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 letter investigates a safe reinforcement learning strategy for grid-forming (GFM) inverter based frequency regulation. To guarantee stability of the inverter based resource (IBR) system under the learned control policy, a model based reinforcement learning (MBRL) technique is combined with Lyapunov approach which determines safe region of states and actions. To obtain near optimal control strategy, 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 MBRL to effectively learn system dynamics from measurements. Numerical simulations validate the effectiveness of the proposed method.**

***Index Terms*—Inverter based resource (IBR), virtual synchronous generator (VSG), safe reinforcement learning, Lyapunov function.**

I. Introduction

POWER system frequency control is critical for maintaining grid stability when imbalance between generation and load occurs. As the penetration of IBR, 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 IBR 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 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 VSG are more flexible than 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 reference [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 above mentioned conventional model-free RL based controllers do not yield any stability guarantees. Therefore, reference [15] proposed a Lyapunov based model-free RL strategy for power system primary frequency control, which can guarantee the system frequency will reach stable equilibrium after disturbances. In [15] and [16], Lyapunov stability theory is 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 region need 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 MBRL strategy for GFM inverter based frequency regulation. Motivated by [17], this strategy addresses the challenges of ensuring controller stability and effectively dealing with system parameter uncertainty. In the developed MBRL controller, GP model is adopted to learn the unknown nonlinear dynamics of the inverter system, and ADP [18], [19] is used to improve the performance of the controller. Moreover, to guarantee the system stability under the learned controller, Lyapunov function is used to obtain the 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 MBRL agent to update during training, leading to an enlarged ROA and improved control performance simultaneously. In addition, the proposed

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

This work was funded in part by the CURENT research center and in part by the NSF grant 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

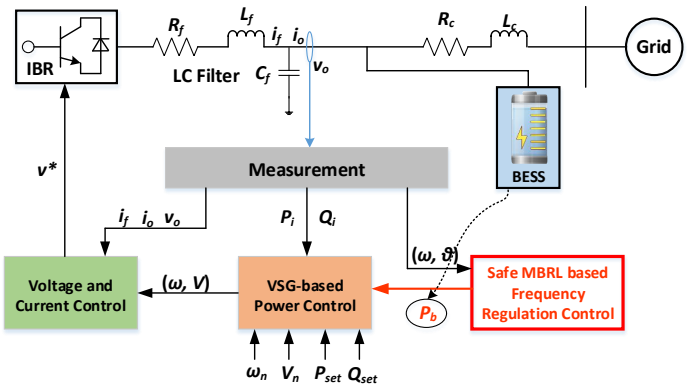

Fig. 1. Diagram of the GFM inverter based primary frequency control.

MBRL controller is adaptive to the adjusting of inverter parameters (i.e., virtual inertia and damping coefficients), which means the controller will be more robust to parameter uncertainty.

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

II. GFM Inverter Based Frequency Regulation

The IBR primary control diagram is depicted in Fig. 1. We assume the bus voltage magnitudes to be 1 per unit (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 equation [5], [16], [20]:

$$\begin{aligned} \frac{d\theta}{dt} &= \omega \\ M \cdot \frac{d\omega}{dt} &= P_{set} - P_i - D \cdot \omega - u(\theta, \omega) \end{aligned} \tag{1}$$

where, 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 denotes the generated angular frequency of the inverter output voltage, and ω_n represents the nominal angular frequency of the inverter. $u(\cdot)$ denotes the control action which is the active charging power (i.e., P_b in Fig. 1) of the battery energy storage systems (BESS). M and D are the virtual inertia and damping constant of the inverter, respectively. P_{set} and P_i are the active power set point and real-time measurement of the inverter's active power output, respectively. In Fig. 1, P_i can be calculated as follows [21]:

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

The element (i, j) of the admittance matrix Y, denoted as Y_{ij}, is defined by $Y_{ij} = G_{ij} + jB_{ij}$, where B_{ij} and G_{ij} represent the susceptance and conductance components, respectively. θ_g is the voltage phase of the main grid. Note that lossy power flow model is adopted in Eq. (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 follows:

$$\begin{aligned} &\min_{\mathbf{u}} (u^T R u + x^T Q x) \\ &s.t. \quad Eq.\ (1),\ and\ \underline{u} \le u \le \bar{u},\ and\ u \text{ is stabilizing} \end{aligned} \tag{3}$$

where $x = (\theta, \omega)$ is the state of the VSG. Q and R are positive definite matrices. $\underline{u}$ and $\bar{u}$ are the lower and upper limitations of the control action, which determined by the BESS maximum power capacity. As shown in Fig. 1, the optimal action $u(\cdot)$ is optimized using the safe MBRL based agent.

III. GFM Inverter Based Frequency Regulation Controller Via Safe Model Based Reinforcement Learning

The primary objective of the controller is to safely learn about the frequency dynamics of VSG from measurements and adapt the control policy π 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's state remains within the ROA. The parameter uncertainty and nonlinearity of the AC power flow, as described in Eq. (2), make the design of controllers for Eq. (1) challenging. The proposed safe MBRL controller for GFM inverter based frequency regulation is depicted in Fig. 2. In the proposed controller, the frequency dynamics of VSG is 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 reinforcement learning approach to expand the ROA. The details of the proposed method are presented below.

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

$$\begin{aligned} \theta_{k+1} &= \theta_k + h \cdot \omega_k \\ \omega_{k+1} &= \omega_k + \frac{h}{M}(P_{set,k} - P_{i,k} - D \cdot \omega_k - u_k) \end{aligned} \tag{4a}$$

where, h is the step size for the discrete simulation. The subscript k denotes the discrete time index. Eq. (4a) can be expressed as

$$x_{k+1} = f(x_k, u_k) = h(x_k, u_k) + g(x_k, u_k) \tag{4b}$$

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

To enable safe learning, we adopt GP model to learn a reliable statistical system model described by Eqs. (1) and (2). GP model is a powerful method in machine learning and statistical modeling. GPs consist of random variables, and any finite group of them follows a joint Gaussian distribution. In system modeling, GPs are 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 $||f(x,u)-\mu_n(x,u)||_1 \leq \beta_n \sigma_n(x,u)$. $\mu_n(\cdot)$ and $\sigma_n(\cdot) = trace(\sum_n^{1/2}(\cdot))$ are the posterior mean and covariance matrix functions of the GP model of the VSG dynamics in Eq. (4b) conditioned on n measurements. In this way, we can use GP models to build confidence intervals on the inverter dynamics which can cover the true dynamics with probability $1-\delta$.

After learned 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's state falls within the boundaries of the ROA, the dynamics outlined in Eq. (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 π. Lyapunov function v is a continuously differentiable function with $v(0) = 0$ and $v(x) > 0$ for all $x \neq 0$ [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 $v(f(x,\pi(x))) < v(x)$ for all x within the level set $\Theta(c) = \{x \in \chi \backslash \{0\} | v(x) \leq c\}$, $c > 0$, then $\Theta(c)$ is a ROA, so that $x_0 \in \Theta(c)$ implies $x_k \in \Theta(c)$ for all $k > 0$ and $\lim_{k\to\infty} x_k = 0$.*

The theorem indicates that when a fixed policy π is employed, applying the dynamics f 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 is achieved by examining a level set of the Lyapunov function, denoted as $\Theta(c)$. To compute ROA, the crucial steps involve identifying an appropriate Lyapunov function and determining $\Theta(c)$ that ensures the condition $v(f(x,\pi(x))) < v(x)$ holds for all $x \in \Theta(c)$.

The dynamics of VSG, denoted as $f(\cdot)$, are uncertain, leading to uncertainty in $v(f(\cdot))$. This introduces an additional challenge in determining $\Theta(c)$ using the above theorem. According to the GP model, $v(f(x,u))$ is contained in $\Upsilon_n(x,u) := [v(\mu_{n-1}(x,u)) \pm L_v \beta_n \sigma_{n-1}(x,u)]$ 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 $v(f(x,u))$ as $u_n(x,u) := \max C_n(x,u)$, where $C_n(x,u) = C_{n-1}(x,u) \bigcap \Upsilon_n(x,u)$. Therefore, in accordance with the aforementioned theorem and considering $v(f(x,u)) \leq u_n(x,u)$, the system's stability in Eq. (1) is assured if $u_n(x,u) < v(x)$ is satisfied for all $x \in \Theta(c)$. Nevertheless, determining $\Theta(c)$ becomes impractical when attempting to identify all states x on the continuous domain that satisfy $u_n(x,u) < v(x)$. To address this challenge, we can discretize the state space into cells denoted as χ_τ, such that $|| x - [x]_\tau ||_1 \leq \tau$. In this context, $[x]_\tau$ represents the cell with the minimal distance to 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 III.

Theorem 2: *If for all $x \in \Theta(c) \bigcap \chi_\tau$ and for some $n \geq 0$ it holds that $u_n(x,u) < v(x) - L_{\Delta v}\tau$, then $v(f(x,\pi(x))) < v(x)$ holds for all $x \in \Theta(c)$ with probability at least $(1-\delta)$, where $L_{\Delta v} = L_v L_f (L_\pi + 1) + L_v$. And $\Theta(c)$ is a region of attraction for the dynamics f under policy π.*

In this way, under a fixed policy π, the ROA can be identified within the discretized state space as follows:

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

It should be noted that the ROA is dependent on the policy. To get the largest possible ROA, we can optimize the policy using:

$$\pi_n, c_n = \underset{\pi \in \Pi_P, c \in R_{>0}}{\arg\max} \; c, \quad \text{for all } x \in \Theta(c) \bigcap \chi_\tau \;, (x,\pi(x)) \in D_n \tag{6}$$

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

$$\begin{aligned}\pi_n = \underset{\pi_W \in \Pi_P}{\arg\min} \sum_{x \in \chi_\tau} & r(x,\pi_W(x)) + \gamma J_{\pi_W}(\mu_{n-1}(x,\pi_W(x))) + \\ & \lambda(u_n(x,\pi_W(x)) - v(x) + L_{\Delta v}\tau)\end{aligned} \tag{7}$$

where, π_W is the policy with parameters W. $r(x,\pi_W(x)) = u^T R u + x^T Q x \geq 0$ is the cost function. $J_{\pi_W}(\cdot)$ is the value function of the Bellman's equation, which is approximated using piecewise linear approximations [18] in this work, and $J(x) = r(x,\pi(x)) + \gamma J(f(x,\pi(x)))$. μ_n is the mean dynamics of the inverter represented by the GP model. Considering the cost function is strictly positive, we use the above $J(\cdot)$ as the Lyapunov function. λ is a Lagrange multiplier for the safety constraint. In Eq. (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 MBRL based frequency regulation controller, 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 Eq. (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:

$$(x_n, u_n) = \underset{(x,u) \in D_n}{\arg\max} [u_n(x,u) - l_n(x,u)] \tag{8}$$

where, $l_n(x,u)$ is the lower bounds of $v(f(x,u))$. The proposed approach is summarized in **Algorithm 1**.

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 safe MBRL algorithm for system frequency regulation. The step size of the system discrete simulation was

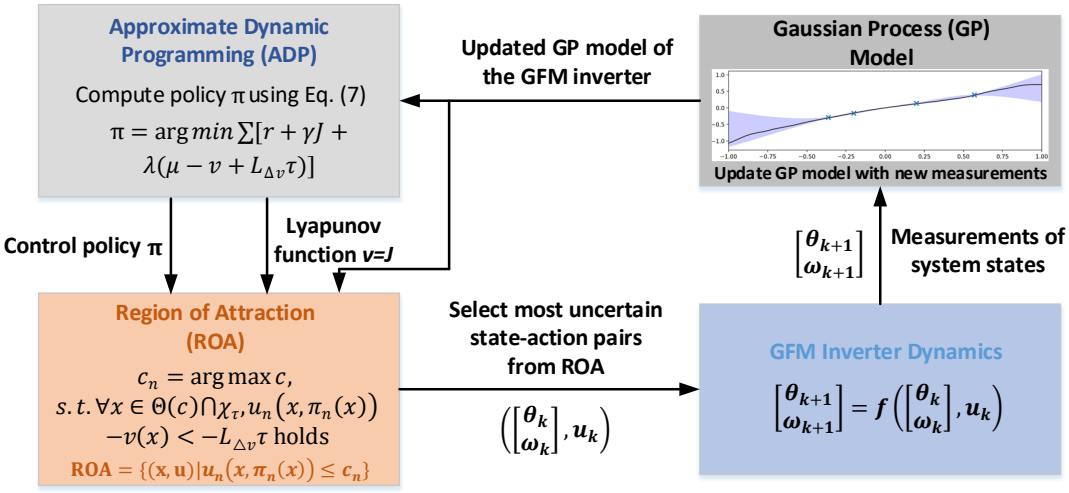

Fig. 2. The developed safe MBRL strategy for GFM inverter based frequency regulation.

Algorithm 1 Safe MBRL based algorithm for GFM inverter based frequency regulation.

1: Load the power system simulation environment; Initialize the LQR based initial policy; Initialize the parameters of the policy π_W; Initialize GP based dynamics model and ADP value functions; Set the total number of episode N_e, and set the training step $n = 1$.
2: Get the initial safe set based on the initial LQR controller and the corresponding initial Lyapunov function;
3: **for** $n \leq N_e$ **do**
4: **for** $i = 1, 2, \cdots, N$ **do**
5: Based on Eq. (8), select a new safe sample of the state-action pair (x, u).
6: Update the GP model for VSG dynamics based on the actively selected new data point.
7: Optimize policy π_n by solving Eq. (7) using the SGD method.
8: Update the Lyapunov function (i.e., value function J).
9: Using the updated policy, calculate c_n in Eq. (6) to ensure that $\forall\, x \in \Theta(c) \bigcap \chi_\tau$, $u_n(x, \pi_n(x)) - v(x) < -L_{\Delta v}\tau$ holds.
10: Compute and update the safe set (i.e., ROA).
11: Return the well-trained policy π_W.

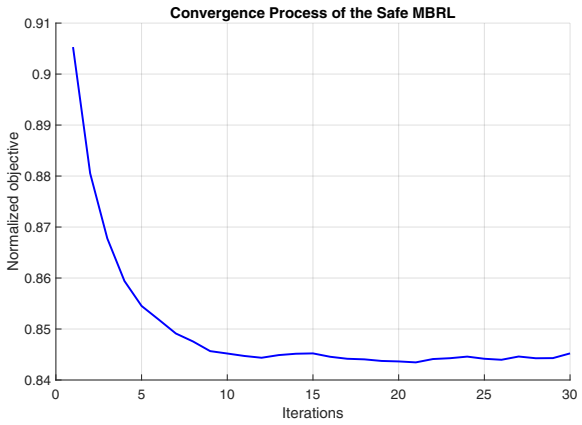

Fig. 3. The convergence of the training process for the developed safe MBRL algorithm.

set to 0.01s and the total simulation time horizon was 15s. 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 (see Eq. (A3) in Appendix II), 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 region of attraction, primarily due to underactuation of the system. We adopted a hybrid approach employing both linear and Matérn kernels (refer to Appendix I) [22], [26]. This combination enabled us to effectively capture model errors stemming from inaccuracies in parameters. As for the policy network, the authors implemented a neural network featuring two hidden layers, 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. The R and Q in Eq. (3) were set to 0.1 and $\begin{bmatrix} 0.1 & 0 \\ 0 & 2 \end{bmatrix}$, respectively.

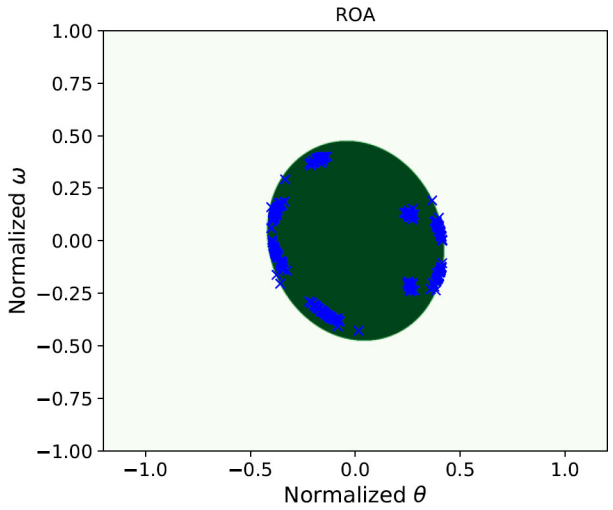

Fig. 4. The ROA under the the proposed MBRL based control policy. The dark green area denotes the safe region. The blue cross marks the data points the agent selected to explore the safe region.

The case study was conducted on an Intel Core i7-8650U @1.90GHz Windows based computer with 16GB RAM. The training convergence process of the developed safe MBRL based controller for the inverter system is illustrated in Fig. 3. The developed algorithm exhibited remarkable convergence,

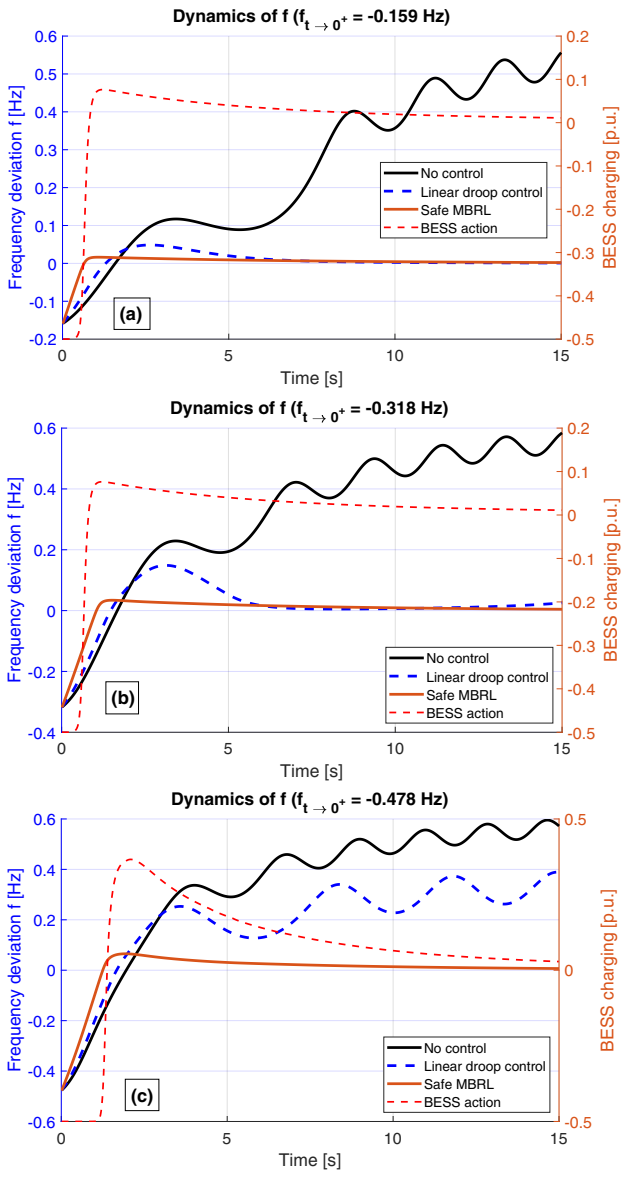

Fig. 5. The frequency control performances of the proposed safe MBRL and the comparing methods. The frequency deviation f is derived from the angular frequency deviation ω, with the relationship expressed as $f = \frac{\omega}{2\pi}$. Consequently, in (a), (b), and (c), the values of $\omega_{t\to 0+}$ are -1 rad/s, -2 rad/s, and -3 rad/s, respectively.

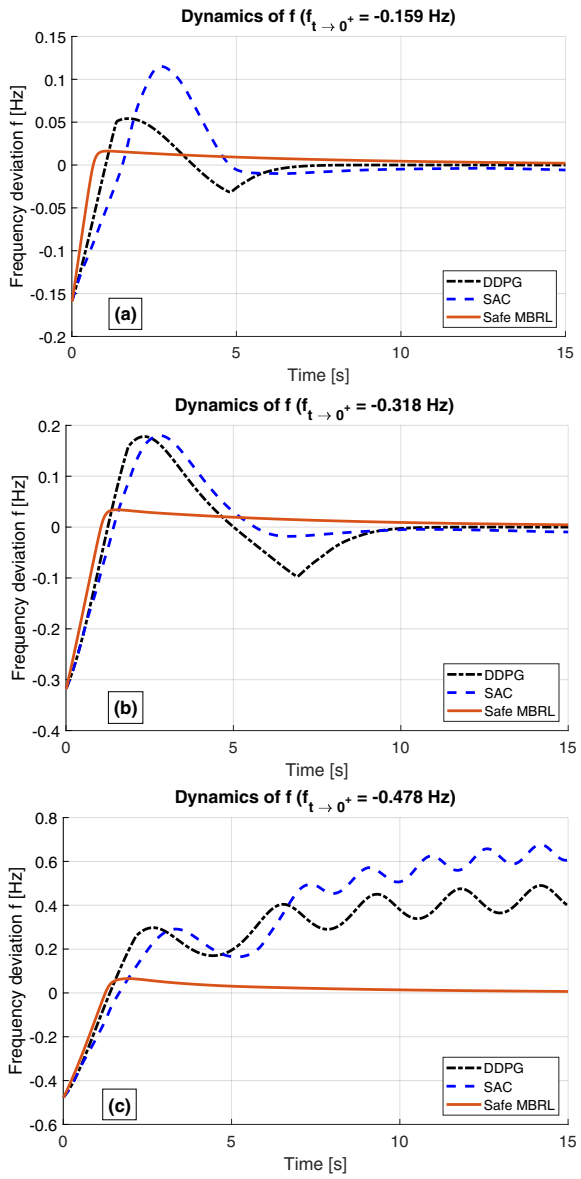

Fig. 6. The frequency control performances of the proposed safe MBRL and the model-free DRL methods. In (a), (b), and (c), the values of $\omega_{t\to 0+}$ are -1 rad/s, -2 rad/s, and -3 rad/s, respectively.

typically requiring only a few tens of iterations. Under the developed safe MBRL based control policy, the ROA is shown in Fig. 4. From the result, the ROA was determined based on the information from multiple measurements. We investigated the control performance of the safe MBRL controller, as depicted in Fig. 5 (a)-(c). From the figures, it is evident that when the inverter experiences frequency deviation, the safe MBRL controller efficiently restores the system to a stable state using BESS. 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 (refer to 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.

To demonstrate the superiority of the developed MBRL-based controller over traditional model-free deep reinforcement learning approaches (e.g., the Deep Deterministic Policy Gradient (DDPG) based method outlined in [14]), we conducted a comparative analysis of our method's performance against those employing DDPG and Soft Actor Critic (SAC) for frequency regulation. The results are depicted in Fig. 6. The analysis reveals that while the DDPG and SAC approaches achieve satisfactory control performance under relatively mild disturbances (see Fig. 6 (a)-(b)), managing to stabilize inverter frequency within several seconds post-disturbance, their effectiveness diminishes with increasing disturbance magnitude. In contrast, the MBRL-based frequency regulation technique not only restores inverter frequency more swiftly than the model-free DRL strategies in scenarios with relatively minor disturbances but also maintains robust frequency control under more significant disturbances (see Fig. 6 (c)). The stable control

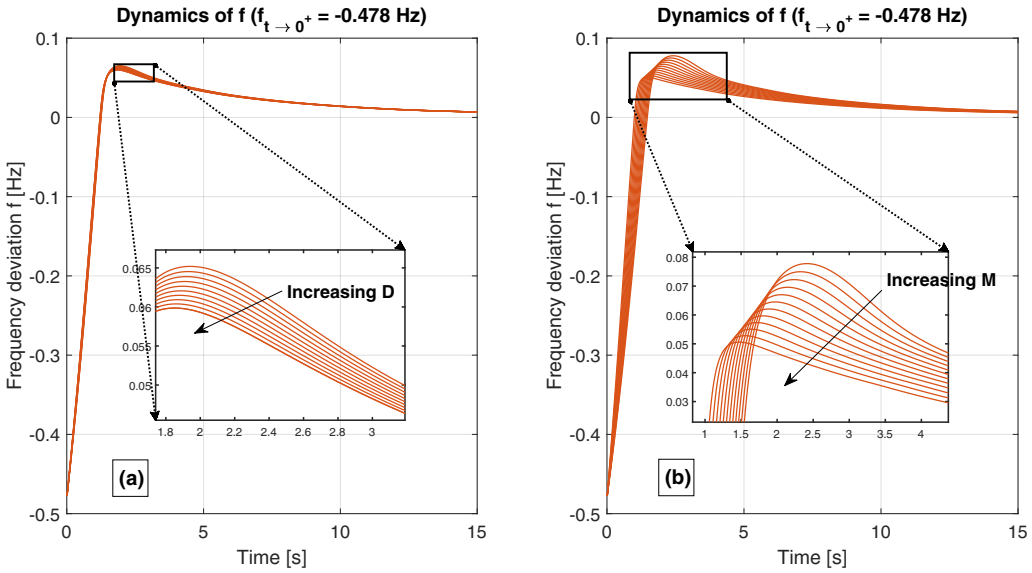

Fig. 7. The robustness of the safe MBRL policy against D and M uncertainties, respectively. In (a) and (b), the value of $\omega_{t\to 0+}$ is -3 rad/s.

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

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

V. Conclusion

In this letter, 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 reinforcement learning to enhance control performance. Additionally, a Gaussian process modeling was employed to capture VSG dynamics and enhance robustness to parameter uncertainty. The proposed approach offers a safe and robust controller for GFM inverter based frequency regulation. Simulation results demonstrated that the performance of the proposed safe MBRL algorithm surpasses that of traditional droop controller and model-free DRL based approaches. Moreover, the proposed MBRL based method requires only the measurements of the inverter's voltage phase and angular frequency, which are easily accessible in modern power systems. The algorithm's ease of implementation enhances its potential for practical applications.

Appendix I

The linear kernel is given by:

$$k_L(x, x') = x^T x' \tag{A1}$$

The Matérn kernel is given by:

$$k_M(x, x') = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l} d(x, x')\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{l} d(x, x')\right) \tag{A2}$$

Here, l represents a length-scale parameter, $d(\cdot, \cdot)$ denotes the Euclidean distance, $K_\nu(\cdot)$ is a modified Bessel function, and $\Gamma(\cdot)$ is the gamma function. The parameter ν regulates the smoothness of the function.

Appendix II

The initial LQR 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

$$\begin{bmatrix} \Delta\dot{\theta} \\ \Delta\dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{M}(B cos\theta - G sin\theta) & -\frac{D}{M} \end{bmatrix} \begin{bmatrix} \Delta\theta \\ \Delta\omega \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{M} \end{bmatrix} u \tag{A3}$$

The eigenvalues of the system are

$$\lambda_{1,2} = \frac{-D \pm \sqrt{D^2 - 4M \cdot (B cos\theta - G sin\theta)}}{2M} \tag{A4}$$

where, $Y_{ig} = G + jB$ 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 follows [21]

$$Y_{ig} = -\frac{1}{R_c + jX_c} = \frac{-R_c}{R_c^2 + X_c^2} + j\frac{X_c}{R_c^2 + X_c^2} \tag{A5}$$

It can be found that the eigenvalues depend on the operating point, virtual inertia and damping coefficients, and line parameters (i.e., R_c and X_c). In this work, $Y_{ig} = -0.495 + j4.95$. The per unit values of M and D are set to 5 and 1, respectively.

APPENDIX III

Lemma 1: *The control policy π_W is Lipschitz continuous with Lipschitz constant L_π.*

Proof. In this work, $\pi_W = \phi(x)$. $\phi(x)$ is the output of a K layer network, which is given by

$$\phi(x) = \phi_K(\phi_{K-1}(\cdots\phi_1(x; W_1); W_2)\cdots; W_K) \tag{A6}$$

In the hidden layers, ReLU activation functions are used. For the kth layer, there exits a constant $L_k > 0$ such that $\| \phi_k(x; W_k) - \phi_k(x + r; W_k) \| \leq L_k \| r \|$ holds for all x, r. The output layer utilizes tanh activation function, thus the network satisfies $\| \phi(x) - \phi(x + r) \| \leq L \| r \|$, with $L_\pi = \prod_{k=1}^{K} L_k$.

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

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

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

Proof. In this work, the Lyapunov function is set as the value function J 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], $v(f(x, \pi(x))) - v(x) < 0$ for all continuous states $x \in \Theta(c)$ with probability higher than $1 - \delta$. So, $\Theta(c)$ is a region of attraction for the system can be concluded based on Theorem 1.

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, 2014.

[2] D. Chen, K. Chen, Z. Li, T. Chu, R. Yao, F. Qiu, and K. Lin, "Powernet: Multi-agent deep reinforcement learning for scalable powergrid control," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1007–1017, 2022.

[3] Z. A. Obaid, L. M. Cipcigan, L. Abrahim, and M. T. Muhssin, "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, 2019.

[4] P. Verma, S. K., 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, 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, 2019.

[6] J. Liu, Y. Miura, H. Bevrani, and T. Ise, "Enhanced virtual synchronous generator control for parallel inverters in microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2268–2277, 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 *8th International Conference on Power Electronics - ECCE Asia*, 2011, pp. 1498–1505.

[8] P. He, Z. Li, H. Jin, C. Zhao, J. Fan, and X. Wu, "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, 2023.

[9] M. Li, W. Huang, N. Tai, L. Yang, D. Duan, and Z. Ma, "A dual-adaptivity inertia control strategy for virtual synchronous generator," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 594–604, 2019.

[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, 2014.

[11] F. Wang, L. Zhang, X. Feng, and H. Guo, "An adaptive control strategy for virtual synchronous generator," *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 5124–5133, 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, 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, 2019.

[14] Y. Li, W. Gao, S. Huang, R. Wang, W. Yan, V. Gevorgian, and D. W. Gao, "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, 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, 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, 2022.

[17] F. Berkenkamp, M. Turchetta, A. Schoellig, and A. Krause, "Safe model-based reinforcement learning with stability guarantees," *Advances in neural information processing systems*, vol. 30, 2017.

[18] H. Shuai, J. Fang, X. Ai, Y. Tang, J. Wen, and H. He, "Stochastic optimization of economic dispatch for microgrid based on approximate dynamic programming," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2440–2452, 2019.

[19] H. Shuai, J. Fang, X. Ai, J. Wen, and H. He, "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, 2019.

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

[21] V. Vittal, J. D. McCalley, P. M. Anderson, and A. Fouad, *Power System Control and Stability, 3rd Edition*. John Wiley & Sons, 2019.

[22] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Springer, 2006, vol. 1.

[23] F. Berkenkamp, R. Moriconi, A. P. Schoellig, and A. Krause, "Safe learning of regions of attraction for uncertain, nonlinear systems with gaussian processes," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 4661–4666.

[24] B. She, J. Liu, F. Qiu, H. Cui, N. Praisuwanna, J. Wang, L. M. Tolbert, and F. Li, "Systematic controller design for inverter-based microgrids with certified large-signal stability and domain of attraction," *IEEE Transactions on Smart Grid*, pp. 1–1, 2023.

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

[26] D. Duvenaud, "The kernel cookbook: Advice on covariance functions," *URL https://www. cs. toronto. edu/duvenaud/cookbook*, 2014.

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, RI, USA, from 2018 to 2019. He was a Postdoctoral Researcher with the URI and University of Tennessee, Knoxville (UTK) from 2019 to 2022. Currently, he is a Research Assistant Professor with the UTK, Knoxville, TN, USA. 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 University, China in 2017 and 2019, and the Ph.D. degree from the University of Tennessee, Knoxville in 2023, all in electrical engineering. He is currently a research engineer in Pacific Northwest National Laboratory (PNNL). He served as a student guest editor of IET-RPG. He was an outstanding reviewer of IEEE OAJPE (2020) and 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 engi-neering from the 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, TN, USA. His research interests include data mining, scientific computation, and power system simulation. He is the author of power system dispatch simulator, which is a key component of the CURENT Large-scale Testbed. He has been coordinating the LTB de-velopment efforts since 2021. He also built and maintains the list Popular Open Source Libraries for Power System Analysis.

Fangxing Li is also known as Fran Li. He 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, VA, USA, in 2001. He is currently the John W. Fisher Professor of electrical engineering and the Campus Director of CURENT with the University of Tennessee, Knoxville, TN, USA. His research interests include resilience, artificial intelligence in power, demand response, distributed generation and microgrid, and electricity markets. 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. Dr. Li 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.