}_n = -\mathbf{A} \mathbb{1}_n \omega^* - \nabla U(\delta^*) + \mathbf{p} + \mathbf{u}^* \quad (4)$$

with  $\mathbf{A} := \text{diag}(\alpha_i, i \in \mathcal{N}) \in \mathbb{R}^{n \times n}$ . Moreover, after premultiplying (4) by  $\mathbb{1}_n^T$ , we can characterize  $\omega^*$  as

$$\omega^* = \frac{\sum_{i=1}^n p_i + \sum_{i=1}^n u_i^*}{\sum_{i=1}^n \alpha_i}, \quad (5)$$

where we have used the zero net power flow balance  $\mathbb{1}_n^T \nabla U(\delta) = 0$  of a lossless power system.

Observe from (5) that, in the absence of frequency control, the system undergoing power disturbances  $\mathbf{p}$  will synchronize to a nonzero frequency deviation, i.e.,  $\omega^* \neq 0$ , since  $\omega^* = 0$  only if  $\sum_{i=1}^n p_i + \sum_{i=1}^n u_i^* = 0$ . Therefore, the main goal of frequency control is to regulate the frequency such that  $\omega^* = 0$  by providing appropriate controllable power injections  $\mathbf{u}$  to meet power disturbances  $\mathbf{p}$ . Note that the existence of equilibrium points  $(\delta^*, \mathbb{1}_n \omega^*)$  such that (4) holds for  $\omega^* = 0$  is equivalent to the feasibility of the power flow equation

$$\nabla U(\delta^*) = \mathbf{p} + \mathbf{u}^*, \quad (6)$$

which is a standing assumption of this paper. In fact, if the feasibility of (6) holds, then there are multiple choices of  $\mathbf{u}$  to meet (6). Consequently, we seek to select among the possible candidates by optimizing certain performance metrics, which we describe next.

## C. PERFORMANCE ASSESSMENT

For the design of frequency control strategies, not only frequency performance but also economic factors must be taken into account. Moreover, the secure and efficient operation of power systems relies on properly controlled frequency and cost in both slow and fast timescales. Thus, we now introduce the frequency and economic performance metrics used in this paper based on different timescales.

<sup>2</sup>The generator buses can have collocated loads.

<sup>3</sup>For a finite set  $\mathcal{G}$ ,  $|\mathcal{G}|$  denotes the cardinality of  $\mathcal{G}$ .

## 1) STEADY-STATE PERFORMANCE METRICS

Our control objective in the long run is to achieve the nominal frequency restoration, i.e.,  $\omega^* = \mathbb{0}_n$ , as well as the lowest steady-state aggregate operational cost  $C(\mathbf{u}^*) := \sum_{i=1}^n C_i(u_i^*)$ , where the cost function  $C_i(u_i)$  quantifies either the generation cost on a generator bus  $i$  or the user disutility on a load bus for contributing  $u_i$ . This results in the following constrained optimization problem called *optimal steady-state economic dispatch problem*:

$$\min_{\mathbf{u}^*} C(\mathbf{u}^*) := \sum_{i=1}^n C_i(u_i^*) \quad (7a)$$

$$\text{s.t.} \quad \sum_{i=1}^n p_i + \sum_{i=1}^n u_i^* = 0, \quad (7b)$$

where (7b) is a necessary constraint on the steady-state controllable power injections  $\mathbf{u}^*$  in order to achieve  $\omega^* = \mathbb{0}_n$  as discussed in Section II-B. Here, we adopt the standard assumption that the cost function  $C_i(u_i)$  is strictly convex and continuously differentiable [9], [13] with respect to  $u_i$ . Then, the optimization problem (7) has a convex objective function and an affine equality constraint. Thus, by the Karush-Kuhn-Tucker conditions [27, Chapter 5.5.3],  $\mathbf{u}^*$  is the unique minimizer of (7) if and only if it ensures *identical marginal costs* [6], [9], [10], [16], [26], i.e.,

$$\nabla C_i(u_i^*) = \nabla C_j(u_j^*), \quad \forall i, j \in \mathcal{N}. \quad (8)$$

## 2) TRANSIENT PERFORMANCE METRICS

Following sudden major power disturbances, the transient frequency dip can be large in the first few seconds, especially in low-inertia power systems. This may trigger undesired protection measures and even cause cascading failures. Thus, besides the steady-state performance, one should also pay attention to the transient frequency performance with moderate economic cost. With this aim, we define the following transient performance metrics evaluated along the trajectories of the system (1):

- *Frequency Nadir* is the maximum frequency deviation from the nominal frequency on each bus during the transient response, i.e.,

$$\|\omega_i\|_\infty := \max_{t \geq 0} |\omega_i(t)|.$$

- *Finite horizon economic cost* measures the average cost on a generator or load bus for its participation in frequency control during a time horizon  $T$ , i.e.,

$$\bar{C}_{i,T} := \frac{1}{T} \int_0^T C_i(u_i(t)) dt.$$

Then the *optimal transient frequency control problem* becomes:

$$\min_{\mathbf{u}} \sum_{i \in \mathcal{G}} \|\omega_i\|_\infty + \rho \sum_{i=1}^n \bar{C}_{i,T} \quad \text{s.t.} \quad \mathbf{u} \text{ stabilize (1)}, \quad (9)$$

where  $\rho > 0$  is the coefficient for tradeoff between the frequency performance and the economic cost. Note that we impose the stability requirement on  $\mathbf{u}$  as a hard constraint in the optimization problem (9), which will play a pivotal role in its design.

## D. REINFORCEMENT LEARNING FOR FREQUENCY CONTROL

Our goal is to design an optimal stabilizing controller that brings the system (1) to an equilibrium that restores the nominal frequency, i.e.,  $\omega^* = \mathbb{0}_n$ , and solves the optimal steady-state economic dispatch problem (7), while solving the optimal transient frequency control problem (9) at the same time. A good starting point to achieve our steady-state control goal is the well-known DAI control. However, it is not straightforward how to also optimize the transient performance. This is hard to be done purely by conventional optimization methods since power systems are nonlinear and power disturbances are unknown. Therefore, we would like to integrate RL into DAI to jointly optimize steady-state and transient performance.

## III. GENERALIZED DISTRIBUTED AVERAGING-BASED INTEGRAL CONTROL

We start this section by briefly reviewing distributed averaging-based integral (DAI) control and then generalizing it to account for nonlinear control laws. We show that the closed-loop system is stable as long as the nonlinearity satisfies conditions on continuity and monotonicity. Interestingly, the steady-state performance objectives are still achieved under our proposed generalized DAI even for nonquadratic cost functions provided that the gradients of the cost functions on individual buses satisfy certain condition.

### A. REVIEW AND GENERALIZATION OF DAI

DAI control [6], [10], [11], [16] is an established choice of frequency control strategy to meet the steady-state performance objectives discussed in Section II-C. However, most works restrict the cost functions to be quadratic, i.e.,

$$C_i(u_i) = \frac{1}{2} c_i u_i^2, \quad \forall i \in \mathcal{N}, \quad (10)$$

where  $c_i > 0$  is the cost coefficient. With the cost functions given by (10), the power injections  $\mathbf{u}$  in the system (2) under DAI are

$$u_i(s_i) = k_i s_i, \quad \forall i \in \mathcal{N}, \quad (11a)$$

with

$$\dot{s}_i = -2\pi f_0 \omega_i - c_i \sum_{j=1}^n Q_{ij} (c_i u_i - c_j u_j), \quad (11b)$$

where  $k_i > 0$  is a tunable control gain and  $Q_{ij} \geq 0$  is the weight associated with an undirected connected communication graph  $(\mathcal{V}, \mathcal{E}_Q)$  (not necessarily the same as the physical system topology graph  $(\mathcal{V}, \mathcal{E})$ ) such that  $Q_{ij} = Q_{ji} > 0$  if buses  $i$  and  $j$  communicate, i.e.,  $\{i, j\} \in \mathcal{E}_Q \subset \{\{i, j\} \mid i, j \in$

$\mathcal{N}, i \neq j$ }, and  $Q_{ij} = Q_{ji} = 0$  otherwise,  $\forall i, j \in \mathcal{N}$ . The core idea behind DAI is that the control dynamics can settle down only if the nominal frequency restoration, i.e.,  $\omega_i = 0$ ,  $\forall i \in \mathcal{N}$ , and the identical marginal costs, i.e.,  $\nabla C_i(u_i) = c_i u_i = c_j u_j = \nabla C_j(u_j)$ ,  $\forall i, j \in \mathcal{N}$ , are simultaneously achieved, which ensures that  $\omega^* = \mathbf{0}_n$  and  $\mathbf{u}^*$  is the solution of the optimal steady-state economic dispatch problem (7).

However, DAI in (11) does not address the transient behavior of the system. Thus, trajectories that eventually reach the equilibrium may have poor transient performances. Therefore, it is desirable to modify DAI so that we have better flexibility in improving the transient performance. Inspired by the fact that the control policy  $u_i(s_i)$  of DAI in (11a) is directly proportional to the integral variable  $s_i$ , we generalize DAI by allowing  $u_i(s_i)$  to be any Lipschitz continuous and strictly increasing function of  $s_i$ , with  $u_i(0) = 0$ . This extends  $u_i(s_i)$  from a linear function to any nonlinear function satisfying these mild requirements. The added nonlinearity provides us with a good opportunity to solve the optimal transient frequency control problem (9) while remaining the optimal steady-state performance.

*Assumption 1 (Continuity and monotonicity of  $\mathbf{u}$ ):*  $\forall i \in \mathcal{N}$ , the function  $u_i(\cdot) : \mathbb{R} \mapsto \mathbb{R}$  is Lipschitz continuous and strictly increasing with  $u_i(0) = 0$ .

In our treatment, we also relax the restriction to quadratic cost functions by allowing DAI to handle any cost functions satisfying the following assumption.

*Assumption 2 (Scaled cost gradient functions):* There exists a strictly convex and continuously differentiable function  $C_o(\cdot) : \mathbb{R} \mapsto \mathbb{R}$  and a group of positive scaling factors  $\zeta := (\zeta_i, i \in \mathcal{N}) \in \mathbb{R}_+^n$  such that

$$\nabla C_i(\cdot) = \nabla C_o(\zeta_i \cdot), \quad \forall i \in \mathcal{N}. \quad (12)$$

Assumption 2 admits any strictly convex cost functions whose gradients are identical up to heterogeneous scaling factors. To look in greater detail at how this generalizes beyond quadratic cost functions, we provide some examples of common strictly convex functions that satisfy Assumption 2. If the cost functions on individual buses are:

- *power functions with positive even integer powers:*

$$C_i(u_i) = \frac{c_i}{r} u_i^r + b_i, \quad (13)$$

where  $c_i > 0$  and  $r$  is a positive even integer, then

$$\nabla C_i(u_i) = c_i u_i^{r-1}, \quad \forall i \in \mathcal{N}. \quad (14)$$

Thus, we can choose

$$C_o(\cdot) = \frac{1}{r} (\cdot)^r \quad \text{and} \quad \zeta_i = c_i^{\frac{1}{r-1}}, \quad (15)$$

which satisfies,  $\forall i \in \mathcal{N}$ ,

$$\nabla C_i(u_i) = c_i u_i^{r-1} = \left( c_i^{\frac{1}{r-1}} u_i \right)^{r-1} = (\zeta_i u_i)^{r-1} = \nabla C_o(\zeta_i u_i).$$

- *arbitrary strictly convex and continuously differentiable functions that are identical up to constant terms:*

$$C_i(u_i) = C_o(u_i) + b_i,$$

then  $\nabla C_i(u_i) = \nabla C_o(u_i) = \nabla C_o(\zeta_i u_i)$  with  $\zeta_i = 1$ ,  $\forall i \in \mathcal{N}$ .

Clearly, the quadratic cost functions in (10) belong to the first scenario since (13) reduces exactly to (10) if  $r = 2$  and  $b_i = 0$ , which validates our generalization through Assumption 2.

*Remark 2 (Cost function extensions and limitations):* Although the generation costs of conventional synchronous generators are usually considered as quadratic functions with respect to their mechanical power output, it need not be the case for inverter-based resources like solar [28] and wind [29] power as well as electrochemical batteries [30]. Thus, it is important to extend the class of cost functions that DAI can handle, especially when non-conventional generation participates in frequency control. Admittedly, our extension via Assumption 2 is still limited, whose further relaxation is a possible direction of future research.

Now, under Assumptions 1 and 2, we are ready to propose the *generalized DAI* as follows:

$$\dot{s}_i = -2\pi f_0 \omega_i - \zeta_i \sum_{j=1}^n Q_{ij} (\nabla C_i(u_i(s_i)) - \nabla C_j(u_j(s_j))). \quad (16)$$

Although not obvious at first sight, the generalized DAI in (16) preserves the ability of the classic DAI to restore the nominal frequency and solves the optimal steady-state economic dispatch problem (7) for cost functions satisfying Assumption 2. In addition, the more general nonlinear form of  $\mathbf{u}$  introduced in Assumption 1 provides us with more degrees of freedom to better deal with the optimal transient frequency control problem (9). All of these statements would become clear as the analysis unfolds.

After combining (2) and (16), we can write the overall closed-loop system dynamics under our proposed generalized DAI compactly as

$$\dot{\delta} = 2\pi f_0 \left( I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \boldsymbol{\omega}, \quad (17a)$$

$$\mathbf{M} \dot{\boldsymbol{\omega}}_G = -\mathbf{A}_G \boldsymbol{\omega}_G - \nabla_G U(\boldsymbol{\delta}) + \mathbf{p}_G + \mathbf{u}_G(\mathbf{s}), \quad (17b)$$

$$\mathbf{0}_{|\mathcal{L}|} = -\mathbf{A}_L \boldsymbol{\omega}_L - \nabla_L U(\boldsymbol{\delta}) + \mathbf{p}_L + \mathbf{u}_L(\mathbf{s}), \quad (17c)$$

$$\dot{\mathbf{s}} = -2\pi f_0 \boldsymbol{\omega} - \mathbf{Z} \mathbf{L}_Q \nabla C(\mathbf{u}(\mathbf{s})), \quad (17d)$$

where  $\mathbf{s} := (s_i, i \in \mathcal{N}) \in \mathbb{R}^n$ ,  $\mathbf{Z} := \text{diag}(\zeta_i, i \in \mathcal{N}) \in \mathbb{R}^{n \times n}$ ,  $\mathbf{L}_Q := [L_{Q,ij}] \in \mathbb{R}^{n \times n}$  is the Laplacian matrix associated with the communication graph whose  $ij$ th element is  $L_{Q,ij} := -Q_{ij}$  if  $i \neq j$  and  $L_{Q,ij} := \sum_{j'=1, j' \neq i}^n Q_{ij'}$  if  $i = j$ ,  $\forall i, j \in \mathcal{N}$ , and  $\mathbf{u}(\mathbf{s}) := (u_i(s_i), i \in \mathcal{N}) \in \mathbb{R}^n$  with each  $u_i(s_i)$  satisfying Assumption 1.

We conclude this subsection by stating a property of the inverse functions of the cost gradient functions on individual

buses, which is a consequence of Assumption 2. It is useful later for the stability analysis of the closed-loop system (17).

*Lemma 1 (Identical scaled controllable power injections):* If Assumption 2 holds, then

$$\xi_i \nabla C_i^{-1}(\cdot) = \nabla C_0^{-1}(\cdot), \quad \forall i \in \mathcal{N}, \quad (18)$$

where  $\nabla C_i^{-1}(\cdot)$  and  $\nabla C_0^{-1}(\cdot)$  denote the inverse functions of  $\nabla C_i(\cdot)$  and  $\nabla C_0(\cdot)$ , respectively.

*Proof:* See the Appendix A.  $\blacksquare$

Provided that the cost functions on individual buses satisfy Assumption 2, Lemma 1 indicates that, for any given marginal price, each generator or load contributes the same amount of controllable power injection up to a scaling factor.

*Remark 3 (Existence and monotonicity of  $\nabla C_i^{-1}(\cdot)$  and  $\nabla C_0^{-1}(\cdot)$ ):* Note that the existence of  $\nabla C_i^{-1}(\cdot)$  is guaranteed by the strict convexity of  $C_i(\cdot)$ . More precisely, by [31, Theorem 2.14], the strict convexity of  $C_i(\cdot)$  implies that  $\nabla C_i(\cdot)$  is a strictly increasing function. According to [32, Chapter 12.9], this further implies that  $\nabla C_i(\cdot)$  has an inverse function  $\nabla C_i^{-1}(\cdot)$  which is also strictly increasing. The same argument holds for  $\nabla C_0^{-1}(\cdot)$ .

## B. CLOSED-LOOP EQUILIBRIUM ANALYSIS

The first result we show is that the steady-state performance objectives are still achieved under the generalized DAI. This is captured by the following theorem.

*Theorem 1 (Closed-loop equilibrium):* Suppose Assumptions 1 and 2 hold. Assume  $\forall \{i, j\} \in \mathcal{E}$ ,  $|\delta_i^* - \delta_j^*| \in [0, \pi/2)$ .<sup>4</sup> Then the equilibrium  $(\delta^*, \omega^*, s^*)$  of the closed-loop system (17) is a unique point satisfying

$$\omega^* = \mathbf{0}_n, \quad (19a)$$

$$\nabla U(\delta^*) = \mathbf{p} + \mathbf{u}(s^*), \quad (19b)$$

$$\mathbf{u}(s^*) = \nabla C_0^{-1}(\gamma) \mathbf{Z}^{-1} \mathbf{1}_n, \quad (19c)$$

with  $\gamma$  uniquely determined by

$$\nabla C_0^{-1}(\gamma) = -\frac{\sum_{i=1}^n P_i}{\sum_{i=1}^n \xi_i^{-1}}. \quad (20)$$

*Proof:* The equilibrium analysis is similar to the one presented in Section II-B, except that the effect of the integral variables  $s$  introduced by the generalized DAI should be considered. Hence, in steady-state, (17) yields

$$\mathbf{0}_n = -\mathbf{A} \mathbf{1}_n \omega^* - \nabla U(\delta^*) + \mathbf{p} + \mathbf{u}(s^*), \quad (21a)$$

$$2\pi f_0 \mathbf{1}_n \omega^* = -\mathbf{Z} \mathbf{L}_Q \nabla C(\mathbf{u}(s^*)), \quad (21b)$$

where we have used the equilibria characterizations  $\omega^* = \mathbf{1}_n \omega^*$  and (4).

We then proceed to investigate the equilibrium by following a similar argument as in [16, Lemma 4.2]. Premultiplying (21b) by  $\mathbf{1}_n^T \mathbf{Z}^{-1}$  yields

$$2\pi f_0 \mathbf{1}_n^T \mathbf{Z}^{-1} \mathbf{1}_n \omega^* = -\mathbf{1}_n^T \mathbf{L}_Q \nabla C(\mathbf{u}(s^*)) = 0, \quad (22)$$

<sup>4</sup>This is the counterpart of the model assumption,  $\forall \{i, j\} \in \mathcal{E}$ ,  $|\theta_i^* - \theta_j^*| \in [0, \pi/2)$  in Remark 1.

where the second equality is due to the property of the Laplacian matrix [33] that  $\mathbf{1}_n^T \mathbf{L}_Q = \mathbf{0}_n^T$ . Note that  $\mathbf{1}_n^T \mathbf{Z}^{-1} \mathbf{1}_n > 0$  since  $\mathbf{Z}^{-1} \succ 0$  by construction. Thus, (22) implies that  $\omega^* = 0$ . Then, (21b) becomes  $\mathbf{Z} \mathbf{L}_Q \nabla C(\mathbf{u}(s^*)) = \mathbf{0}_n$ , which indicates that  $\nabla C(\mathbf{u}(s^*)) \in \text{range}(\mathbf{1}_n)$ , i.e.,  $\nabla C(\mathbf{u}(s^*)) = \gamma \mathbf{1}_n$  for some constant  $\gamma$ . Thus,  $\mathbf{u}(s^*) = (\nabla C_i^{-1}(\gamma), i \in \mathcal{N})$ . Note that  $\mathbf{Z} \mathbf{u}(s^*) = (\xi_i \nabla C_i^{-1}(\gamma), i \in \mathcal{N}) = \nabla C_0^{-1}(\gamma) \mathbf{1}_n$  by Lemma 1, which further implies that  $\mathbf{u}(s^*) = \nabla C_0^{-1}(\gamma) \mathbf{Z}^{-1} \mathbf{1}_n$ .

Now, applying  $\omega^* = 0$  and  $\mathbf{u}(s^*) = \nabla C_0^{-1}(\gamma) \mathbf{Z}^{-1} \mathbf{1}_n$  to (21a) yields

$$\nabla U(\delta^*) = \mathbf{p} + \nabla C_0^{-1}(\gamma) \mathbf{Z}^{-1} \mathbf{1}_n. \quad (23)$$

Then, premultiplying (23) by  $\mathbf{1}_n^T$  yields the equation that determines  $\gamma$  as shown in (20), where  $\mathbf{1}_n^T \nabla U(\delta) = 0$  is used again. We can show that the solution  $\gamma$  to (20) is unique by way of contradiction. Suppose that both  $\gamma$  and  $\tilde{\gamma}$  satisfy (20), where  $\gamma \neq \tilde{\gamma}$ . Then,  $\nabla C_0^{-1}(\gamma) - \nabla C_0^{-1}(\tilde{\gamma}) = 0$ , which after being multiplied by  $(\gamma - \tilde{\gamma})$  yields

$$0 = (\gamma - \tilde{\gamma}) (\nabla C_0^{-1}(\gamma) - \nabla C_0^{-1}(\tilde{\gamma})) > 0, \quad (24)$$

where the inequality is due to the fact that  $\nabla C_0^{-1}(\cdot)$  is strictly increasing as discussed in Remark 3. Clearly, (24) is a contradiction. Thus,  $\gamma$  is unique.

It remains to show that the equilibrium is unique. We focus on the uniqueness of  $s^*$ . By way of contradiction, suppose that both  $s^*$  and  $s^*$  satisfy (19c), where  $s^* \neq s^*$ . Then,  $\mathbf{u}(s^*) - \mathbf{u}(s^*) = \mathbf{0}_n$ , which after being multiplied by  $(s^* - s^*)^T$  yields

$$\begin{aligned} 0 &= (s^* - s^*)^T (\mathbf{u}(s^*) - \mathbf{u}(s^*)) \\ &= \sum_{i=1}^n (s_i^* - s_i^*) (u_i(s_i^*) - u_i(s_i^*)) > 0, \end{aligned} \quad (25)$$

where the inequality results from our requirement that  $u_i(s_i)$  is strictly increasing with respect to  $s_i$ . Clearly, (25) is a contradiction. Thus,  $s^*$  is unique. By [34, Lemma 1], the solution  $\delta^*$  to (23) is unique. This concludes the proof of the uniqueness of the equilibrium.  $\blacksquare$

Theorem 1 verifies that the generalized DAI preserves the steady-state performance of the normal DAI. It forces the closed-loop system (17) to settle down at a unique equilibrium point where the frequency is nominal, i.e.,  $\omega^* = \mathbf{0}_n$ , and the controllable power injections  $\mathbf{u}^*$  meet the identical marginal cost requirement (8) since,  $\forall i \in \mathcal{N}$ ,

$$\nabla C_i(u_i^*) = \nabla C_i(u_i(s_i^*)) \stackrel{(19c)}{=} \nabla C_i(\nabla C_0^{-1}(\gamma) \xi_i^{-1})$$

$$\stackrel{(18)}{=} \nabla C_i(\xi_i \nabla C_i^{-1}(\gamma) \xi_i^{-1}) = \nabla C_i(\nabla C_i^{-1}(\gamma)) = \gamma.$$

Thus, the objectives of nominal steady-state frequency and optimal steady-state economic dispatch for a broader range of cost functions are both achieved.

## C. LYAPUNOV STABILITY ANALYSIS

Having characterized the equilibrium point and confirmed the steady-state performance of the closed-loop system (17) under the generalized DAI, we are now ready to investigate the

system stability by performing Lyapunov stability analysis. More precisely, the stability under the generalized DAI can be certified by finding a well-defined ‘‘weak’’ Lyapunov function that is nonincreasing along the trajectories of the closed-loop system (17). The main result of this whole subsection is presented below, whose proof is enabled by a sequence of smaller results that we discuss next.

*Theorem 2 (Asymptotic stability):* Under Assumptions 1 and 2, any trajectory of (17) starting from the neighborhood  $\{(\delta, \omega, s) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \mid |\delta_i - \delta_j| \in [0, \pi/2), \forall \{i, j\} \in \mathcal{E}\}$  of the equilibrium characterized by (19) with  $|\delta_i^* - \delta_j^*| \in [0, \pi/2), \forall \{i, j\} \in \mathcal{E}$ , converges asymptotically to the equilibrium.

First note that the algebraic equation (17c) fully determines  $\omega_{\mathcal{L}}$  as a function of  $\delta$  and  $s$ , i.e.,

$$\omega_{\mathcal{L}} = A_{\mathcal{L}}^{-1} (-\nabla_{\mathcal{L}} U(\delta) + p_{\mathcal{L}} + u_{\mathcal{L}}(s)), \quad (26)$$

which means that we only need to explicitly consider the evolution of  $(\delta, \omega_{\mathcal{G}}, s)$ . Thus, in what follows, we only focus on the state component  $(\delta, \omega_{\mathcal{G}}, s)$  for a neighborhood  $\mathcal{X} := \{(\delta, \omega_{\mathcal{G}}, s) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{G}|} \times \mathbb{R}^n \mid |\delta_i - \delta_j| \in [0, \pi/2), \forall \{i, j\} \in \mathcal{E}\}$  of the equilibrium  $(\delta^*, \omega_{|\mathcal{G}|}, s^*)$ . The next lemma formalizes this thought by showing that the distance from the whole state  $(\delta, \omega, s)$  to  $(\delta^*, \omega_n, s^*)$  is lower and upper bounded by the distance from the partial state  $(\delta, \omega_{\mathcal{G}}, s)$  to  $(\delta^*, \omega_{|\mathcal{G}|}, s^*)$ .

*Lemma 2 (Bounds on whole state distance):* Let  $\|\cdot\|_2$  denote the Euclidean norm. Under Assumptions 1 and 2, there exists some constant  $\nu > 0$  such that,  $\forall (\delta, \omega_{\mathcal{G}}, s) \in \mathcal{X}$ ,  $\|\delta - \delta^*\|_2^2 + \|\omega_{\mathcal{G}}\|_2^2 + \|s - s^*\|_2^2 \leq \|\delta - \delta^*\|_2^2 + \|\omega\|_2^2 + \|s - s^*\|_2^2 \leq \nu(\|\delta - \delta^*\|_2^2 + \|\omega_{\mathcal{G}}\|_2^2 + \|s - s^*\|_2^2)$ , where  $(\delta^*, \omega_n, s^*)$  is the equilibrium of the closed-loop system (17) characterized by (19) with  $|\delta_i^* - \delta_j^*| \in [0, \pi/2), \forall \{i, j\} \in \mathcal{E}$ .

*Proof:* See Appendix B. ■

Lemma 2 indicates that it suffices to investigate the evolution of  $(\delta, \omega_{\mathcal{G}}, s)$ . To do so, we first find a well-defined Lyapunov function  $W(\delta, \omega_{\mathcal{G}}, s)$  such that  $W(\delta^*, \omega_{|\mathcal{G}|}, s^*) = 0$  and  $W(\delta, \omega_{\mathcal{G}}, s) > 0, \forall (\delta, \omega_{\mathcal{G}}, s) \in \mathcal{X} \setminus (\delta^*, \omega_{|\mathcal{G}|}, s^*)$ , and then verify  $\dot{W}(\delta, \omega_{\mathcal{G}}, s) \leq 0, \forall (\delta, \omega_{\mathcal{G}}, s) \in \mathcal{X}$ .

Inspired by [9], we begin our construction of a Lyapunov function by defining the following integral function:

$$L(s) := \sum_{i=1}^n \int_0^{s_i} u_i(\xi) d\xi, \quad (27)$$

which clearly satisfies  $\nabla L(s) = u(s)$ . Another useful property related to  $L(s)$  is given in the next lemma.

*Lemma 3 (Strict convexity of  $L(s)$ ):* If Assumption 1 holds, then the function  $L(s)$  defined in (27) is strictly convex.

*Proof:* See the Appendix C. ■

Having characterized the properties of  $L(s)$ , we consider the following Lyapunov function candidate:

$$\begin{aligned} W(\delta, \omega_{\mathcal{G}}, s) := & \pi f_0 \omega_{\mathcal{G}}^T M \omega_{\mathcal{G}} \\ & + U(\delta) - U(\delta^*) - \nabla U(\delta^*)^T (\delta - \delta^*) \\ & + L(s) - L(s^*) - \nabla L(s^*)^T (s - s^*), \end{aligned} \quad (28)$$

where  $(\delta^*, \omega_{|\mathcal{G}|}, s^*)$  corresponds to the unique equilibrium point of the closed-loop system (17) satisfying (19). The next result shows that this is a well-defined Lyapunov function candidate on  $\mathcal{X}$ .

*Lemma 4 (Well-defined Lyapunov function):* Let Assumptions 1 and 2 hold, the function  $W(\delta, \omega_{\mathcal{G}}, s)$  defined in (28) satisfies  $W(\delta^*, \omega_{|\mathcal{G}|}, s^*) = 0$  and  $W(\delta, \omega_{\mathcal{G}}, s) > 0, \forall (\delta, \omega_{\mathcal{G}}, s) \in \mathcal{X} \setminus (\delta^*, \omega_{|\mathcal{G}|}, s^*)$ .

*Proof:* See the Appendix D. ■

Next, we examine the system stability through the derivative of  $W(\delta, \omega_{\mathcal{G}}, s)$  along the trajectories of the closed-loop system (17), whose expression is provided by the next result.

*Lemma 5 (Directional derivative of Lyapunov function):* Let Assumptions 1 and 2 hold. Then the derivative of the Lyapunov function  $W(\delta, \omega_{\mathcal{G}}, s)$  defined in (28) along the trajectories of the closed-loop system (17) is given by

$$\begin{aligned} \dot{W}(\delta, \omega_{\mathcal{G}}, s) & = -2\pi f_0 \omega_{\mathcal{G}}^T A_{\mathcal{G}} \omega_{\mathcal{G}} - u(s)^T \mathbf{ZL}_Q \nabla C(u(s)) \\ & \quad - 2\pi f_0 (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + u_{\mathcal{L}}(s) - u_{\mathcal{L}}(s^*))^T A_{\mathcal{L}}^{-1} \\ & \quad \cdot (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + u_{\mathcal{L}}(s) - u_{\mathcal{L}}(s^*)). \end{aligned} \quad (29)$$

*Proof:* See the Appendix E. ■

Our next step is to show that  $\dot{W}(\delta, \omega_{\mathcal{G}}, s)$  is nonpositive. From (29), we observe that the quadratic terms defined by  $A_{\mathcal{G}}$  and  $A_{\mathcal{L}}^{-1}$  are clearly negative. Thus, it only remains to determine the sign of the cross-term  $u(s)^T \mathbf{ZL}_Q \nabla C(u(s))$  defined by the scaled Laplacian matrix  $\mathbf{ZL}_Q$ . The first thing we notice is that this term as a bilinear form with respect to the scaled Laplacian matrix  $\mathbf{ZL}_Q$  can be expanded as the expression provided below.

*Lemma 6 (Bilinear form on  $\mathbb{R}^n$  for scaled  $\mathbf{L}_Q$ ):*  $\forall \mathbf{x} := (x_i, i \in \mathcal{N}) \in \mathbb{R}^n, \mathbf{y} := (y_i, i \in \mathcal{N}) \in \mathbb{R}^n$ , we have  $\mathbf{x}^T \mathbf{ZL}_Q \mathbf{y} = \sum_{\{i,j\} \in \mathcal{E}_Q} Q_{ij} (y_i - y_j)(\zeta_i x_i - \zeta_j x_j)$ .

*Proof:* See the Appendix F. ■

With the help of Lemma 6, the problem of determining the sign of the cross-term  $u(s)^T \mathbf{ZL}_Q \nabla C(u(s))$  can be solved using the following corollary.

*Corollary 1 (Sign of cross-term):* If Assumption 2 holds, then  $u(s)^T \mathbf{ZL}_Q \nabla C(u(s)) \geq 0$  with equality holding if and only if  $\nabla C(u(s)) \in \text{range}(\mathbf{1}_n)$ .

*Proof:* See the Appendix G. ■

We now have all the elements necessary to establish the stability of the equilibrium as summarized in Theorem 2 by using  $W$  as a Lyapunov function. Therefore, we are ready to finish the proof of Theorem 2. Using the expression of  $\dot{W}(\delta, \omega_{\mathcal{G}}, s)$  in (29), it follows directly from the fact that  $A_{\mathcal{G}} > 0, A_{\mathcal{L}}^{-1} > 0$ , and  $u(s)^T \mathbf{ZL}_Q \nabla C(u(s)) \geq 0$  by Corollary 1 that  $\dot{W}(\delta, \omega_{\mathcal{G}}, s) \leq 0$ . Observe from (29) that  $\dot{W}(\delta, \omega_{\mathcal{G}}, s) \equiv 0$  directly enforces  $\omega_{\mathcal{G}} \equiv \omega_{|\mathcal{G}|}, u(s)^T \mathbf{ZL}_Q \nabla C(u(s)) \equiv 0$ , and  $\nabla_{\mathcal{L}} U(\delta) - u_{\mathcal{L}}(s) \equiv \nabla_{\mathcal{L}} U(\delta^*) - u_{\mathcal{L}}(s^*) = p_{\mathcal{L}}$ . Clearly,  $\omega_{\mathcal{G}} \equiv \omega_{|\mathcal{G}|}$  implies that  $\dot{\omega}_{\mathcal{G}} \equiv \omega_{|\mathcal{G}|}$ . By (17c),  $\nabla_{\mathcal{L}} U(\delta) - u_{\mathcal{L}}(s) \equiv p_{\mathcal{L}}$  ensures that  $\omega_{\mathcal{L}} \equiv \omega_{|\mathcal{L}|}$ . It follows from  $\omega_{\mathcal{G}} \equiv \omega_{|\mathcal{G}|}$  and  $\omega_{\mathcal{L}} \equiv \omega_{|\mathcal{L}|}$  that  $\omega \equiv \omega_n$ , which further indicates that  $\dot{\delta} \equiv \omega_n$  by (17a). Finally, by Corollary 1,  $u(s)^T \mathbf{ZL}_Q \nabla C(u(s)) \equiv 0$  is equivalent

to  $\nabla C(\mathbf{u}(s)) \in \text{range}(\mathbb{1}_n)$  constantly. Hence,  $\mathbf{ZL}_Q \nabla C(\mathbf{u}(s)) \equiv \mathbb{0}_n$ , which together with  $\boldsymbol{\omega} \equiv \mathbb{0}_n$  implies that  $\dot{s} \equiv \mathbb{0}_n$  based on (17d). Thus, we have shown that  $\dot{W}(\boldsymbol{\delta}, \boldsymbol{\omega}_G, \mathbf{s}) \equiv 0$  implies  $\dot{\boldsymbol{\delta}} \equiv \mathbb{0}_n$ ,  $\dot{\boldsymbol{\omega}}_G \equiv \mathbb{0}_{|\mathcal{G}|}$ , and  $\dot{s} \equiv \mathbb{0}_n$ . This means that the largest invariant set contained in the set of points where  $\dot{W}$  vanishes is actually the set of equilibria. Therefore, by the LaSalle invariance principle [35, Theorem 4.4], every trajectory of the closed-loop system (17) starting within  $\mathcal{X}$  converges to the equilibrium set, which by Theorem 1 consists of a unique point satisfying (19). This concludes the proof of local asymptotic stability.

Theorem 2 shows that the closed-loop system (17) under the generalized DAI is locally asymptotically stabilized to the unique equilibrium characterized by Theorem 1, where the steady-state performance objectives are achieved even for a broader range of cost functions beyond quadratic ones.

#### IV. REINFORCEMENT LEARNING FOR OPTIMAL TRANSIENT FREQUENCY CONTROL

In this section, we focus on integrating reinforcement learning (RL) into our proposed generalized DAI for the purpose of further improving the transient performance of the system without jeopardizing its stability. Basically, after parameterizing the control policy  $u_i(s_i)$  that maps the integral variable  $s_i$  of the generalized DAI to the controllable power injection  $u_i$  on each bus as a monotonic neural network, we train those neural networks by a RL algorithm based on a recurrent neural network (RNN). We now discuss this in more detail by mostly following the approach from [24].

##### A. MONOTONIC NEURAL NETWORKS FOR STABILITY GUARANTEE

Recall from Section III that the stability of the closed-loop system (17) under the generalized DAI is ensured by any nonlinear control policy  $u_i(s_i)$  that is a Lipschitz continuous and strictly increasing function of  $s_i$  with  $u_i(0) = 0$  as summarized by Assumption 1, provided that the cost functions on individual buses satisfy Assumption 2. Within this class of stabilizing controllers, we want to find one that has the best transient performance when the frequencies are recovering to their nominal values. In principle, optimizing over all controllers satisfying Assumption 1 is an infinite-dimensional problem. To make the problem tractable, we need to parameterize  $u_i(s_i)$  in some way. Neural networks emerge as a natural candidate due to their universal approximation property [36]. According to [24, Theorem 2], any function of our interest can be approximated accurately enough by a single hidden layer fully-connected neural network with rectified linear unit (ReLU) activation functions for properly designed weights and biases of the neural network. Such a neural network is called a *stacked ReLU monotonic neural network* [24, Lemma 5] whose architecture is shown in Fig. 1.<sup>5</sup>

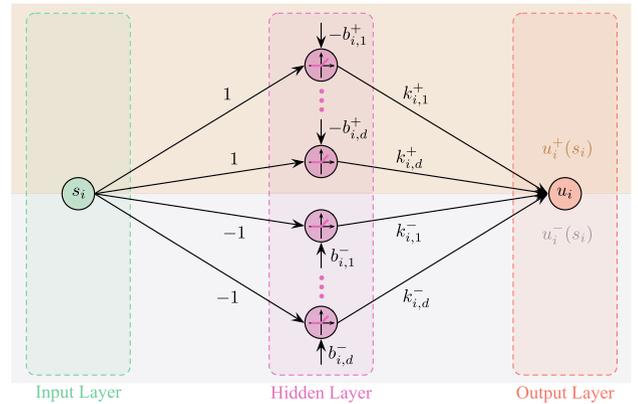


FIGURE 1. Stacked ReLU monotonic neural network.

As illustrated in Fig. 1, each stacked ReLU monotonic neural network is essentially a parameterized mapping from the state  $s_i$  to the control  $u_i$  with parameters  $\mathbf{k}_i^+ := (k_{i,j}^+, j \in \mathcal{D})$ ,  $\mathbf{b}_i^+ := (b_{i,j}^+, j \in \mathcal{D})$ ,  $\mathbf{k}_i^- := (k_{i,j}^-, j \in \mathcal{D})$ , and  $\mathbf{b}_i^- := (b_{i,j}^-, j \in \mathcal{D})$ , where  $\mathcal{D} := \{1, \dots, d\}$  and  $2d$  is the number of neurons in the hidden layer. Here, the vectors  $\mathbf{k}_i^+$  and  $\mathbf{k}_i^-$  are referred to as the “weight” vectors and the vectors  $\mathbf{b}_i^+$  and  $\mathbf{b}_i^-$  are referred to as the “bias” vectors. To see how this neural network architecture could provide the desired properties of the mapping  $u_i(s_i)$ , we now explain the operations it performs in more detail. Firstly, observe from Fig. 1 that the top  $d$  and the bottom  $d$  hidden neurons perform similar operations to produce  $u_i^+(s_i)$  and  $u_i^-(s_i)$ , respectively, where  $u_i(s_i) = u_i^+(s_i) + u_i^-(s_i)$ . Thus, it is natural to investigate the neural network by this partition. For the top part, the state  $s_i$  is transformed linearly as  $(s_i - b_{i,j}^+)$  and then processed by the ReLU in the  $j$ th hidden neuron,  $\forall j \in \mathcal{D}$ . Denote ReLU operator as  $\sigma(\cdot)$  for simplicity. Then, the output of the  $j$ th hidden neuron in the top part becomes  $\sigma(s_i - b_{i,j}^+) = \max(0, s_i - b_{i,j}^+)$ , which is linearly weighted by  $k_{i,j}^+$  before contributing to the final output. Therefore, the contribution from the  $j$ th hidden neuron to the final output is  $k_{i,j}^+ \sigma(s_i - b_{i,j}^+) = k_{i,j}^+ \max(0, s_i - b_{i,j}^+)$ , which implies that the  $j$ th hidden neuron is activated only if the state  $s_i$  exceeds the threshold  $b_{i,j}^+$ , as illustrated in Fig. 2(a). At the output layer, the contributions from all  $d$  hidden neurons in the top part yield

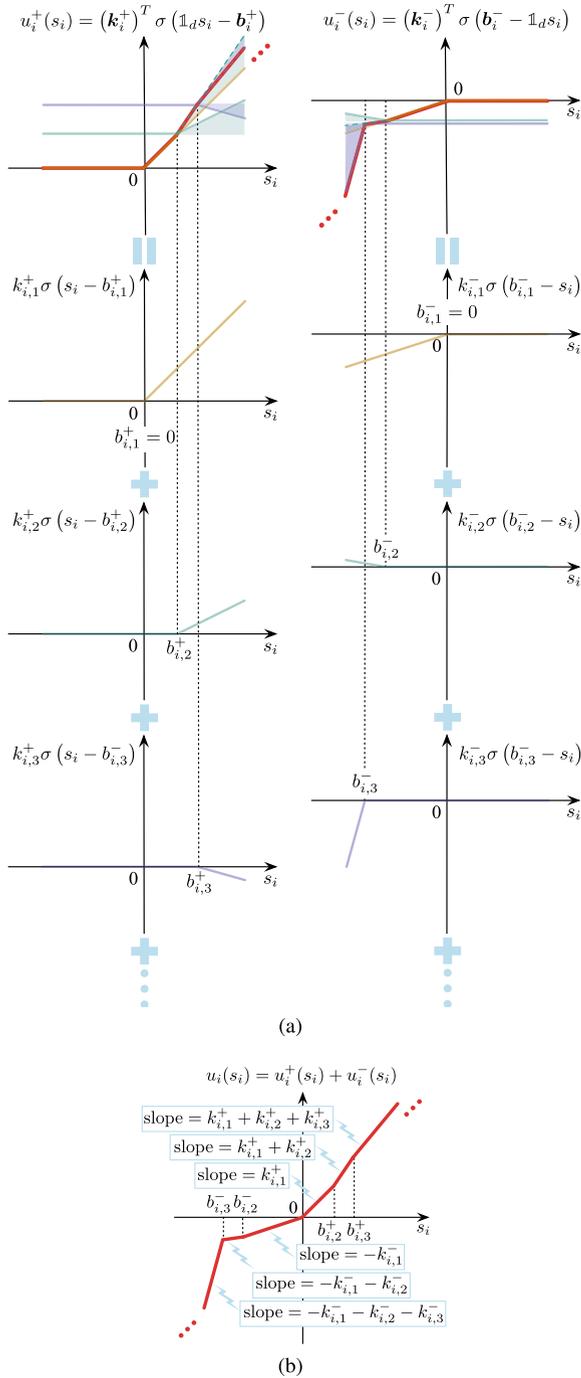
$$u_i^+(s_i) := \sum_{j=1}^d k_{i,j}^+ \sigma(s_i - b_{i,j}^+) = (\mathbf{k}_i^+)^T \sigma(\mathbb{1}_d s_i - \mathbf{b}_i^+).$$

The bottom part follows the similar logic, where the  $d$  hidden neurons produce

$$u_i^-(s_i) := \sum_{j=1}^d k_{i,j}^- \sigma(b_{i,j}^- - s_i) = (\mathbf{k}_i^-)^T \sigma(\mathbf{b}_i^- - \mathbb{1}_d s_i).$$

In order to let  $u_i(s_i)$  satisfy Assumption 1, we need to ensure that its net slope is always positive and finite. This gives rise to the following requirements on weights and biases [24,

<sup>5</sup>ReLU is a widely used activation function in neural networks, which is an element-wise operator applying the function  $\max(0, x)$  to any element  $x$  of its argument.



**FIGURE 2.** Sketch of how hidden neurons in the  $i$ th stacked ReLU monotonic neural network form  $u_i(s_i)$ . (a) Principle of stacking hidden neurons. (b) Piece-wise linear control policy.

Lemma 5]:

$$b_{i,d}^- \leq \dots \leq b_{i,2}^- \leq b_{i,1}^- = 0 = b_{i,1}^+ \leq b_{i,2}^+ \leq \dots \leq b_{i,d}^+, \quad (30a)$$

$$-\infty < \sum_{j=1}^l k_{i,j}^- < 0 < \sum_{j=1}^l k_{i,j}^+ < \infty, \quad \forall l \in \mathcal{D}, \quad (30b)$$

which can be illustrated via Fig. 2. Observe from Fig. 2(a) that the requirements on biases in (30a) assign the top  $d$  and bottom  $d$  hidden neurons to take charge of the control policy  $u_i(s_i)$  for  $s_i \geq 0$  and  $s_i \leq 0$ , respectively, since they are deactivated for the other way around, where a by-product of  $b_{i,1}^- = 0 = b_{i,1}^+$  is  $u_i(0) = 0$ . Observe from Fig. 2(b) that the requirements on weights in (30b) simply ensure that the slope of the piece-wise linear curve  $u_i(s_i)$  is always positive and finite since the  $l$ th piece of  $u_i(s_i)$  for  $s_i \geq 0$  and  $s_i \leq 0$  has the slope  $\sum_{j=1}^l k_{i,j}^+$  and  $-\sum_{j=1}^l k_{i,j}^-$ , respectively,  $\forall l \in \mathcal{D}$ . Thus, under (30), the stacked ReLU monotonic neural network shown in Fig. 1 produces a control policy  $u_i(s_i)$  satisfying Assumption 1 with a shape as sketched in Fig. 2(b).

The inequality constraints in (30) are not trivial to enforce when training the neural networks. A simple trick to address this challenge is to introduce a group of intermediate parameters  $\mu_i^+ := (\mu_{i,j}^+, j \in \mathcal{D}) \in \mathbb{R}^d$ ,  $\chi_i^+ := (\chi_{i,j}^+, j \in \mathcal{D} \setminus \{d\}) \in \mathbb{R}^{d-1}$ ,  $\mu_i^- := (\mu_{i,j}^-, j \in \mathcal{D}) \in \mathbb{R}^d$ , and  $\chi_i^- := (\chi_{i,j}^-, j \in \mathcal{D} \setminus \{d\}) \in \mathbb{R}^{d-1}$  that are unconstrained<sup>6</sup> such that the original parameters  $k_i^+$ ,  $b_i^+$ ,  $k_i^-$ , and  $b_i^-$  are parameterized as

$$k_{i,1}^+ = (\mu_{i,1}^+)^2, \quad k_{i,j}^+ = (\mu_{i,j}^+)^2 - (\mu_{i,j-1}^+)^2, \quad \forall j \in \mathcal{D} \setminus \{1\},$$

$$k_{i,1}^- = -(\mu_{i,1}^-)^2, \quad k_{i,j}^- = -(\mu_{i,j}^-)^2 + (\mu_{i,j-1}^-)^2,$$

$$b_{i,1}^+ = b_{i,1}^- = 0, \quad b_{i,j}^+ = \sum_{l=1}^{j-1} (\chi_{i,l}^+)^2, \quad b_{i,j}^- = -\sum_{l=1}^{j-1} (\chi_{i,l}^-)^2.$$

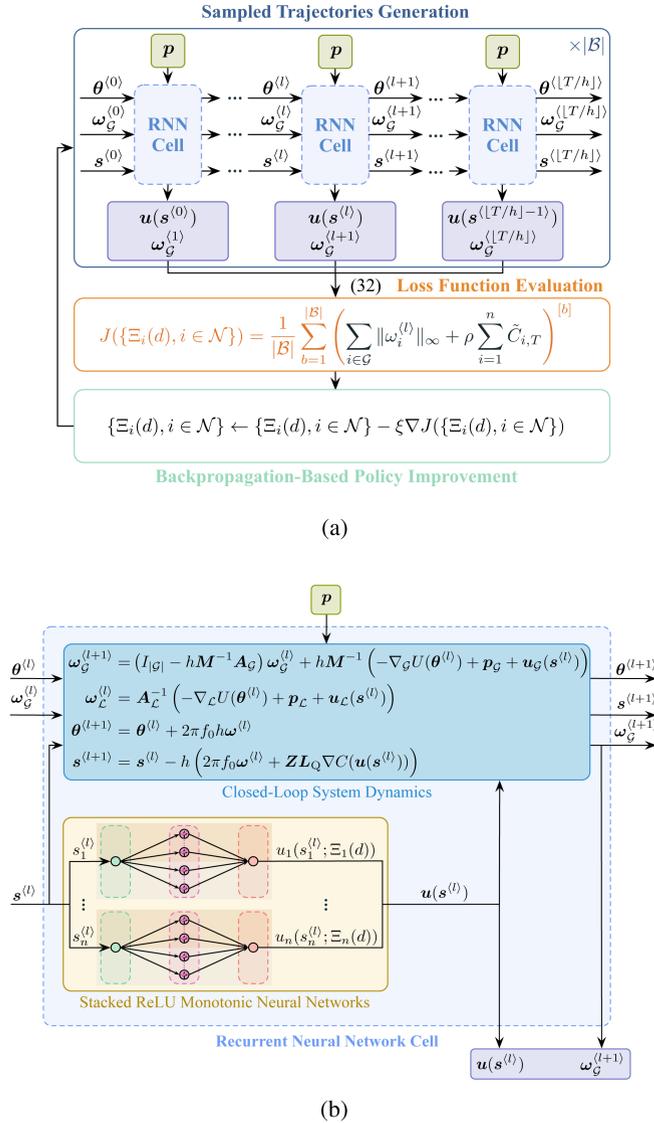
In this way, the requirements in (30) naturally hold. Thus, the remaining problem is just the search of optimal parameters  $\mu_i^+$ ,  $\chi_i^+$ ,  $\mu_i^-$ , and  $\chi_i^-$  through the training of neural networks. To highlight the dependence of the control policy  $u_i(s_i)$  generated by the  $i$ th neural network on parameters  $\mu_i^+$ ,  $\chi_i^+$ ,  $\mu_i^-$ , and  $\chi_i^-$ , we will use the notation  $u_i(s_i; \Xi_i(d))$  in the rest of this paper, where  $\Xi_i(d)$  denotes a set of parameters  $\Xi_i := \{\mu_i^+, \chi_i^+, \mu_i^-, \chi_i^-\}$  associated with a stacked ReLU monotonic neural network that has  $2d$  hidden neurons.

## B. RECURRENT NEURAL NETWORK-BASED REINFORCEMENT LEARNING

A recurrent neural network (RNN) is a class of neural networks that have a looping mechanism to allow information learned from the previous step to flow to the next step. This feature makes a RNN suitable for learning parameters related to dynamic systems, where the prior states usually affect the following states. Thus, to efficiently train the stacked ReLU monotonic neural networks for parameters  $\Xi_i(d)$ ,  $\forall i \in \mathcal{N}$ , we adopt the RNN-based RL algorithm proposed by [24], whose scheme is shown in Fig. 3(a).

To fit into the RNN architecture, we have to discretize the problem (9). Let the  $l$ th sampling instant after the initial time

<sup>6</sup>Strictly speaking, we require  $\mu_{i,j}^+ \neq 0$  and  $\mu_{i,j}^- \neq 0$ ,  $\forall j \in \mathcal{D}$ , which, however, is true almost for sure in the training process. Thus, we can simply ignore this constraint in practice.



**FIGURE 3.** Illustration of the RNN-based RL algorithm. (a) RL scheme for training stacked ReLU monotonic neural networks. (b) Structure of the recurrent neural network cell.

$t_0$  be  $(t_0 + lh)$ , where  $h$  is the sampling interval. Then, using voltage angles  $\theta := (\theta_i, i \in \mathcal{N}) \in \mathbb{R}^n$  as an example, we would like to approximate the continuous-time state at the  $l$ th sampling instant  $\theta(t_0 + lh)$  as a discrete-time state  $\theta^{(l)}$ . Provided that  $\theta^{(0)} = \theta(t_0)$ , the simplest way to obtain  $\theta^{(l)} \approx \theta(t_0 + lh)$  is through the recurrence relation via Euler method, i.e.,  $\theta^{(l+1)} = \theta^{(l)} + 2\pi f_0 h \omega^{(l)}$ . In general, Euler method works well for small enough sampling interval  $h$ , which means that the discretized system dynamics are sufficiently close to the original system dynamics. This allows us to solve the discretized problem below instead:

$$\begin{aligned} \min_{\Xi_i(d), i \in \mathcal{N}} \quad & \sum_{i \in \mathcal{G}} \|\omega_i^{(l)}\|_\infty + \rho \sum_{i=1}^n \tilde{C}_{i,T} \\ \text{s.t.} \quad & \omega_G^{(l+1)} = (I_{|\mathcal{G}|} - hM^{-1}A_G)\omega_G^{(l)} \end{aligned}$$

$$\begin{aligned} & + hM^{-1} \left( -\nabla_{\mathcal{G}}U(\theta^{(l)}) + p_G + u_G(s^{(l)}) \right), \\ \omega_L^{(l)} &= A_L^{-1} \left( -\nabla_L U(\theta^{(l)}) + p_L + u_L(s^{(l)}) \right), \\ \theta^{(l+1)} &= \theta^{(l)} + 2\pi f_0 h \omega^{(l)}, \\ s^{(l+1)} &= s^{(l)} - h \left( 2\pi f_0 \omega^{(l)} + ZL_Q \nabla C(u(s^{(l)})) \right), \\ u(s^{(l)}) &:= \left( u_i(s_i^{(l)}; \Xi_i(d)), i \in \mathcal{N} \right), \end{aligned} \quad (31)$$

with

$$\|\omega_i^{(l)}\|_\infty := \max_{l \in \{0, \dots, \lfloor T/h \rfloor - 1\}} |\omega_i^{(l+1)}|, \quad (32a)$$

$$\tilde{C}_{i,T} := \frac{1}{\lfloor T/h \rfloor} \sum_{l=0}^{\lfloor T/h \rfloor - 1} C_i(u_i(s_i^{(l)}; \Xi_i(d))), \quad (32b)$$

where  $\lfloor T/h \rfloor$  denotes the floor of  $T/h$ .

After a random initialization of  $\{\Xi_i(d), i \in \mathcal{N}\}$ , the RNN-based RL algorithm solves the problem (31) by gradually learning the optimal  $\{\Xi_i(d), i \in \mathcal{N}\}$  through an iterative process mainly composed of three parts — sampled trajectories generation, loss function evaluation, and backpropagation-based policy improvement — as illustrated in Fig. 3(a). For a given operating point  $(\theta^{(0)}, \omega_G^{(0)}, s^{(0)})$  of the power system, a sample of system trajectory over a time horizon  $T$  following sudden power disturbances can be generated by randomly initializing the constant input  $p$  to the RNN cell and then simulating forward  $\lfloor T/h \rfloor$  time steps. The detailed structure of the RNN cell is shown in Fig. 3(b), where the control signals  $u(s^{(l)})$  produced by  $n$  stacked ReLU monotonic neural networks with parameters  $\{\Xi_i(d), i \in \mathcal{N}\}$  are applied to the discretized closed-loop system with states  $(\theta^{(l)}, \omega_G^{(l)}, s^{(l)})$  for the generation of  $(\theta^{(l+1)}, \omega_G^{(l+1)}, s^{(l+1)})$ . In this way, the RNN cell evolves forward for  $l = 0, \dots, \lfloor T/h \rfloor - 1$ , which yields one trajectory under the parameters  $\{\Xi_i(d), i \in \mathcal{N}\}$  that are shared temporally. The loss function related to this trajectory is defined as the objective function of the problem (31), which can be computed by applying (32) to the stored historical outputs of the RNN cell. Usually, to improve the efficiency of training, multiple trajectories are generated parallelly, where each of them is excited by a randomly initialized constant input  $p$ . The collection of these trajectories is called a batch and the number of such trajectories is called the batch size. Then, the loss function related to a batch of size  $|\mathcal{B}|$  is defined as the mean of the losses related to individual trajectories in the batch, i.e.,

$$J(\{\Xi_i(d), i \in \mathcal{N}\}) := \frac{1}{|\mathcal{B}|} \sum_{b=1}^{|\mathcal{B}|} \left( \sum_{i \in \mathcal{G}} \|\omega_i^{(l)}\|_\infty + \rho \sum_{i=1}^n \tilde{C}_{i,T} \right)^{[b]},$$

where, with abuse of notation, we simply introduce a superscript  $[b]$  to denote the loss along the  $b$ th trajectory in the batch rather than accurately distinguish  $\|\omega_i^{(l)}\|_\infty$  and  $\tilde{C}_{i,T}$  along different trajectories to avoid complicating the notations too much. Once the loss of the batch  $J(\{\Xi_i(d), i \in \mathcal{N}\})$  has

been calculated, the parameters  $\{\Xi_i(d), i \in \mathcal{N}\}$  are adjusted according to the batch gradient descent method with a suitable learning rate  $\xi$ , i.e.,

$$\{\Xi_i(d), i \in \mathcal{N}\} \leftarrow \{\Xi_i(d), i \in \mathcal{N}\} - \xi \nabla J(\{\Xi_i(d), i \in \mathcal{N}\}),$$

where the gradient is approximated efficiently through backward gradient propagation. This ends one epoch of the iterative training process. The procedure is repeated until the number of epochs  $E$  is reached, which produces the optimal  $\{\Xi_i(d), i \in \mathcal{N}\}$  for the stacked ReLU monotonic neural networks so that the corresponding control policy  $u_i(s_i; \Xi_i(d))$  on each bus helps the generalized DAI to best handle the optimal transient frequency control problem (9).

## V. NUMERICAL ILLUSTRATIONS

In this section, we provide numerical validation for the performance of our RL-DAI on the 39-bus New England system [37]. First, we will train the optimal control policy  $\mathbf{u}(s) = (u_i(s_i; \Xi_i(d)), i \in \mathcal{N})$  for the RL-DAI on the ideal power system model as described in Section IV-B using TensorFlow framework. Then, we will validate the performance of the trained RL-DAI on a more realistic power system test case using Power System Toolbox (PST) [38] for Matlab.

The dynamic model of the 39-bus New England system contains 10 generator buses and 29 load buses, whose union is denoted as  $\mathcal{N}$ . Each of the 10 generator buses is distinctly indexed by some  $i \in \{30, \dots, 39\} := \mathcal{G}$  and each of the 29 load buses is distinctly indexed by some  $i \in \{1, \dots, 29\} := \mathcal{L}$ . The generator inertia constant  $m_i$ , the voltage magnitude  $v_i$ , and the susceptance  $B_{ij}$  are directly obtained from the dataset. Given that the values of frequency sensitivity coefficients are not provided by the dataset, we set  $\alpha_i = 150$  p.u.,  $\forall i \in \mathcal{G}$ , and  $\alpha_i = 100$  p.u.,  $\forall i \in \mathcal{L}$ .<sup>7</sup>

We then add frequency control  $u_i$  governed by our RL-DAI to each bus, where the underlying communication graph  $(\mathcal{V}, \mathcal{E}_Q)$  has 38 edges forming a path that connects all buses with  $Q_{ij} = 125$ ,  $\forall \{i, j\} \in \mathcal{E}_Q$ . The operational cost functions are assumed to be  $C_i(u_i) = c_i u_i^4 / 4 + b_i$ ,  $\forall i \in \mathcal{N}$ , where the cost coefficients  $c_i$  and  $b_i$  are generated randomly from (0,1) and (0,0.001), respectively. Thus,  $\forall i \in \mathcal{N}$ ,  $\nabla C_i(u_i) = c_i u_i^3$  by (14) and  $\zeta_i = c_i^{1/3}$  by (15), which determines the evolution of the integral variable  $s_i$  in RL-DAI according to (16). In the remaining of this section, we will first train the control policy  $\mathbf{u}(s) = (u_i(s_i; \Xi_i(d)), i \in \mathcal{N})$  on the ideal power system model and then apply the optimal control policy obtained from the training to the more realistic power system test case.

### A. TRAINING OF STACKED RELU MONOTONIC NEURAL NETWORKS

Following the RNN-based RL algorithm illustrated in Section IV-B, we can train the stacked ReLU monotonic neural networks that construct control policy  $\mathbf{u}(s) =$

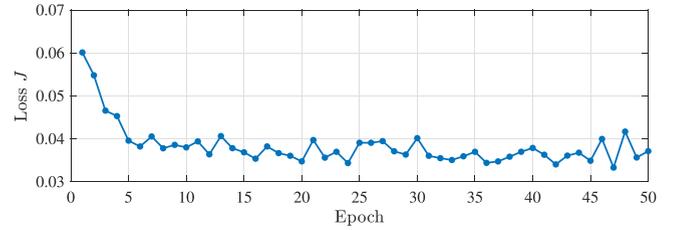


FIGURE 4. Loss over epochs for training RL-DAI.

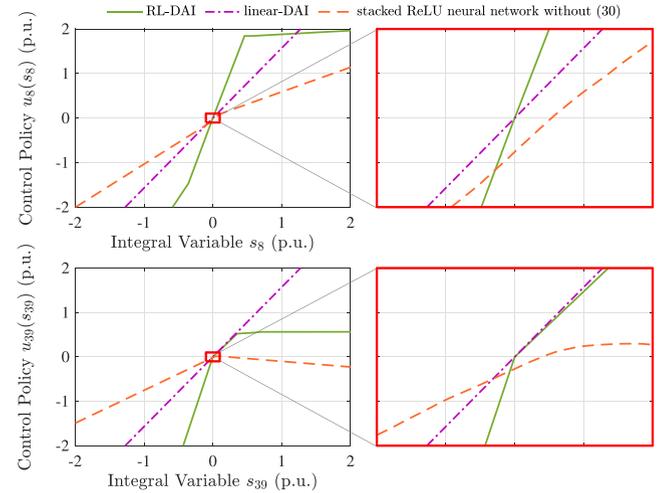


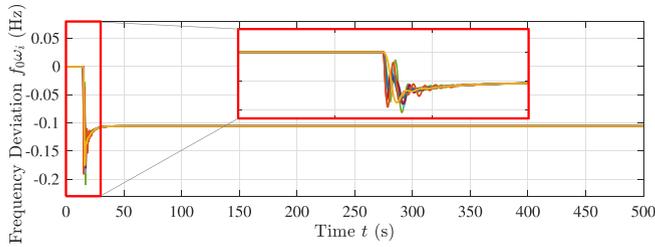
FIGURE 5. Visualization of control policy at the load bus 8 and the generator bus 39.

$(u_i(s_i; \Xi_i(d)), i \in \mathcal{N})$  satisfying Assumption 1 for RL-DAI. Specifically, each sample of system trajectories is generated by disturbing random step power changes  $\mathbf{p}$  to the ideal system model that is initially at a given supply-demand balanced setpoint, where each  $p_i$  is drawn uniformly from  $(-5, 5)$  p.u.. The values of all the hyperparameters mentioned in Section IV-B are summarized as follows:  $\rho = 0.01$ ,  $d = 20$ ,  $h = 0.5$  ms,  $T = 2.5$  s,  $|\mathcal{B}| = 64$ ,  $E = 50$ , and  $\xi = 0.5$  (exponentially decayed). With above hyperparameters, the training of the proposed RL-DAI takes around 7200 s in TensorFlow framework on MacBook Pro personal laptop with Apple M1 Pro chip. As shown in Fig. 4, the loss  $J$  is reduced by around 37.5% after the training. To provide a better insight into the learned control policy, Fig. 5 plots  $u_8(s_8)$  at the load bus 8 and  $u_{39}(s_{39})$  at the generator bus 39 as examples which clearly confirm that the learned  $u_i(s_i)$  satisfy Assumption 1 as expected.

### B. TIME-DOMAIN RESPONSES FOLLOWING POWER IMBALANCES

We now proceed to test the performance of our trained RL-DAI on the 39-bus New England system test case provided by PST. Although the simplified model is used in our theoretical analysis and training process, the test case adopts a more detailed power system model including line losses and voltage dynamics. As for the generators, instead of the

<sup>7</sup>All per unit values are on the system base, where the system power base is  $S_0 = 100$  MVA and the nominal system frequency is  $f_0 = 50$  Hz.



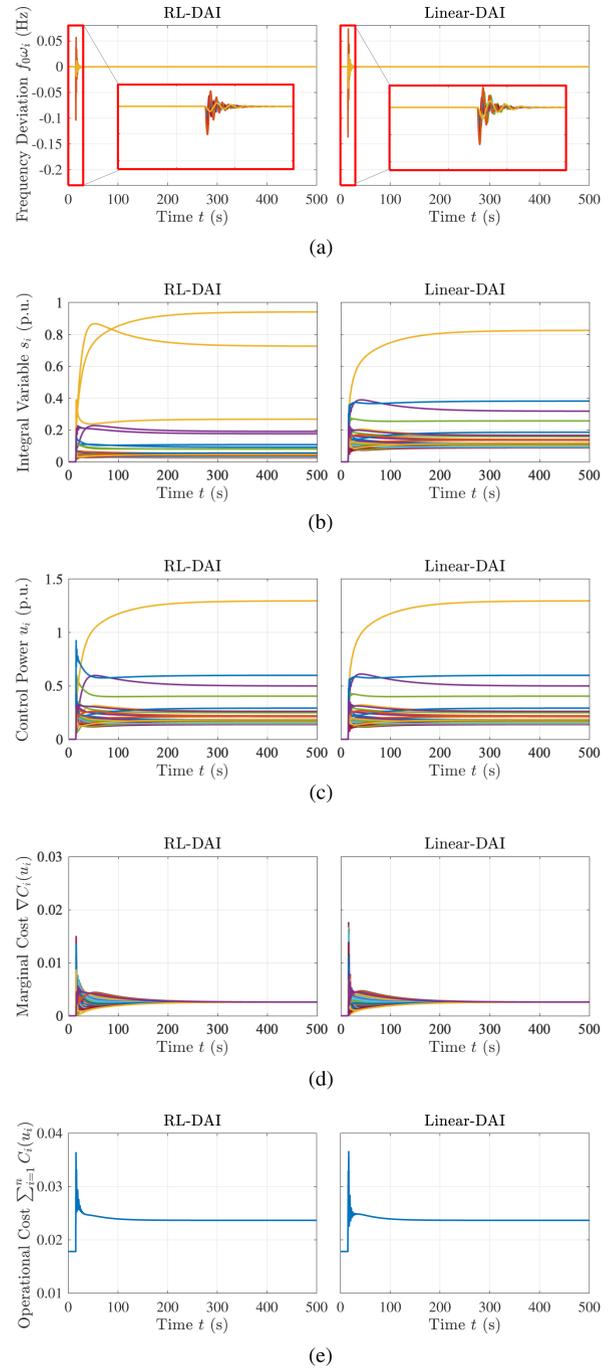
**FIGURE 6.** Original system performance without additional frequency control when a  $-3$  p.u. step power change occurs on buses 15, 23, and 29.

ideal swing dynamics, 2-axis sub-transient reactance models equipped with multi-stage turbines are adopted in our simulations, where the parameters of the turbines are chosen to be somewhat heterogenous to make the case more realistic. Moreover, instead of controllable power injections that follow our commands instantly and perfectly, the frequency control is implemented through inverter-interfaced resources on individual buses in our simulations, where we explicitly model dynamics of interfacing converters with the phase-locked loop and inner current control loop as in [39].

For the purpose of comparison, the frequency deviations of the original system without additional frequency control  $\mathbf{u}$  when there is a step change of 3 p.u. in power consumption at buses 15, 23, and 29 at time  $t = 15$  s are provided in Fig. 6. Clearly, in the absence of proper control, the transient frequency dip is large and there exists noticeable steady-state frequency deviation from the nominal frequency.

The performance of the system under RL-DAI in the same scenario is given in Fig. 7. Some observations are in order. First, a comparison between Figs. 6 and 7(a) shows that RL-DAI significantly improves the frequency Nadir and perfectly restores the frequencies to the nominal value as predicted by Theorem 1. Second, Fig. 7(d) confirms that RL-DAI help generators and loads asymptotically achieve identical marginal costs and thus the optimal steady-state economic dispatch problem is solved. Third, the nuances between the shapes of the trajectories of  $\mathbf{s}$  and  $\mathbf{u}$  in Fig. 7(b) and (c) are indicators of the nonlinearity of the learned control policy  $\mathbf{u}(\mathbf{s})$ . Clearly, this nonlinearity has not jeopardized the stability of the system.

To serve as the baseline, the performance of the system under linear-DAI in the same scenario is also provided in Fig. 7. Here, the linear-DAI denotes the DAI with the linear control policy  $u_i(s_i) = k_i s_i$  as in (11a) but modified dynamics of the integral variable  $s_i$  according to (16) in order to handle the biquadratic operational cost function. Trivially, linear-DAI is able to restore the nominal frequency, achieve the identical marginal costs, and ensure the closed-loop system stability by Theorems 1 and 2. Thus, we are more curious about its transient performance. To make a fair comparison, we gradually increase the linear control gain  $k_i$  uniformly for linear-DAI until the settling time under linear-DAI is the same as that under RL-DAI. Here, by settling time, we mean the time required for all states to reach and stay within 1% of their



**FIGURE 7.** Comparison of system performance under RL-DAI and linear-DAI when a  $-3$  p.u. step power change occurs on buses 15, 23, and 29. (a) Frequency deviations on generator buses. (b) Integral variables. (c) Controllable power injections. (d) Marginal costs. (e) Aggregate operational cost.

final values, which is around 343 s in our case. This procedure yields  $k_i = 1.57, \forall i \in \mathcal{N}$ . Thus, when generating the dynamic simulation results for linear-DAI in Fig. 7, we adopt the linear control policy  $u_i(s_i) = 1.57s_i, \forall i \in \mathcal{N}$ , whose shape is shown in Fig. 5. Observe from Fig. 7 that the trajectories of the aggregate economic cost are similar under RL-DAI and

**TABLE 1** Transient Performance of the System

Controller	Frequency Nadir $\sum_{i \in \mathcal{G}} \ \omega_i\ _\infty$	Economic cost $\sum_{i=1}^n \bar{C}_{i,T}$
RL-DAI	0.0093	0.0240
Linear-DAI	0.0150	0.0239

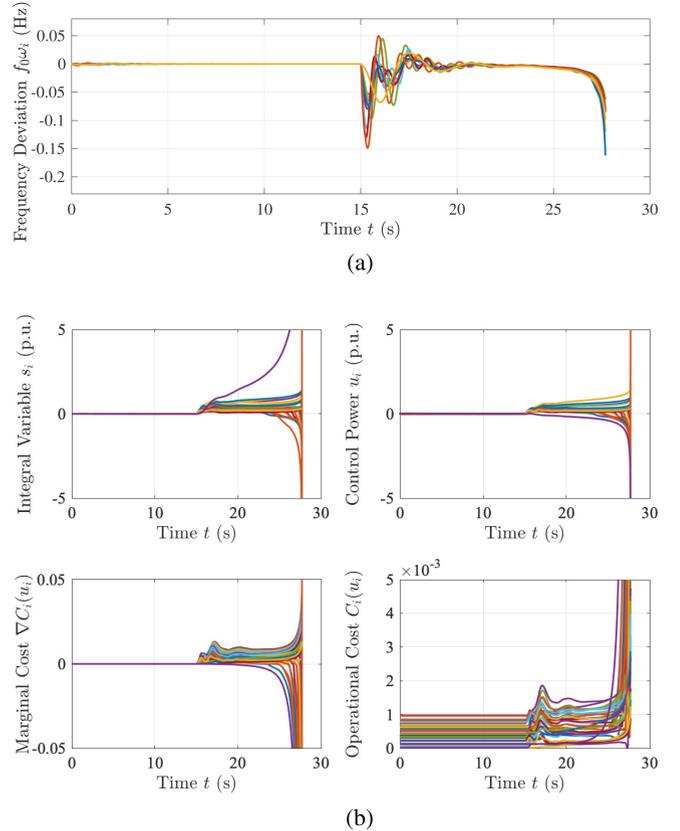
linear-DAI but RL-DAI largely improves the frequency Nadir than linear-DAI does. This observation can be confirmed by Table 1, which summarizes the transient performance metrics under two controllers. Thus, we can see, with the practically same economic cost, RL-DAI improves the frequency dynamics better than linear-DAI does, which shows the advantage of our proposed RL-DAI.

Finally, in order to better illustrate the importance of constraints (30) on weights and biases of stacked ReLU monotonic neural networks shown in Fig. 1 to stability of the closed-loop system under RL-DAI, we also test the case when constraints (30) are relaxed in the training of RL-DAI. Examples of the resulting control policy  $u_i(s_i)$  are plotted in Fig. 5, which clearly shows that each  $u_i(s_i)$  obtained neither necessarily passes through the origin nor necessarily increases with  $s_i$ . This implies that Assumption 1, one of the key assumptions for Theorems 1 and 2, may be violated if constraints (30) are not enforced. Thus, there is no stability guarantee in this case. This is in line with the dynamic simulation results shown in Fig. 8, where the system is obviously unstable since all states end up blowing up. Last but not least, it should be noted that there exist slight frequency oscillations immediately following  $t = 0$  s in Fig. 8(a), which is a weird action caused by the failure to satisfy  $u_i(0) = 0, \forall i \in \mathcal{N}$ . This straightforwardly explains the necessity of the requirement that each control policy  $u_i(s_i)$  must pass through the origin since otherwise the controllers will take actions even at the equilibrium point.

## VI. CONCLUSIONS AND OUTLOOK

This paper develops a framework that bridges the gap between the optimization of steady-state and transient frequency control performance for power systems undergoing power disturbances. Particularly, we propose to generalize DAI by adding properly designed nonlinearity to it. This provides us with the freedom to integrate RL techniques to DAI for finding the stable nonlinear controller that optimizes the transient frequency behavior while taking advantage of the DAI mechanism to optimize the steady-state cost of restoring frequency to its nominal value.

Several interesting research directions are open. First, we would like to expand the characterization of the properties of the learned controller beyond its stabilizing nature to address questions about robustness to unmodelled dynamics. In addition, although the proposed controllers only require local information when they are implemented, they are trained in a centralized manner. Understanding how they can be trained in a distributed way is also very relevant. Finally, in this paper, we have extended the class of cost functions from quadratic



**FIGURE 8.** System performance when control policy  $u(s)$  of RL-DAI are constructed from stacked ReLU neural networks without constraints (30) when  $-3$  p.u. step power changes occur at buses 15, 23, and 29. (a) Frequency deviations on generator buses. (b) Control and cost variables.

ones to those that satisfy a certain scaling property. It would be interesting to see if DAI control can be applied to an even larger class of cost functions.

## APPENDIX

### A. Proof of Lemma 1

$\forall y \in \mathbb{R}$ , let  $\tilde{u}_i = \nabla C_i^{-1}(y)$ . Clearly,  $y = \nabla C_i(\tilde{u}_i)$  by the definition of the inverse function. Also, under Assumption 2,  $\nabla C_i(\tilde{u}_i) = \nabla C_o(\zeta_i \tilde{u}_i)$ ,  $\forall i \in \mathcal{N}$ . Thus,  $y = \nabla C_i(\tilde{u}_i) = \nabla C_o(\zeta_i \tilde{u}_i)$ , which implies that  $\zeta_i \tilde{u}_i = \nabla C_o^{-1}(y)$ . Therefore, we have  $\zeta_i \nabla C_i^{-1}(y) = \zeta_i \tilde{u}_i = \nabla C_o^{-1}(y)$ ,  $\forall i \in \mathcal{N}$ , which concludes the proof.

### B. Proof of Lemma 2

By a similar argument as in the proof of [11, Lemma 4], we can easily get the first inequality and

$$\|\omega_{\mathcal{L}}\|_2 \leq \tilde{\nu} (\|\delta - \delta^*\|_2 + \|u(s) - u(s^*)\|_2) \quad (33)$$

for some constant  $\tilde{\nu} > 0$ . Under Assumption 1, each  $u_i(s_i)$  is Lipschitz continuous, which indicates that there exists some constant  $\eta > 0$  such that,  $\forall i \in \mathcal{N}$ ,

$$|u_i(s_i) - u_i(s_i^*)| \leq \eta |s_i - s_i^*|, \quad \forall s_i \in \mathbb{R}. \quad (34)$$

Combining (33) and (34) yields

$$\begin{aligned} \|\omega_{\mathcal{L}}\|_2 &\leq \tilde{v} (\|\delta - \delta^*\|_2 + \eta\|s - s^*\|_2) \\ &\leq \tilde{v} (\eta + 1) (\|\delta - \delta^*\|_2^2 + \|\omega_{\mathcal{G}}\|_2^2 + \|s - s^*\|_2^2)^{\frac{1}{2}}. \end{aligned} \quad (35)$$

Using (35), we can get

$$\begin{aligned} &\|\delta - \delta^*\|_2^2 + \|\omega\|_2^2 + \|s - s^*\|_2^2 \\ &= \|\delta - \delta^*\|_2^2 + \|\omega_{\mathcal{G}}\|_2^2 + \|s - s^*\|_2^2 + \|\omega_{\mathcal{L}}\|_2^2 \\ &\leq [1 + \tilde{v}^2 (\eta + 1)^2] (\|\delta - \delta^*\|_2^2 + \|\omega_{\mathcal{G}}\|_2^2 + \|s - s^*\|_2^2), \end{aligned}$$

which concludes the proof with  $\nu := 1 + \tilde{v}^2 (\eta + 1)^2$ .

### C. Proof of Lemma 3

Let  $\mathbf{x} := (x_i, i \in \mathcal{N}) \in \mathbb{R}^n$  and  $\mathbf{y} := (y_i, i \in \mathcal{N}) \in \mathbb{R}^n$ . We note that,  $\forall \mathbf{x} \neq \mathbf{y}$ ,

$$\begin{aligned} (\nabla L(\mathbf{x}) - \nabla L(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) &= (\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) \\ &= \sum_{i=1}^n (u_i(x_i) - u_i(y_i)) (x_i - y_i) \\ &> 0, \end{aligned} \quad (36)$$

where the inequality results from the monotonicity of each  $u_i(s_i)$  under Assumption 1. Thus, by [31, Theorem 2.14], (36) indicates the strict convexity of  $L(s)$ .

### D. Proof of Lemma 4

We can check that  $W(\delta, \omega_{\mathcal{G}}, s)$  only vanishes at the equilibrium term by term. Clearly, the kinetic energy term  $\pi f_0 \omega_{\mathcal{G}}^T \mathbf{M} \omega_{\mathcal{G}} > 0, \forall \omega_{\mathcal{G}} \neq \mathbf{0}_{|\mathcal{G}|}$ . By [11, Lemma 4], the potential energy term  $U(\delta) - U(\delta^*) - \nabla U(\delta^*)^T (\delta - \delta^*) \geq \beta \|\delta - \delta^*\|_2^2 > 0, \forall \delta \neq \delta^*$  satisfying  $|\delta_i - \delta_j| \in [0, \pi/2), \forall \{i, j\} \in \mathcal{E}$ , for some constant  $\beta > 0$ . By [31, Theorem 2.14], the strict convexity of  $L(s)$  shown in Lemma 3 ensures that  $L(s) - L(s^*) - \nabla L(s^*)^T (s - s^*) > 0, \forall s \neq s^*$ . It follows directly that  $W(\delta^*, \mathbf{0}_{|\mathcal{G}|}, s^*) = 0$  and  $W(\delta, \omega_{\mathcal{G}}, s) > 0, \forall (\delta, \omega_{\mathcal{G}}, s) \in \mathcal{X} \setminus (\delta^*, \mathbf{0}_{|\mathcal{G}|}, s^*)$ .

### E. Proof of Lemma 5

We calculate the derivative of  $W(\delta, \omega_{\mathcal{G}}, s)$  along the trajectories of the closed-loop system (17) as

$$\begin{aligned} &\dot{W}(\delta, \omega_{\mathcal{G}}, s) \\ &= \begin{bmatrix} 2\pi f_0 \mathbf{M} \omega_{\mathcal{G}} \\ \nabla U(\delta) - \nabla U(\delta^*) \\ \nabla L(s) - \nabla L(s^*) \end{bmatrix}^T \begin{bmatrix} \dot{\omega}_{\mathcal{G}} \\ \dot{\delta} \\ \dot{s} \end{bmatrix} \\ &\stackrel{\textcircled{1}}{=} 2\pi f_0 \omega_{\mathcal{G}}^T (-\mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}} - \nabla_{\mathcal{G}} U(\delta) + \mathbf{p}_{\mathcal{G}} + \mathbf{u}_{\mathcal{G}}(s)) \end{aligned}$$

$$\begin{aligned} &+ 2\pi f_0 (\nabla U(\delta) - \nabla U(\delta^*))^T \left( \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \omega \\ &+ (\mathbf{u}(s) - \mathbf{u}(s^*))^T (-2\pi f_0 \omega - \mathbf{Z} \mathbf{L}_{\mathcal{Q}} \nabla C(\mathbf{u}(s))) \\ &- 2\pi f_0 \omega_{\mathcal{G}}^T (-\nabla_{\mathcal{G}} U(\delta^*) + \mathbf{p}_{\mathcal{G}} + \mathbf{u}_{\mathcal{G}}(s^*)) \\ &\stackrel{\textcircled{2}}{=} -2\pi f_0 \omega_{\mathcal{G}}^T \mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}} - (\mathbf{u}(s) - \mathbf{u}(s^*))^T \mathbf{Z} \mathbf{L}_{\mathcal{Q}} \nabla C(\mathbf{u}(s)) \\ &- 2\pi f_0 (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + \mathbf{u}_{\mathcal{L}}(s) - \mathbf{u}_{\mathcal{L}}(s^*))^T \omega_{\mathcal{L}} \\ &\stackrel{\textcircled{3}}{=} -2\pi f_0 \omega_{\mathcal{G}}^T \mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}} - \mathbf{u}(s)^T \mathbf{Z} \mathbf{L}_{\mathcal{Q}} \nabla C(\mathbf{u}(s)) \\ &- 2\pi f_0 (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + \mathbf{u}_{\mathcal{L}}(s) - \mathbf{u}_{\mathcal{L}}(s^*))^T \mathbf{A}_{\mathcal{L}}^{-1} \\ &\quad \cdot (-\nabla_{\mathcal{L}} U(\delta) + \mathbf{p}_{\mathcal{L}} + \mathbf{u}_{\mathcal{L}}(s)) \\ &\stackrel{\textcircled{4}}{=} -2\pi f_0 \omega_{\mathcal{G}}^T \mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}} - \mathbf{u}(s)^T \mathbf{Z} \mathbf{L}_{\mathcal{Q}} \nabla C(\mathbf{u}(s)) \\ &- 2\pi f_0 (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + \mathbf{u}_{\mathcal{L}}(s) - \mathbf{u}_{\mathcal{L}}(s^*))^T \mathbf{A}_{\mathcal{L}}^{-1} \\ &\quad \cdot [-\nabla_{\mathcal{L}} U(\delta) + \mathbf{p}_{\mathcal{L}} + \mathbf{u}_{\mathcal{L}}(s) - (-\nabla_{\mathcal{L}} U(\delta^*) + \mathbf{p}_{\mathcal{L}} + \mathbf{u}_{\mathcal{L}}(s^*))] \\ &= -2\pi f_0 \omega_{\mathcal{G}}^T \mathbf{A}_{\mathcal{G}} \omega_{\mathcal{G}} - \mathbf{u}(s)^T \mathbf{Z} \mathbf{L}_{\mathcal{Q}} \nabla C(\mathbf{u}(s)) \\ &- 2\pi f_0 (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + \mathbf{u}_{\mathcal{L}}(s) - \mathbf{u}_{\mathcal{L}}(s^*))^T \mathbf{A}_{\mathcal{L}}^{-1} \\ &\quad \cdot (\nabla_{\mathcal{L}} U(\delta^*) - \nabla_{\mathcal{L}} U(\delta) + \mathbf{u}_{\mathcal{L}}(s) - \mathbf{u}_{\mathcal{L}}(s^*)), \end{aligned}$$

which is exactly (29). Here, some tricks are used to add or remove terms for constructing a quadratic format without affecting the original value of  $\dot{W}(\delta, \omega_{\mathcal{G}}, s)$ . In  $\textcircled{1}$ , the property that  $-\nabla_{\mathcal{G}} U(\delta^*) + \mathbf{p}_{\mathcal{G}} + \mathbf{u}_{\mathcal{G}}(s^*) = \mathbf{0}_{|\mathcal{G}|}$  by (19b) of Theorem 1 is used. In  $\textcircled{2}$ , the fact that  $\nabla U(\delta) \mathbf{1}_n = \mathbf{0}$  is used. In  $\textcircled{3}$ , the fact that  $\omega_{\mathcal{L}}$  is determined by (26) and the property that  $\mathbf{u}(s^*)^T \mathbf{Z} \mathbf{L}_{\mathcal{Q}} = \nabla C_{\mathcal{O}}^{-1}(\gamma) \mathbf{1}_n^T \mathbf{Z}^{-1} \mathbf{Z} \mathbf{L}_{\mathcal{Q}} = \nabla C_{\mathcal{O}}^{-1}(\gamma) \mathbf{1}_n^T \mathbf{L}_{\mathcal{Q}} = \mathbf{0}_n^T$  obtained through (19c) of Theorem 1 are used. In  $\textcircled{4}$ , the property that  $-\nabla_{\mathcal{L}} U(\delta^*) + \mathbf{p}_{\mathcal{L}} + \mathbf{u}_{\mathcal{L}}(s^*) = \mathbf{0}_{|\mathcal{L}|}$  by (19b) of Theorem 1 is used.

### F. Proof of Lemma 6

This is a direct extension of the standard Laplacian potential function [33]:

$$\begin{aligned} &\mathbf{x}^T \mathbf{Z} \mathbf{L}_{\mathcal{Q}} \mathbf{y} \\ &= \sum_{i=1}^n \zeta_i x_i \left( \sum_{j=1}^n L_{\mathcal{Q},ij} y_j \right) = \sum_{i=1}^n \zeta_i x_i \left( L_{\mathcal{Q},ii} y_i + \sum_{j=1, j \neq i}^n L_{\mathcal{Q},ij} y_j \right) \\ &= \sum_{i=1}^n \zeta_i x_i \left( \sum_{j=1, j \neq i}^n Q_{ij} y_i - \sum_{j=1, j \neq i}^n Q_{ij} y_j \right) \\ &= \sum_{i=1}^n \sum_{j=1, j \neq i}^n Q_{ij} \zeta_i x_i (y_i - y_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ij}}{2} \zeta_i x_i (y_i - y_j) + \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ij}}{2} \zeta_j x_j (y_i - y_j) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ij}}{2} \zeta_i x_i (y_i - y_j) + \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ji}}{2} \zeta_j x_j (y_j - y_i) \\
 &\stackrel{\textcircled{1}}{=} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ij}}{2} \zeta_i x_i (y_i - y_j) + \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ij}}{2} \zeta_j x_j (y_j - y_i) \\
 &= \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{ij}}{2} (y_i - y_j) (\zeta_i x_i - \zeta_j x_j) \\
 &= \sum_{(i,j) \in \mathcal{E}_Q} Q_{ij} (y_i - y_j) (\zeta_i x_i - \zeta_j x_j),
 \end{aligned}$$

where  $\textcircled{1}$  uses the symmetry of the Laplacian matrix  $\mathbf{L}_Q$ .

### G. Proof of Corollary 1

It follows directly from Lemma 6 by setting  $\mathbf{x} = \mathbf{u}(\mathbf{s})$  and  $\mathbf{y} = \nabla C(\mathbf{u}(\mathbf{s}))$  that

$$\begin{aligned}
 &\mathbf{u}(\mathbf{s})^T \mathbf{Z} \mathbf{L}_Q \nabla C(\mathbf{u}(\mathbf{s})) \\
 &= \sum_{(i,j) \in \mathcal{E}_Q} Q_{ij} [\nabla C_i(u_i(s_i)) - \nabla C_j(u_j(s_j))] [\zeta_i u_i(s_i) - \zeta_j u_j(s_j)].
 \end{aligned} \tag{37}$$

We claim that,  $\forall i, j \in \mathcal{N}$ , there is parity between the signs of  $(\nabla C_i(u_i(s_i)) - \nabla C_j(u_j(s_j)))$  and  $(\zeta_i u_i(s_i) - \zeta_j u_j(s_j))$ , i.e.,

$$\text{sign}(\nabla C_i(u_i) - \nabla C_j(u_j)) = \text{sign}(\zeta_i u_i - \zeta_j u_j), \tag{38}$$

where  $\text{sign}$  denotes the sign function. To see this, without loss of generality,  $\forall i, j \in \mathcal{N}$ , we only consider the case where  $\nabla C_i(u_i) - \nabla C_j(u_j) > 0$ . Since  $\nabla C_i(u_i) > \nabla C_j(u_j)$ , we have

$$\begin{aligned}
 &\zeta_i u_i = \zeta_i \nabla C_i^{-1}(\nabla C_i(u_i)) > \zeta_i \nabla C_i^{-1}(\nabla C_j(u_j)) \\
 &\stackrel{\textcircled{1}}{=} \nabla C_0^{-1}(\nabla C_j(u_j)) \stackrel{\textcircled{2}}{=} \zeta_j \nabla C_j^{-1}(\nabla C_j(u_j)) = \zeta_j u_j.
 \end{aligned}$$

Here, the inequality results from the fact that  $\zeta_i > 0$  and  $\nabla C_i^{-1}(\cdot)$  is strictly increasing as mentioned in Remark 3. By Lemma 1,  $\textcircled{1}$  and  $\textcircled{2}$  hold. The case where  $\nabla C_i(u_i) - \nabla C_j(u_j) \leq 0$  follows from an analogous argument. Therefore, our claim that (38) holds is true, which together with the fact that  $Q_{ij} > 0, \forall \{i, j\} \in \mathcal{E}_Q$ , implies that (37) is nonnegative. Notably, since each term of the sum in (37) is nonnegative, as the sum vanishes, each term must be 0. Thus,  $\mathbf{u}(\mathbf{s})^T \mathbf{Z} \mathbf{L}_Q \nabla C(\mathbf{u}(\mathbf{s})) = 0$  if and only if,  $\forall \{i, j\} \in \mathcal{E}_Q$ ,

$$[\nabla C_i(u_i(s_i)) - \nabla C_j(u_j(s_j))] [\zeta_i u_i(s_i) - \zeta_j u_j(s_j)] = 0,$$

which is equivalent to  $\nabla C_i(u_i(s_i)) = \nabla C_j(u_j(s_j)), \forall \{i, j\} \in \mathcal{E}_Q$  by (38). Recall that the communication graph  $(\mathcal{V}, \mathcal{E}_Q)$  is assumed to be connected, which implies that  $\nabla C_1(u_1(s_1)) = \dots = \nabla C_n(u_n(s_n))$ , i.e.,  $\nabla C(\mathbf{u}(\mathbf{s})) \in \text{range}(\mathbb{1}_n)$ .

## REFERENCES

- [1] D. S. Kirschen and G. Strbac, *Fundamentals of Power System Economics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2019.
- [2] J. H. Eto et al., "Use of frequency response metrics to assess the planning and operating requirements for reliable integration of variable renewable generation," Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep. LBNL-4142E, 2010.
- [3] B. Kroposki et al., "Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy," *IEEE Power Energy Mag.*, vol. 15, no. 2, pp. 61–73, Mar./Apr. 2017.
- [4] S. Nalley and A. LaRose, "Annual energy outlook 2021 with projections to 2050," U.S. Energy Inf. Admin., Tech. Rep., 2021.
- [5] L. Bird, M. Milligan, and D. Lew, "Integrating variable renewable energy: Challenges and solutions," National Renewable Energy Laboratory, Golden, CO, USA, Tech. Rep. NREL/TP-6A20-60451, 2013.
- [6] C. Zhao, E. Mallada, and F. Dörfler, "Distributed frequency control for stability and economic dispatch in power networks," in *Proc. Amer. Control Conf.*, 2015, pp. 2359–2364.
- [7] N. Li, C. Zhao, and L. Chen, "Connecting automatic generation control and economic dispatch from an optimization view," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 254–264, Sep. 2016.
- [8] E. Mallada, C. Zhao, and S. Low, "Optimal load-side control for frequency regulation in smart grids," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6294–6309, Dec. 2017.
- [9] F. Dörfler and S. Grammatico, "Gather-and-broadcast frequency control in power systems," *Automatica*, vol. 79, pp. 296–305, May 2017.
- [10] J. Schiffer, F. Dörfler, and E. Fridman, "Robustness of distributed averaging control in power systems: Time delays & dynamic communication topology," *Automatica*, vol. 80, pp. 261–271, Jun. 2017.
- [11] E. Weitenberg, C. D. Persis, and N. Monshizadeh, "Exponential convergence under distributed averaging integral frequency control," *Automatica*, vol. 98, pp. 103–113, Dec. 2018.
- [12] Z. Wang, F. Liu, S. H. Low, C. Zhao, and S. Mei, "Distributed frequency control with operational constraints, Part II: Network power balance," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 53–64, Jan. 2019.
- [13] P. You, Y. Jiang, E. Yeung, D. F. Gayme, and E. Mallada, "On the stability, economic efficiency and incentive compatibility of electricity market dynamics," 2021, *arXiv:2112.05811*.
- [14] A. Cherukuri, T. Stegink, C. D. Persis, A. J. van der Schaft, and J. Cortés, "Frequency-driven market mechanisms for optimal dispatch in power networks," *Automatica*, vol. 133, Nov. 2021, Art. no. 109861.
- [15] Z. Miao and L. Fan, "Achieving economic operation and secondary frequency regulation simultaneously through local feedback control," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 85–93, Jan. 2017.
- [16] J. Schiffer and F. Dörfler, "On stability of a distributed averaging PI frequency and active power controlled differential-algebraic power system model," in *Proc. Eur. Control Conf.*, 2016, pp. 1487–1492.
- [17] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: The MIT Press, 2018.
- [18] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: Reinforcement learning framework," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 427–435, Feb. 2004.
- [19] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Reinforcement learning for selective key applications in power systems: Recent advances and future challenges," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2935–2958, Jul. 2022.
- [20] D. Ye, M. Zhang, and D. Sutanto, "A hybrid multiagent framework with Q-learning for power grid systems restoration," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2434–2441, Nov. 2011.
- [21] 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, Jun. 2017.
- [22] Z. Yan and Y. Xu, "A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4599–4608, Nov. 2020.
- [23] C. Chen, M. Cui, F. Li, S. Yin, and X. Wang, "Model-free emergency frequency control based on reinforcement learning," *IEEE Trans. Ind. Inform.*, vol. 17, no. 4, pp. 2336–2346, Apr. 2021.
- [24] W. Cui, Y. Jiang, and B. Zhang, "Reinforcement learning for optimal primary frequency control: A Lyapunov approach," *IEEE Trans. Power Syst.*, early access, May 2022, doi: [10.1109/TPWRS.2022.3176525](https://doi.org/10.1109/TPWRS.2022.3176525).
- [25] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1177–1189, May 2014.
- [26] E. Weitenberg, Y. Jiang, C. Zhao, E. Mallada, C. De Persis, and F. Dörfler, "Robust decentralized secondary frequency control in power systems: Merits and tradeoffs," *IEEE Trans. Autom. Control*, vol. 64, no. 10, pp. 3967–3982, Oct. 2019.
- [27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

- [28] A. Walker et al., “Model of operation-and-maintenance costs for photovoltaic systems,” Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep., NREL/TP-5C00-74840, 2020.
- [29] T. Stehly, P. Beiter, and P. Duffy, “2019 cost of wind energy review,” National Renewable Energy Laboratory, Golden, CO, USA, Tech. Rep., NREL/TP-5000-78471, 2020.
- [30] B. Xu, A. Oudalov, A. Ulbig, G. Andersson, and D. S. Kirschen, “Modeling of Lithium-ion battery degradation for cell life assessment,” *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1131–1140, Mar. 2018.
- [31] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. Berlin, Germany: Springer, 1998.
- [32] K. Binmore, *Mathematical Analysis: A Straightforward Approach*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1977.
- [33] F. Bullo, *Lectures on Network Systems*, 1st ed. Seattle, WA, USA: Kindle Direct Publishing, 2022.
- [34] E. Weitenberg, C. D. Persis, and N. Monshizadeh, “Exponential convergence under distributed averaging integral frequency control,” 2017, *arXiv:1608.03798*.
- [35] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.
- [36] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The expressive power of neural networks: A view from the width,” in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017.
- [37] T. Athay, R. Podmore, and S. Virmani, “A practical method for the direct analysis of transient stability,” *IEEE Trans. Power App. Syst.*, vol. PAS-98, no. 2, pp. 573–584, Mar. 1979.
- [38] J. H. Chow and K. W. Cheung, “A toolbox for power system dynamics and control engineering education and research,” *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1559–1564, Nov. 1992.
- [39] Y. Jiang, E. Cohn, P. Vorobev, and E. Mallada, “Storage-based frequency shaping control,” *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5006–5019, Nov. 2021.

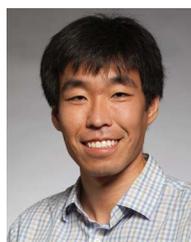


**YAN JIANG** (Member, IEEE) received the B.Eng. degree in electrical engineering and automation from the Harbin Institute of Technology, Harbin, China, in 2013, the M.S. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2016, and the Ph.D. degree in electrical engineering with the M.S.E. degree in applied mathematics and statistics from Johns Hopkins University, Baltimore, MD, USA, in 2021. She is currently a Postdoctoral Scholar with the Department of Electrical and

Computer Engineering, University of Washington, Seattle, WA, USA. Her research interests include control of power systems.



**WENQI CUI** (Graduate Student Member, IEEE) received the B.Eng. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, and Zhejiang University, Hangzhou, China, in 2016 and 2019, respectively. She is currently working toward the Ph.D. degree in electrical engineering with the University of Washington, Seattle, WA, USA. She works on control, optimization, and machine learning, with applications in cyberphysical systems.



**BAOSEN ZHANG** (Member, IEEE) received the Bachelor of Applied Science degree in engineering science from the University of Toronto, Toronto, ON, Canada, in 2008, and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, Berkeley, CA, USA, in 2013. He was a Postdoctoral Scholar with Stanford University, Stanford, CA. He is currently an Associate Professor in electrical and computer engineering with the University of Washington, Seattle, WA, USA. His research interests include

control, optimization, and learning applied to power systems and other cyberphysical systems. He was the recipient of the NSF CAREER Award and also several best paper awards.



**JORGE CORTÉS** (Fellow, IEEE) received the Licenciatura degree in mathematics from the Universidad de Zaragoza, Zaragoza, Spain, in 1997, and the Ph.D. degree in engineering mathematics from Universidad Carlos III de Madrid, Madrid, Spain, in 2001. He held Postdoctoral positions with the University of Twente, Twente, The Netherlands, and the University of Illinois at Urbana-Champaign, Urbana, IL, USA. He was an Assistant Professor with the Department of Applied Mathematics and Statistics, University of California,

Santa Cruz, CA, USA, from 2004 to 2007. He is currently a Professor with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA. His research interests include distributed control and optimization, network science, nonsmooth analysis, reasoning and decision making under uncertainty, network neuroscience, and multi-agent coordination in robotic, power, and transportation networks. He is a Fellow of SIAM.