Safe Reinforcement Learning for Grid-forming Inverter Based Frequency Regulation with Stability Guarantee

Hang Shuai, *Member*, *IEEE*, Buxin She, *Student Member*, *IEEE*, Jinning Wang, *Student Member*, *IEEE*, and Fangxing Li, *Fellow*, *IEEE*

***Abstract*—This study investigates a safe reinforcement learning algorithm for grid-forming (GFM) inverter based frequency regulation. To guarantee the stability of the inverter-based resource (IBR) system under the learned control policy, a model-based reinforcement learning (MBRL) algorithm is combined with Lyapunov approach, which determines the safe region of states and actions. To obtain near optimal control policy, the control performance is safely improved by approximate dynamic programming (ADP) using data sampled from the region of attraction (ROA). Moreover, to enhance the control robustness against parameter uncertainty in the inverter, a Gaussian process (GP) model is adopted by the proposed algorithm to effectively learn system dynamics from measurements. Numerical simulations validate the effectiveness of the proposed algorithm.**

***Index Terms*—Inverter-based resource (IBR), virtual synchronous generator (VSG), safe reinforcement learning, Lyapunov function, frequency regulation, grid-forming inverter.**

I. Introduction

POWER system frequency control is critical for maintaining grid stability when imbalance between generation and load occurs. As the penetration of inverter-based resources (IBRs) such as renewable energy and battery storage continues to increase, modern power systems are facing significant challenges due to reduced mechanical inertia and increased disturbances. Therefore, power system stability control has recently spurred much interest from both academia and industry [1], [2].

Various control methods have been proposed for IBRs to provide frequency regulation services [1], [3], [4]. For instance, both conventional synchronous generators (SGs) and IBR employ the frequency droop control strategy, which adjusts the active power output in response to frequency deviations. Droop-control-based inverters barely provide inertia support to the grid. Consequently, a droop-control-based network is typically characterized by a lack of inertia and being sensitive to faults [5]. In the event of a disturbance, the system frequency may undergo abrupt changes, potentially leading to the tripping of generators or the unnecessary shedding of loads. To alleviate the negative impact of low inertia, the virtual synchronous generator (VSG) [6], [7] control was developed. This control strategy emulates the frequency response characteristics of SGs, augmenting the system with virtual inertia and damping properties. Additionally, the values of inertia and damping in VSGs are more flexible than those in SGs, which are not limited by physical conditions such as rotating mass. Therefore, IBRs can adjust the inertia adaptively to obtain faster and more stable power output [8]-[10]. However, traditional frequency regulation strategies for IBRs were usually designed based on linearized small-signal models [8], [9], [11], which makes the control performance deteriorate quickly when frequency deviations are large. Due to the challenges posed by the low inertia and nonlinearity of IBRs, advanced controls are needed to ensure grid stability.

To deal with the challenges, various advanced frequency controllers are developed recently [12] - [14]. Among these methods, reinforcement learning (RL) technique is one of the most promising approaches. In [13], a model-free deep reinforcement learning (DRL) based load frequency control method was designed. The challenge of designing DRL-based power system stability controller lies in guaranteeing the control strategy won't lead to unstable condition after disturbances. However, the conventional model-free RL-based controllers mentioned above do not yield any stability guarantees. Therefore, [15] proposed a Lyapunov-based model-free RL strategy for primary frequency control of the power system, which can guarantee that the system frequency reaches stable equilibrium after disturbances. In [15] and [16], Lyapunov stability theory was utilized to design the architecture of recurrent neural network (RNN) controllers for power networks. However, the system parameters (e.g., inertia of SGs) need to be known in prior in order to train the neural Lyapunov function [16], and whether the learned function satisfies the Lyapunov conditions for all points in a re-

Manuscript received: November 13, 2023; revised: February 11, 2024; accepted: March 28, 2024. Date of CrossCheck: March 28, 2024. Date of online publication: April 9, 2024.

This work was funded in part by the CURENT Research Center and in part by the National Science Foundation (NSF) (No. ECCS-2033910).

H. Shuai, B. She, J. Wang, and F. Li (corresponding author) are with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, 37996, USA (e-mail: hshuai1@utk.edu; bshe@vols.utk.edu; jwang175@vols.utk.edu; fli6@utk.edu).

DOI: 10.35833/MPCE.2023.000882

gion needs further investigation. Given the frequent adjustments of virtual inertia and damping parameters in IBRs, the development of a robust DRL-based frequency regulation controller for IBRs could enhance their integration into the power system.

The primary contribution of this work is the development of a safe model-based reinforcement learning (MBRL) algorithm for grid-forming (GFM) inverter based frequency regulation. Inspired by [17], this algorithm addresses the challenges of ensuring controller stability and effectively dealing with system parameter uncertainty. In the proposed algorithm, a Gaussian process (GP) model is adopted to learn the unknown nonlinear dynamics of the inverter system, and approximate dynamic programming (ADP) [18], [19] is used to improve the control performance of the algorithm. Moreover, to guarantee the system stability under the learned control policy, Lyapunov function is used to obtain the region of attraction (ROA). Different from pre-training a neural Lyapunov function according to system dynamics in [16], we design the Lyapunov function as the value function of the Bellman's equation in ADP. This allows both the Lyapunov function and the control policy to update during training, leading to an enlarged ROA and improved control performance simultaneously. In addition, the controller based on the proposed algorithm is adaptive to the adjustment of inverter parameters (i.e., virtual inertia and damping coefficients), which means the controller will be more robust to parameter uncertainty.

This paper is organized as follows. Section II formulates the GFM inverter based frequency regulation problem. In Section III, the GFM inverter based frequency regulation via the safe MBRL controller is designed. The numerical simulations are presented in Section IV. Section V concludes the paper.

II. Formulation of GFM Inverter Based Frequency Regulation Problem

The diagram of GFM inverter based primary frequency control is depicted in Fig. 1.

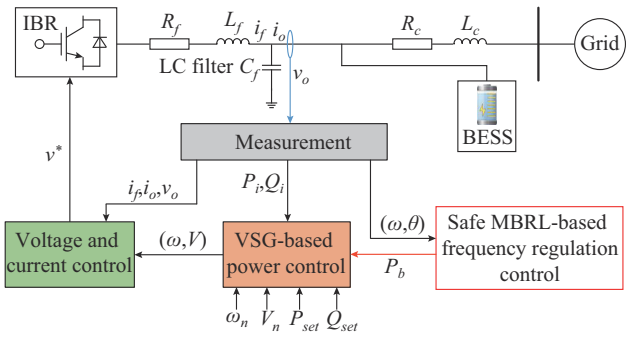

Fig. 1. Diagram of GFM inverter based primary frequency control.

We assume the bus voltage magnitudes to be 1 p.u., and neglect the reactive power flows. The frequency dynamics of VSG-based power control loop of the GFM inverter can be given by the swing equations [5], [16], [20]:

$$\begin{cases} \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega \\ M\dfrac{\mathrm{d}\omega}{\mathrm{d}t} = P_{set} - P_i - D\omega - u(\theta,\omega) \end{cases} \tag{1}$$

where $u(\cdot)$ is the control action function of the battery energy storage systems (BESSs), which denotes the active charging power (i.e., P_b in Fig. 1) of the BESS; M and D are the virtual inertia and damping constant of the inverter, respectively; P_{set} and P_i are the set point and real-time measurement of the active power output of the inverter, respectively; and θ and ω are the voltage phase and angular frequency deviation of the inverter, respectively. More specifically, $\omega = \omega_i - \omega_n$, where ω_i is the generated angular frequency of the inverter output voltage, and ω_n is the nominal angular frequency of the inverter. In Fig. 1, P_i can be calculated as [21]:

$$P_i = \sum_{j\in\{i,g\}} V_i V_j (B_{ij}\sin(\theta_i-\theta_j) + G_{ij}\cos(\theta_i-\theta_j)) \tag{2}$$

where B_{ij} and G_{ij} are the susceptance and conductance components of the (i,j) element of the admittance matrix $\boldsymbol{Y}$, respectively; V_i and θ_i are the voltage magnitude and phase of node i, respectively; and θ_g is the voltage phase of the main grid. Note that lossy power flow model is adopted in (2).

We aim to propose a control policy to improve the dynamic performance of VSG after disturbances with the minimal cost. The optimal control problem can be formulated as:

$$\begin{cases} \min\limits_{\boldsymbol{u}} \ (\boldsymbol{u}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{u} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}) \\ \text{s.t. (1)} \\ \quad \underline{\boldsymbol{u}} \le \boldsymbol{u} \le \bar{\boldsymbol{u}} \\ \quad \boldsymbol{u} \text{ is stabilizing} \end{cases} \tag{3}$$

where $\boldsymbol{x} = (\theta,\omega)$ is the state of the VSG; $\boldsymbol{Q}$ and $\boldsymbol{R}$ are the positive definite matrices; $\underline{\boldsymbol{u}}$ and $\bar{\boldsymbol{u}}$ are the lower and upper limitations of the control actions $\boldsymbol{u}$, respectively, which are determined by the maximum charging and discharging capacities of BESSs; and $\boldsymbol{u}$ is the vector of $u(\cdot)$. As shown in Fig. 1, the control action is optimized using the proposed algorithm.

III. GFM Inverter Based Frequency Regulation via Safe MBRL Controller

The primary objective of the controller is to safely learn about the frequency dynamics of VSG from measurements and adapt the control policy $\boldsymbol{\pi}$ for optimal performance, without encountering unstable system states. This implies that the adjustment of the control policy throughout the learning process must be performed in such a way that the system state remains within the ROA. The parameter uncertainty and nonlinearity of the AC power flow, as described in (2), make the design of controllers for (1) challenging. The proposed controller for GFM inverter based frequency regulation is depicted in Fig. 2. In the proposed controller, the frequency dynamics of VSG are learned by the GP model with system measurements. The ROA for a fixed policy is determined using Lyapunov functions. And the control policy is updated by ADP-based RL approach to expand the ROA. The details of the proposed policy are presented below.

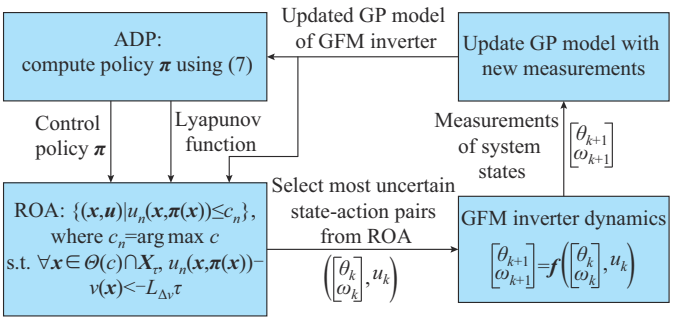

Fig. 2. Proposed algorithm for GFM inverter based frequency regulation.

By discretizing the dynamic model shown in (1) and (2), the dynamics can be reformulated as the following nonlinear discrete-time system:

$$\begin{cases}\theta_{k+1}=\theta_k+h\omega_k\\ \omega_{k+1}=\omega_k+\dfrac{h}{M}\left(P_{set,k}-P_{i,k}-D\omega_k-u_k\right)\end{cases} \tag{4a}$$

where h is the step size for the discrete simulation; and the subscript k denotes the discrete time index.

Equation (4a) can be expressed as:

$$\boldsymbol{x}_{k+1}=\boldsymbol{f}(\boldsymbol{x}_k,\boldsymbol{u}_k)=\boldsymbol{h}(\boldsymbol{x}_k,\boldsymbol{u}_k)+\boldsymbol{g}(\boldsymbol{x}_k,\boldsymbol{u}_k) \tag{4b}$$

where $\boldsymbol{f}(\cdot)$ denotes the true dynamics of the VSG, comprising two components: a known model represented by $\boldsymbol{h}(\cdot)$, and a priori unknown model errors denoted by $\boldsymbol{g}(\cdot)$. In inverters, the parameter (e.g., M and D in (1)) can undergo dynamic changes, which introduces uncertainties. To ensure the stability and predictability of the system, we assume the dynamic of the VSG is L_f-Lipschitz continuous, which means that the dynamic does not change too rapidly between any two points in its domain. This assumption holds true for the VSG system as described in (1), with a supporting proof provided in Appendix A.

To enable safe learning, we adopt GP model to learn a reliable statistical system model described by (1) and (2). GP model is a powerful method in machine learning and statistical modeling. GP consists of random variables, and any finite group of them follows a joint Gaussian distribution. In system modeling, GP is often used to capture complex relationships in data [22]. According to the GP theory [23], there exists a parameter $\beta_n>0$ such that with probability at least $1-\delta$ it holds for all $n\geq 0$ that $\|\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})-\boldsymbol{\mu}_n(\boldsymbol{x},\boldsymbol{u})\|_1\leq \beta_n\boldsymbol{\sigma}_n(\boldsymbol{x},\boldsymbol{u})$. $\boldsymbol{\mu}_n(\cdot)$ and $\boldsymbol{\sigma}_n(\cdot)=trace\left(\sqrt{\sum_n(\cdot)}\right)$ are the posterior mean and covariance matrix functions of the GP model of the VSG dynamics in (4b) conditioned on n measurements, respectively. In this way, we can use a GP model to build confidence intervals on the inverter dynamics, which can cover the true dynamics with probability $1-\delta$.

After learning about the inverter dynamics from measurements, the goal is to safely adapt the optimal control policy without leading to unstable system conditions. The safety of the controller is characterized by the safe region of states and actions, commonly referred to as the ROA [24]. When the system state falls within the boundaries of the ROA, the dynamics outlined in (1) will remain stable. Conversely, if the state ventures outside this region, the system is prone to instability. We can use Lyapunov function v to determine ROA for a fixed control policy $\boldsymbol{\pi}$. Lyapunov function v is a continuously differentiable function with $v(0)=0$ and $v(\boldsymbol{x})>0$ for all $\boldsymbol{x}\neq\boldsymbol{0}$ [25]. Therefore, Lyapunov function is L_v-Lipschitz continuous. Based on the Lyapunov stability theory, we have the following theorem [23], [25].

Theorem 1 If $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\pi}(\boldsymbol{x})))<v(\boldsymbol{x})$ for all $\boldsymbol{x}$ within the level set $\boldsymbol{\Theta}(c)=\{\boldsymbol{x}\in\boldsymbol{\chi}\backslash\{\boldsymbol{0}\}|v(\boldsymbol{x})\leq c\}$ ($\boldsymbol{\chi}$ is the state space, $c>0$), then $\boldsymbol{\Theta}(c)$ is an ROA, so that $\boldsymbol{x}_0\in\boldsymbol{\Theta}(c)$ implies $\boldsymbol{x}_k\in\boldsymbol{\Theta}(c)$ for all $k>0$ and $\lim_{k\to\infty}\boldsymbol{x}_k=\boldsymbol{0}$.

The theorem indicates that when a fixed policy $\boldsymbol{\pi}$ is employed, applying the dynamics $\boldsymbol{f}(\cdot)$ to the state consistently results in decreasing values in the Lyapunov function. Consequently, the system state is assured to converge inevitably towards the equilibrium point. Further details can be found in [23]. According to the theorem, the determination of the ROA $\boldsymbol{\Theta}(c)$ is achieved by examining a level set of the Lyapunov function. To compute ROA, the crucial steps involve identifying an appropriate Lyapunov function and determining $\boldsymbol{\Theta}(c)$ that ensures the condition $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\pi}(\boldsymbol{x})))<v(\boldsymbol{x})$ holds for all $\boldsymbol{x}\in\boldsymbol{\Theta}(c)$.

The dynamics of VSG $\boldsymbol{f}(\cdot)$ are uncertain, leading to uncertainty in $v(\boldsymbol{f}(\cdot))$. This introduces an additional challenge in determining $\boldsymbol{\Theta}(c)$ using the above theorem. According to the GP model, $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}))$ is contained in $\Upsilon_n(\boldsymbol{x},\boldsymbol{u}):=[v(\boldsymbol{\mu}_{n-1}(\boldsymbol{x},\boldsymbol{u}))\pm L_v\beta_n\boldsymbol{\sigma}_{n-1}(\boldsymbol{x},\boldsymbol{u})]$ with probability higher than $1-\delta$. L_v is the Lipschitz constant of the Lyapunov function $v(\cdot)$. To ensure safe state-actions are always safe, we define the upper bound of $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}))$ as $u_n(\boldsymbol{x},\boldsymbol{u}):=\max C_n(\boldsymbol{x},\boldsymbol{u})$, where $C_n(\boldsymbol{x},\boldsymbol{u})=C_{n-1}(\boldsymbol{x},\boldsymbol{u})\cap\Upsilon_n(\boldsymbol{x},\boldsymbol{u})$. Therefore, in accordance with the aforementioned theorem and considering $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}))\leq u_n(\boldsymbol{x},\boldsymbol{u})$, the system stability in (1) is assured if $u_n(\boldsymbol{x},\boldsymbol{u})<v(\boldsymbol{x})$ is satisfied for all $\boldsymbol{x}\in\boldsymbol{\Theta}(c)$. Nevertheless, determining $\boldsymbol{\Theta}(c)$ becomes impractical when attempting to identify all states $\boldsymbol{x}$ on the continuous domain that satisfy $u_n(\boldsymbol{x},\boldsymbol{u})<v(\boldsymbol{x})$. To address this challenge, we can discretize the state space into cells denoted as $\boldsymbol{\chi}_\tau$, such that $\|\boldsymbol{x}-[\boldsymbol{x}]_\tau\|_1\leq\tau$. In this context, $[\boldsymbol{x}]_\tau$ represents the cell with the minimal distance to $\boldsymbol{x}$. Considering the system dynamic is L_f-Lipschitz continuous and the control policy is L_π-Lipschitz continuous, we can get the following theorem [17]. The proof is discussed in Appendix A.

Theorem 2 If $u_n(\boldsymbol{x},\boldsymbol{u})<v(\boldsymbol{x})-L_{\Delta v}\tau$ holds for all $\boldsymbol{x}\in\boldsymbol{\Theta}(c)\cap\boldsymbol{\chi}_\tau$ and for some $n\geq 0$, then $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\pi}(\boldsymbol{x})))<v(\boldsymbol{x})$ holds for all $\boldsymbol{x}\in\boldsymbol{\Theta}(c)$ with probability at least $1-\delta$, where $L_{\Delta v}=L_vL_f(L_\pi+1)+L_v$. And $\boldsymbol{\Theta}(c)$ is an ROA for the dynamics $\boldsymbol{f}$ under policy $\boldsymbol{\pi}$.

In this way, under a fixed policy $\boldsymbol{\pi}$, the ROA can be identified within the discretized state space as follows:

$$\boldsymbol{D}_n=\{(\boldsymbol{x},\boldsymbol{u})|u_n(\boldsymbol{x},\boldsymbol{\pi}(\boldsymbol{x}))-v(\boldsymbol{x})<-L_{\Delta v}\tau\} \tag{5}$$

It should be noted that the ROA is dependent on the policy. To get the largest possible ROA, we can optimize the policy using (6). The corresponding optimal policy for c_n is $\boldsymbol{\pi}_n$.

$$c_n = \max_{\boldsymbol{\pi} \in \Pi_P, c \in \mathbf{R}_{>0}} c \quad \forall \boldsymbol{x} \in \boldsymbol{\Theta}(c) \bigcap \boldsymbol{\chi}_\tau, (\boldsymbol{x}, \boldsymbol{\pi}(\boldsymbol{x})) \in \boldsymbol{D}_n \tag{6}$$

where Π_P is the set of safe policies.

The ROA optimized by (6) is contained in true ROA with probability at least $1-\delta$ for all $n>0$. Precisely solving (6) is intractable, thus we adopt the ADP [18] technique to improve the performance of the policy from data, as shown below:

$$\boldsymbol{\pi}_n = \arg\min_{\boldsymbol{\pi}_W \in \Pi_P} \sum_{x \in \boldsymbol{\chi}_\tau} r(\boldsymbol{x}, \boldsymbol{\pi}_W(\boldsymbol{x})) + \gamma J_{\pi_W}(\boldsymbol{\mu}_{n-1}(\boldsymbol{x}, \boldsymbol{\pi}_W(\boldsymbol{x}))) + \lambda(u_n(\boldsymbol{x}, \boldsymbol{\pi}_W(\boldsymbol{x})) - v(\boldsymbol{x}) + L_{\Delta v}\tau) \tag{7}$$

where $\boldsymbol{\pi}_W$ is the policy with parameters $\boldsymbol{W}$; γ is the discount factor; λ is a Lagrange multiplier for the safety constraint; $r(\boldsymbol{x}, \boldsymbol{\pi}_W(\boldsymbol{x})) = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{R} \boldsymbol{u} + \boldsymbol{x}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{x} \geq 0$ is the cost function; and $J_{\pi_W}(\cdot)$ is the value function of the Bellman's equation, which is approximated using piecewise linear approximations [18] in this work, and $J_{\pi_W}(\boldsymbol{x}) = r(\boldsymbol{x}, \boldsymbol{\pi}(\boldsymbol{x})) + \gamma J_{\pi_W}(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\pi}(\boldsymbol{x})))$. Considering the cost function is strictly positive, we use $J_{\pi_W}(\cdot)$ as the Lyapunov function. In (7), the objective of the optimization is to minimize the cost and make sure the safety constraint holds, and stochastic gradient descent (SGD) based optimization method can be utilized.

For the proposed algorithm, a safe initial point is essential for initiating the learning process. Consequently, an initial policy is required, ensuring the asymptotic stability of the system origin in (1) within a confined set of states. In this work, we utilize a linear-quadratic regulator (LQR) controller as our initial policy. In addition, to expand the ROA throughout the learning process, the agent strategically explores the state-action pairs for which the system dynamics are most uncertain. To achieve this, we meticulously choose measurement data points based on:

$$(\boldsymbol{x}_n, \boldsymbol{u}_n) = \arg\max_{(\boldsymbol{x}, \boldsymbol{u}) \in \boldsymbol{D}_n} (u_n(\boldsymbol{x}, \boldsymbol{u}) - l_n(\boldsymbol{x}, \boldsymbol{u})) \tag{8}$$

where $l_n(\boldsymbol{x}, \boldsymbol{u})$ is the lower bound of $v(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}))$.

The proposed algorithm is summarized in Algorithm 1.

Algorithm 1: safe MBRL algorithm for GFM inverter based frequency regulation

Load the power system simulation environment; initialize the LQR-based initial policy; initialize the parameters of the policy $\boldsymbol{\pi}_W$; initialize the GP model for VSG dynamics and ADP value functions; set the total number of episodes N_e; and set the training step $n=1$

Get the initial safe set based on the initial LQR controller and the corresponding initial Lyapunov function

for $n \leq N_e$ **do**

  for $i = 1, 2, \ldots, N$ **do**

    Based on (8), select a new safe sample of the state-action pair $(\boldsymbol{x}, \boldsymbol{u})$

    Update the GP model for VSG dynamics based on the actively selected new data point

    Optimize policy $\boldsymbol{\pi}_n$ by solving (7) using the SGD-based optimization method

    Update the Lyapunov function (i.e., value function $J_{\pi_W}(\cdot)$)

    Using the updated policy, calculate c_n in (6) to ensure that $\forall \boldsymbol{x} \in \boldsymbol{\Theta}(c) \bigcap \boldsymbol{\chi}_\tau$, $u_n(\boldsymbol{x}, \boldsymbol{\pi}(\boldsymbol{x})) - v(\boldsymbol{x}) < -L_{\Delta v}\tau$ holds

    Compute and update the safe set (i.e., ROA)

Return the well-trained policy $\boldsymbol{\pi}_W$

IV. Simulation Results

A case study was conducted on a GFM inverter system, as shown in Fig. 1, to demonstrate the effectiveness of the proposed algorithm for system frequency regulation. The step size of the discrete simulation of the system was set to be 0.01 s and the total simulation time horizon was 15 s. We used GP model to learn the frequency dynamics of the VSG. The mean dynamics of the VSG were characterized by a linearized model of the true dynamics, as shown in (B1) in Appendix B, accounting for inaccuracies in the values of M and D. Consequently, the optimal policy designed for the mean dynamics exhibited suboptimal performance with a limited ROA, primarily due to underactuation of the system. We adopted a hybrid approach employing both linear and Matérn kernels (refer to Appendix C) [22], [26]. This combination enabled us to effectively capture model errors stemming from inaccuracies in parameters. As for the policy network, a neural network featuring two hidden layers was implemented, each comprising 32 neurons with rectified linear unit (ReLU) activation functions. The states θ and ω were discretized into 2000 and 1500 intervals, respectively. The action space was discretized into 55 intervals. $\boldsymbol{R}$ and $\boldsymbol{Q}$ in (3) were set to be $\mathbf{0.1}$ and $\begin{bmatrix} 0.1 & 0 \\ 0 & 2 \end{bmatrix}$, respectively.

The case study was conducted on an Intel Core i7-8650U @ 1.90 GHz Windows based computer with 16 GB RAM. The convergence process of the training for the proposed algorithm is illustrated in Fig. 3.

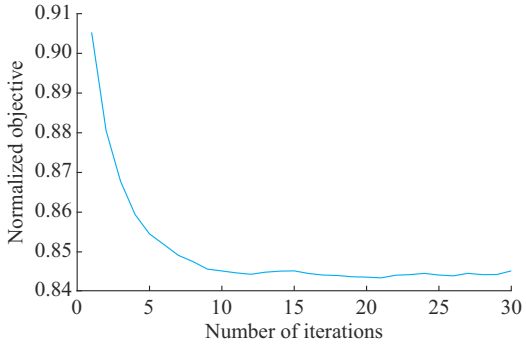

Fig. 3. Convergence process of training for proposed algorithm.

The proposed algorithm exhibited remarkable convergence, typically requiring only a few tens of iterations. Under the obtained control policy, the ROA is shown in Fig. 4 by the dark green area, where the light green area denotes the state space and the blue cross marks denote the data points the agent selected to explore the safe region. From the result, the ROA was determined based on the information from multiple measurements.

We investigated the frequency control performance of the proposed algorithm, as depicted in Fig. 5, where the frequency deviation f is derived from the angular frequency deviation ω, with the relationship expressed as $f = \omega/(2\pi)$, so in Fig. 5(a)-(c), the values of $\omega_{t \to 0^+}$ are -1, -2, and -3 rad/s, respectively. From Fig. 5, it is evident that when the inverter experiences frequency deviation, the proposed algorithm efficiently restores the system to a stable state using BESSs. In

contrast, without any control, the system became unstable after the disturbance. Additionally, while traditional linear droop control can stabilize the system under certain levels of disturbance, it fails to maintain stability when the disturbance is significant, as shown in Fig. 5(c). The results also indicate that the linearized control policy could lead to a rapid deterioration in control performance in the presence of large frequency deviations.

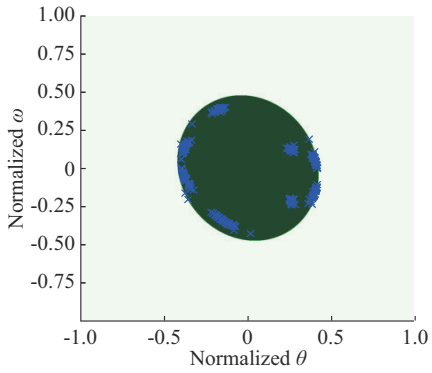

Fig. 4. ROA under safe MBRL-based control policy.

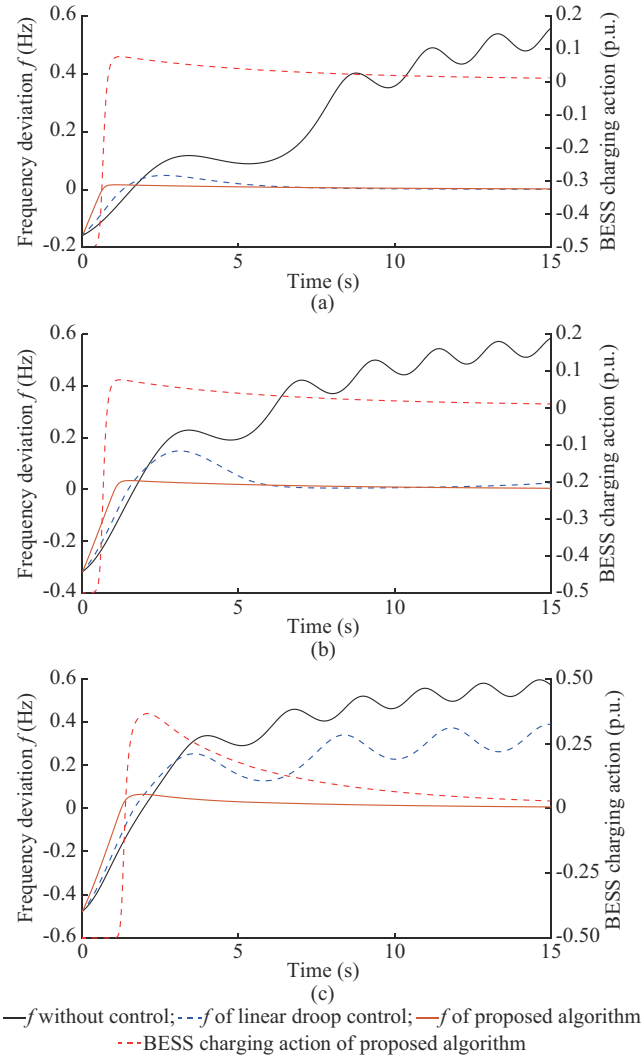

Fig. 5. BESS charging action of proposed algorithm and frequency control performance of proposed algorithm and comparing algorithms. (a) $f_{t\to 0^+}$ = −0.159 Hz. (b) $f_{t\to 0^+}$ = −0.318 Hz. (c) $f_{t\to 0^+}$ = −0.478 Hz.

To demonstrate the superiority of the proposed algorithm over traditional model-free DRL algorithms (e. g., the deep deterministic policy gradient (DDPG) algorithm outlined in [14]), we conducted a comparative analysis of the proposed algorithm against DDPG and soft actor critic (SAC) algorithms for frequency regulation. The results are depicted in Fig. 6. The analysis reveals that while the DDPG and SAC algorithms achieve satisfactory control performance under relatively mild disturbances, as shown in Fig. 6(a) and (b), managing to stabilize the inverter frequency within several seconds after the disturbance, their effectiveness diminishes with increasing disturbance magnitude. In contrast, the proposed algorithm not only restores inverter frequency more swiftly than the model-free algorithms in scenarios with relatively minor disturbances but also maintains robust frequency control under more significant disturbances, as shown in Fig. 6(c).

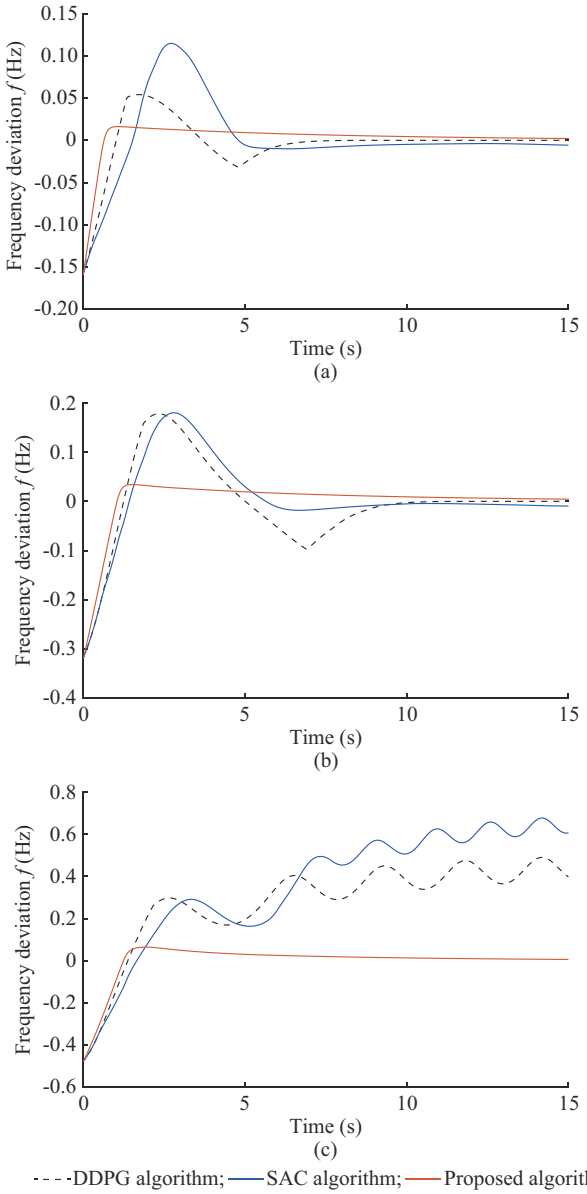

Fig. 6. Frequency control performance of proposed algorithm and model-free DRL algorithms. (a) $f_{t\to 0^+}$ = −0.159 Hz. (b) $f_{t\to 0^+}$ = −0.318 Hz. (c) $f_{t\to 0^+}$ = −0.478 Hz.

The stable control performance of the proposed algorithm can be largely attributed to the integration of Lyapunov stability theory into the learning process, which provides a safety guarantee characteristic. More specifically, the proposed algorithm selects optimal control actions within the ROA, ensuring a level of safety that model-free DRL algorithms cannot guarantee for the learned policy.

Furthermore, to test the robustness of the proposed algorithm against inverter parameter variations, such as M and D in (1), we evaluated the performance of the well-trained safe MBRL controller under different parameter settings. Figure 7(a) illustrates the frequency response of the inverter with varying D values ($70\% \cdot D_{base} \le D \le 130\% \cdot D_{base}$) while the virtual inertia setting was held constant at M_{base}. In Fig. 7(b), the frequency response to varying M values, deviating by ±30% from the base value, was examined, while maintaining the damping coefficient steady at D_{base}. Observations from Fig. 7(a) and (b) indicate that the safe MBRL-based control policy was able to effectively and safely control the BESS to provide frequency regulation, regardless of the M and D adjustments. This adaptability underscores the capability of the controller to handle dynamic changes and uncertainties within the system, affirming its robustness against a wide range of operational conditions.

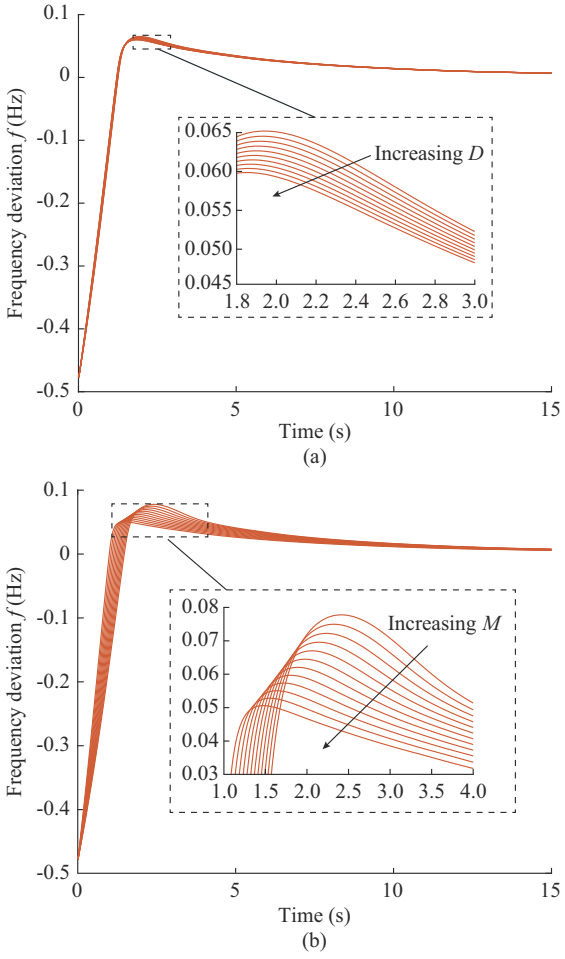

Fig. 7. Robustness of safe MBRL-based control policy against D and M uncertainties with $f_{t\to 0^+}$=−0.478 Hz. (a) Robustness of safe MBRL-based control policy against D uncertainty. (b) Robustness of safe MBRL-based control policy against M uncertainty.

V. Conclusion

In this paper, we presented a novel safe MBRL algorithm for GFM inverter based frequency regulation with stability guarantee. The proposed algorithm ensures stability by learning a Lyapunov function and utilizes ADP-based RL to enhance control performance. Additionally, the GP modeling was employed to capture VSG dynamics and enhance robustness to parameter uncertainty. The proposed algorithm offers a safe and robust controller for GFM inverter based frequency regulation. Simulation results demonstrated that the performance of the proposed algorithm surpasses that of traditional droop control and model-free DRL algorithms. Moreover, the proposed algorithm only requires the measurements of the voltage phase and angular frequency of the inverter, which are easily accessible in modern power systems. The ease of implementation of the proposed algorithm enhances its potential for practical applications.

Appendix A

Lemma 1 The control policy $\boldsymbol{\pi}_{\boldsymbol{W}}$ is Lipschitz continuous with Lipschitz constant L_π.

Proof 1 In this work, $\boldsymbol{\pi}_{\boldsymbol{W}}=\boldsymbol{\phi}(\boldsymbol{x})$. $\boldsymbol{\phi}(\boldsymbol{x})$ is the output of a K-layer network, which is given by:

$$\boldsymbol{\phi}(\boldsymbol{x})=\boldsymbol{\phi}_K(\boldsymbol{\phi}_{K-1}(\ldots\boldsymbol{\phi}_1(\boldsymbol{x};\boldsymbol{W}_1);\boldsymbol{W}_2)\ldots;\boldsymbol{W}_K) \tag{A1}$$

In the hidden layers, ReLU activation functions are used. For the k^{th} layer, there exits a constant $L_k>0$ such that $\|\boldsymbol{\phi}_k(\boldsymbol{x};\boldsymbol{W}_k)-\boldsymbol{\phi}_k(\boldsymbol{x}+\boldsymbol{r};\boldsymbol{W}_k)\|\le L_k\|\boldsymbol{r}\|$ holds for all $\boldsymbol{x}$ and $\boldsymbol{r}$. Here, $\boldsymbol{r}$ is a vector satisfying $\|\boldsymbol{r}\|\le\varepsilon$ and ε is a small enough positive number. The output layer utilizes tanh activation function, thus the network satisfies $\|\boldsymbol{\phi}(\boldsymbol{x})-\boldsymbol{\phi}(\boldsymbol{x}+\boldsymbol{r})\|\le L_\pi\|\boldsymbol{r}\|$, with $L_\pi=\prod_{k=1}^{K}L_k$. This means the control policy $\boldsymbol{\pi}_{\boldsymbol{W}}$ is Lipschitz continuous with Lipschitz constant L_π.

Lemma 2 The closed-loop dynamics of the VSG given in (4b) are Lipschitz continuous with Lipschitz constant L_f.

Proof 2 From the dynamics given in (1) and Lemma 1, the dynamic function of VSG is a continuously differentiable function. Any continuously differentiable function is locally Lipschitz. Therefore, the closed-loop dynamics of VSG given in (4b) are Lipschitz continuous with Lipschitz constant L_f.

Lemma 3 The Lyapunov function v is Lipschitz continuous with Lipschitz constant L_v.

Proof 3 In this work, the Lyapunov function is set as the value function $J_{\boldsymbol{\pi}_{\boldsymbol{W}}}$ of the ADP method. The value function is approximated using a piecewise linear function that is continuous. Given that the slopes of this piecewise linear function are bounded, the Lyapunov function exhibits Lipschitz continuity with a Lipschitz constant denoted by L_v.

Theorem 2 can be proofed as follows. According to Lemma 1 of [17], $v(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\pi}(\boldsymbol{x})))-v(\boldsymbol{x})<0$ for all continuous states $\boldsymbol{x}\in\boldsymbol{\Theta}(c)$ with probability higher than $1-\delta$. So, it can be concluded based on Theorem 1 that $\boldsymbol{\Theta}(c)$ is an ROA for the system.

Appendix B

The LQR-based initial policy is designed based on the linearized VSG dynamics. According to formulas (1) and (2), the

linearized small-signal model of VSG around an given operating point is obtained as:

$$\begin{bmatrix} \Delta\dot{\theta} \\ \Delta\dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{M}(B\cos\theta - G\sin\theta) & -\frac{D}{M} \end{bmatrix} \begin{bmatrix} \Delta\theta \\ \Delta\omega \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{M} \end{bmatrix} \boldsymbol{u} \tag{B1}$$

The eigenvalues of the system are:

$$\lambda_{1,2} = \frac{-D \pm \sqrt{D^2 - 4M(B\cos\theta - G\sin\theta)}}{2M} \tag{B2}$$

where $G + \mathrm{j}B = Y_{ig}$ is the mutual admittance between the IBR node and the main grid. As shown in Fig. 1, the mutual admittance can be calculated using the line parameters as [21]:

$$Y_{ig} = -\frac{1}{R_c + \mathrm{j}X_c} = \frac{-R_c}{R_c^2 + X_c^2} + \mathrm{j}\frac{X_c}{R_c^2 + X_c^2} \tag{B3}$$

It can be found that the eigenvalues depend on the operating point, virtual inertia, and damping coefficients, and line parameters R_c and X_c. In this work, $Y_{ig} = -0.495 + \mathrm{j}4.95$. The per unit values of M and D are set to be 5 and 1, respectively.

APPENDIX C

The linear kernel is given by:

$$k_L(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}' \tag{C1}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are two distinct states.

The Matérn kernel is given by:

$$k_M(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{\Gamma(\hat{v})2^{v-1}} \left(\frac{\sqrt{2\hat{v}}}{l} d(\boldsymbol{x}, \boldsymbol{x}') \right)^{\hat{v}} K_v \left(\frac{\sqrt{2\hat{v}}}{l} d(\boldsymbol{x}, \boldsymbol{x}') \right) \tag{C2}$$

where l is a length-scale parameter; $d(\cdot)$ is the Euclidean distance; $K_v(\cdot)$ is a modified Bessel function; $\Gamma(\cdot)$ is the Gamma function; and $\hat{v}$ is the parameter that regulates the smoothness of the function.

REFERENCES

[1] A. Bidram, A. Davoudi, and F. L. Lewis, "A multiobjective distributed control framework for islanded AC microgrids," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 3, pp. 1785-1798, Aug. 2014.

[2] D. Chen, K. Chen, Z. Li *et al*., "PowerNet: multi-agent deep reinforcement learning for scalable power grid control," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1007-1017, Mar. 2022.

[3] Z. A. Obaid, L. M. Cipcigan, L. Abrahim *et al*., "Frequency control of future power systems: reviewing and evaluating challenges and new control methods," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 1, pp. 9-25, Jan. 2019.

[4] P. Verma, K. Seethalekshmi, and B. Dwivedi, "A cooperative approach of frequency regulation through virtual inertia control and enhancement of low voltage ride-through in DFIG-based wind farm," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 6, pp. 1519-1530, Nov. 2022.

[5] X. Meng, J. Liu, and Z. Liu, "A generalized droop control for grid-supporting inverter based on comparison between traditional droop control and virtual synchronous generator control," *IEEE Transactions on Power Electronics*, vol. 34, no. 6, pp. 5416-5438, Jun. 2019.

[6] J. Liu, Y. Miura, H. Bevrani *et al*., "Enhanced virtual synchronous generator control for parallel inverters in microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2268-2277, Sept. 2017.

[7] K. Sakimoto, Y. Miura, and T. Ise, "Stabilization of a power system with a distributed generator by a virtual synchronous generator function," in *Proceedings of 8th International Conference on Power Electronics*, Jeju, South Korea, Jun. 2011, pp. 1498-1505.

[8] P. He, Z. Li, H. Jin *et al*., "An adaptive VSG control strategy of battery energy storage system for power system frequency stability enhancement," *International Journal of Electrical Power & Energy Systems*, vol. 149, p. 109039, Jul. 2023.

[9] M. Li, W. Huang, N. Tai *et al*., "A dual-adaptivity inertia control strategy for virtual synchronous generator," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 594-604, Jan. 2020.

[10] J. Alipoor, Y. Miura, and T. Ise, "Power system stabilization using virtual synchronous generator with alternating moment of inertia," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 3, no. 2, pp. 451-458, Jun. 2015.

[11] F. Wang, L. Zhang, X. Feng *et al*., "An adaptive control strategy for virtual synchronous generator," *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 5124-5133, Sept. 2018.

[12] A. Ademola-Idowu and B. Zhang, "Frequency stability using MPC-based inverter power control in low-inertia power systems," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1628-1637, Mar. 2021.

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

[14] Y. Li, W. Gao, W. Yan *et al*., "Data-driven optimal control strategy for virtual synchronous generator via deep reinforcement learning approach," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 4, pp. 919-929, Aug. 2021.

[15] W. Cui, Y. Jiang, and B. Zhang, "Reinforcement learning for optimal primary frequency control: a Lyapunov approach," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1676-1688, Mar. 2023.

[16] W. Cui and B. Zhang, "Lyapunov-regularized reinforcement learning for power system transient stability," *IEEE Control Systems Letters*, vol. 6, pp. 974-979, Jun. 2022.

[17] F. Berkenkamp, M. Turchetta, A. Schoellig *et al*., "Safe model-based reinforcement learning with stability guarantees," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, California, USA, Dec. 2017, pp. 908-919.

[18] H. Shuai, J. Fang, X. Ai *et al*., "Stochastic optimization of economic dispatch for microgrid based on approximate dynamic programming," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2440-2452, May 2019.

[19] H. Shuai, J. Fang, X. Ai *et al*., "Optimal real-time operation strategy for microgrid: an ADP-based stochastic nonlinear optimization approach," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 931-942, Apr. 2019.

[20] D. Raisz, D. Deepak, F. Ponci *et al*., "Linear and uniform swing dynamics in multimachine converter-based power systems," *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106475, Feb. 2021.

[21] V. Vittal, J. D. McCalley, P. M. Anderson *et al*., *Power System Control and Stability*. New York: John Wiley & Sons, 2019.

[22] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge: The MIT Press, 2005.

[23] F. Berkenkamp, R. Moriconi, A. P. Schoellig *et al*., "Safe learning of regions of attraction for uncertain, nonlinear systems with Gaussian processes," in *Proceedings of 2016 IEEE 55th Conference on Decision and Control*, Las Vegas, USA, Dec. 2016, pp. 4661-4666.

[24] B. She, J. Liu, F. Qiu *et al*., "Systematic controller design for inverter-based microgrids with certified large-signal stability and domain of attraction," *IEEE Transactions on Smart Grid*, doi: 10.1109/TSG.2023.3330705

[25] H. K. Khalil, *Nonlinear Systems*. London: Prentice Hall, 1996.

[26] D. Duvenaud. (2014, May). The kernel cookbook: advice on covariance functions. [Online]. Available: https://www.cs.toronto.edu/duvenaud/cookbook

Hang Shuai received the B.Eng. degree from Wuhan Institute of Technology (WIT), Wuhan, China, in 2013, and the Ph.D. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2019. He was also a Visiting Student Researcher with the University of Rhode Island (URI), Kingston, USA, from 2018 to 2019. He was a Postdoctoral Researcher with the URI and University of Tennessee (UTK), Knoxville, USA, from 2019 to 2022. Currently, he is a Research Assistant Professor with the UTK. His research interests include reinforcement learning for power system, microgrid operation and control, and bulk power system resilience.

Buxin She received the B.S.E.E and M.S.E.E degrees from Tianjin Universi-

ty, Tianjin, China, in 2017 and 2019, respectively, and the Ph. D. degree from the University of Tennessee, Knoxville, USA, in 2023, all in electrical engineering. He is currently a Research Engineer in Pacific Northwest National Laboratory (PNNL), Richland, USA. He served as a Student Guest Editor of IET-RPG. He was an outstanding reviewer of IEEE Access Journal of Power and Energy (OAJPE) (2020) and Journal of Modern Power Systems and Clean Energy (MPCE) (2022 and 2023). His research interests include microgrid operation and control, machine learning in power systems, distribution system operation and plan, and power grid resilience.

Jinning Wang received the B.S. and M.S. degrees in electrical engineering from Taiyuan University of Technology, Taiyuan, China, in 2017 and 2020, respectively. He is currently pursuing a Ph.D. degree in electrical engineering at the University of Tennessee, Knoxville, USA. He is the author of AMS, a power system dispatch simulator, which is a key component of the CURENT Large-scale Testbed. He also curates the list Popular Open Source Libraries for Power System Analysis. His research interests include data mining, scientific computation, and power system simulation.

Fangxing Li (also known as Fran Li) received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree from Virginia Tech, Blacksburg, USA, in 2001. He is currently the John W. Fisher Professor of electrical engineering and the Director of CURENT with the University of Tennessee, Knoxville, USA. From 2020 to 2021, he was the Chair of IEEE PES Power System Operation, Planning and Economics (PSOPE) Committee. He has been the Chair of IEEE WG on Machine Learning for Power Systems since 2019 and the Editor-in-Chief of IEEE Open Access Journal of Power and Energy (OAJPE) since 2020. He was the recipient of numerous awards and honors, including R&D 100 Award in 2020, IEEE PES Technical Committee Prize Paper awards in 2019 and 2024, five best or prize paper awards at international journals, and seven best papers/posters at international conferences. His research interests include resilience, artificial intelligence in power, demand response, distributed generation and microgrid, and electricity market.