

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

# Robust Stability Control of Droop-Controlled Microgrids under Worst-Case Parameter Uncertainty via Adversarial Deep Reinforcement Learning

Wenhao Chen, *Member, IEEE*, Yu Wang, *Member, IEEE* and Bikash Pal, *Fellow, IEEE*

**Abstract**—Microgrids based on droop-controlled inverters can achieve small-signal stability control by adjusting the droop coefficients. However, due to discrepancies between modeling parameters and actual system parameters in inverter-based microgrids, the droop settings determined from the model often fail to deliver the expected control performance in real-world applications. To address this critical issue, this paper proposes a droop coefficient tuning strategy based on adversarial reinforcement learning, which exhibits strong robustness against system parameter uncertainties. The uncertain parameter behavior and the robust tuning of droop coefficients are respectively modeled as a Markov Decision Process (MDP) and an Adversarial Markov Decision Process (AMDP). The twin delayed deep deterministic policy gradient (TD3) algorithm is employed to solve these processes, yielding the worst-case uncertainty policy and the corresponding robust droop control strategy. Finally, case studies are conducted on a microgrid system with four inverter-based resources (IBRs) and a modified IEEE 33-bus distribution system to validate the effectiveness and scalability of the proposed method.

**Index Terms**—Inverter-based microgrids, parameter uncertainty, robust stability, deep reinforcement learning, adversarial training.

## NOMENCLATURE

|                        |                                                                                                          |
|------------------------|----------------------------------------------------------------------------------------------------------|
| $\delta_i$             | Phase angle of inverter $i$                                                                              |
| $\omega_{com}$         | Common reference frequency                                                                               |
| $\omega_c$             | Cut-off frequency                                                                                        |
| $\omega_n, V_n$        | Nominal values of frequency and voltage                                                                  |
| $\omega_i, V_i$        | Frequency and voltage of inverter $i$                                                                    |
| $P_i, Q_i$             | Measured active and reactive power of inverter $i$                                                       |
| $k_{p,i}, k_{q,i}$     | Active and reactive droop coefficient of inverter $i$                                                    |
| $I_{d,i}, I_{q,i}$     | $d$ - and $q$ -axis current components of inverter $i$                                                   |
| $V_{d,i}, V_{q,i}$     | $d$ - and $q$ -axis voltage components of inverter $i$                                                   |
| $I_{D,i}, I_{Q,i}$     | $d$ -axis and $q$ -axis current components of the $i$ -th inverter in the common reference frame         |
| $V_{D,i}, V_{Q,i}$     | $d$ -axis and $q$ -axis voltage components of the $i$ -th inverter in the common reference frame         |
| $I_{D,ik}, I_{Q,ik}$   | $D$ - and $Q$ -axis current components of line between bus $i$ and bus $k$ in the common reference frame |
| $I_{D,L_i}, I_{Q,L_i}$ | $D$ - and $Q$ -axis current components of load $i$                                                       |
| $R_{ik}, L_{ik}$       | Resistance and inductance of the line between bus $i$ and bus $k$                                        |

|                                  |                                                     |
|----------------------------------|-----------------------------------------------------|
| $R_{L_i}, L_{L_i}$               | Resistance and inductance of load $i$               |
| $x_{sys}$                        | State vector of microgrid system                    |
| $A_{sys}$                        | State matrix of microgrid system                    |
| $\mathbf{q}$                     | Vector of uncertain parameters                      |
| $\mathcal{Q}$                    | Hyper-rectangular parameter space                   |
| $\varphi$                        | half-opening angle of damping-ratio conic region    |
| $\lambda_i$                      | The $i$ -th eigenvalues of state matrix             |
| $\alpha_i, \beta_i$              | Real and imaginary parts of the $i$ -th eigenvalue  |
| $\zeta_i$                        | Damping ratio of the $i$ -th mode                   |
| $\zeta_{min}$                    | Minimum damping ratio among all modes               |
| $\underline{\zeta}$              | Optimal minimum damping ratio                       |
| $\bar{\zeta}, \underline{\zeta}$ | Upper and lower bounds of the minimum damping ratio |
| $s_t, s'_t$                      | MDP and AMDP state at timestep $t$                  |
| $a_t, a'_t$                      | MDP and AMDP action at timestep $t$                 |
| $r_t, r'_t$                      | MDP and AMDP reward at timestep $t$                 |
| $\theta, \phi$                   | Parameter of the critic and actor network           |
| $\theta', \phi'$                 | Parameters of the target critic and actor networks  |
| $\tilde{\sigma}$                 | Standard deviation of the Gaussian function         |
| $\tau$                           | Soft update rate                                    |

## I. INTRODUCTION

THE global energy crisis and environmental pollution have driven the accelerated development of power-electronics-based distributed generation [1]. As a result, ensuring the stable and efficient operation of inverter-based microgrids has become one of the key research focuses. As a commonly used control strategy in inverter-based microgrids, droop control emulates the behavior of synchronous generators to not only ensure stable system operation but also achieve power sharing and voltage regulation [2], [3]. Stability boundaries for droop control coefficients have been identified through eigenvalue analysis methods [4]. Particle swarm optimization has been applied to improve droop coefficients settings, leading to a marked enhancement in the small-signal stability of microgrids [5]. An efficient event-triggered damping recovery strategy was proposed in [6], enabling dynamic restoration of system damping and balanced power sharing by optimizing droop coefficients. Reference [7] introduced an online tuning approach that updates droop coefficients in real time under

This work was supported by science and technology projects of State Grid Corporation of China. Project No.: 5400-202399376A-2-2-ZB. (Corresponding author: Yu Wang, email: [yu\\_wang@cqu.edu.cn](mailto:yu_wang@cqu.edu.cn))

Wenhao Chen and Yu Wang are with the School of Electrical Engineering, Chongqing University, Chongqing 400044, China.

Bikash Pal is with the Department of Electrical and Electronic Engineering, Imperial College London, London, 639798, UK.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

extreme disturbances, thereby improving the overall stability and reliability of the system. It is worth noting that all of the above studies are conducted based on accurately constructed small-signal models.

However, due to the presence of parameter uncertainties, discrepancies arise between the models used in research and the actual system, which significantly reduce the effectiveness of droop coefficients optimization for inverter-based microgrids and may even render it ineffective. Circuit parameter uncertainties and all external interactions were treated as a lumped disturbance, and a nonlinear disturbance observer was introduced to estimate and compensate for this disturbance, thereby enhancing the system's robustness [8]. In [9], an integrator was added to the Q-V droop control loop to accumulate the deviation between the output voltage and the reference voltage, which was then fed back to form a closed-loop droop structure, resulting in strong robustness against parameter drifts and mismatches. An  $H_\infty$  minimum-norm and  $\mu$ -synthesis-based controller was designed in [10], with pole placement performed using a polynomial method to maintain autonomous power sharing and voltage stability under variable droop gains and wide-range parameter perturbations. In [11], a parameter-dependent Lyapunov-function-based filter is proposed, which guarantees that the expected value of the trace of the state estimation error covariance remains within a bounded range under arbitrary parameter uncertainties. Although the above methods enhance the system's adaptability to parameter uncertainties from different perspectives, they still involve fundamental trade-offs among model structural errors, statistical assumptions, online computational complexity, and robustness in both dynamic and steady-state performance. As a result, their applicability and generalizability remain limited in large-scale and highly dynamic microgrid environments.

By interacting with the environment to obtain feedback and iteratively optimize the control policy, model-free optimization is gradually becoming a highly promising alternative. To address voltage and power coordination control in AC/DC interconnected microgrids, a hybrid method combining dual-droop control with data-driven model-free adaptive control was first proposed in [12]. In [13], an event-triggered, distributed model-free adaptive control strategy was developed to efficiently achieve control objectives with reduced communication overhead. A data-driven Volt/VAR curve optimization approach was presented to minimize system voltage deviations and reactive power losses in complex distribution networks [14]. An optimization-based control framework utilizing sampled data was established in [15] to guide inverter inputs for rapid dynamic voltage support through online iterative updates under fault disturbances. In [16], the authors proposed a data-driven droop control method based on Q-learning to achieve frequency synchronization and voltage recovery. Although the above control strategies eliminate the reliance on accurate system models, their performance may degrade when system parameters vary rapidly or when the training data fail to capture the underlying uncertainties.

In recent years, the outstanding performance of deep reinforcement learning (DRL) across various fields has led to its

increasing application in microgrids, including frequency control [17], energy management [18], and dynamic networking [19]. As an interaction-based intelligent decision-making approach, DRL employs deep neural networks (DNNs) to approximate the policy, value function, or system model. The agent explores the complex and high-dimensional state space through continuous trial-and-error to discover the optimal control strategy. To enhance the robustness of control policies in uncertain environments, researchers have developed adversarial learning mechanisms based on DRL. In [20], an adversarial training method was proposed for robust control in power systems, significantly enhancing the stability and generalization capability of the control policy when faced with unknown disturbances such as line outages. To address cybersecurity challenges in droop-controlled distributed inverter-based microgrids, an adversarial deep reinforcement learning framework was developed in [21], offering new insights into microgrid cybersecurity. A two-stage deep reinforcement learning approach was introduced in [22], where the offline agent undergoes adversarial training to gain robustness against transfer gaps, and is then deployed in the online phase, substantially improving the efficiency and safety of online exploration.

This work builds on our earlier conference paper [23] with substantial extensions and enhancements. The previous study examined only a least-effort destabilizing attack on droop gains using a single-agent DDPG framework applied to a linearized 3-IBR model. In contrast, the present study moves from an attacker-only setting to a defender-attacker formulation that explicitly searches for the worst-case uncertain-parametric behavior (WCUPB) while learning an adaptive, robust droop control policy that preserves stability. Furthermore, we replace the attack-oriented, conventional stability criterion used in [23] with a quantitative stability objective based on the minimum damping-ratio criterion, thereby improving dynamic stability under parametric uncertainty. Finally, we extend the single-agent DDPG framework in [23] to an adversarial TD3 framework and demonstrate the effectiveness and scalability of the proposed method through case studies on a 4-IBR system and a modified IEEE-33-bus network. Collectively, these upgrades transform [23] from an attack-identification prototype into a comprehensive robust-stability control framework.

Motivated by the above discussion, this paper introduces an adversarial deep reinforcement learning approach to enhance the small-signal damping margin and dynamic stability of microgrids under parametric uncertainty. To this end, we propose the concept of robust D-stability for microgrids, which is characterized by the system's minimum damping ratio. An analysis is conducted to evaluate the impact of uncertain parameters on the robust D-stability of microgrids under various control parameter combinations. The results preliminarily demonstrate the feasibility of tuning droop coefficients to preserve desirable robust D-stability in the presence of parametric uncertainties. Building on our previous work [23], the behavior of uncertain parameters is modeled as a MDP to introduce the worst-case uncertainty scenarios. Subsequently, adversarial training based on the AMDP is employed to search for robust control policies in uncertain environments. Finally, case studies on a four-IBR

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

microgrid and the modified IEEE 33-bus system corroborate the proposed method's effectiveness and its superiority over conventional robust-control baselines. Compared with existing studies on adversarial reinforcement learning in power systems [20]–[22], this paper makes the following distinct contributions:

1) This paper is the first to formulate the robust small-signal stability control of inverter-based microgrids under parametric uncertainty as an AMDP, in which the protagonist optimizes the droop coefficients while the adversary perturbs uncertain system parameters to emulate the worst-case dynamics.

2) This paper introduces a D-stability based minimum damping ratio criterion to quantitatively evaluate dynamic stability margins. This criterion endows the reward function with a clear physical interpretation and enables the control policy to directly optimize the system's small-signal stability performance.

3) Building upon the TD3 algorithm, we design an adversarial TD3 framework that trains the protagonist and adversary in continuous state–action spaces. The proposed method can successfully identify the worst-case uncertain-parametric behavior (WCUPB) and simultaneously learn an adaptive robust control policy.

## II. MICROGRID MODELING

This section first provides a brief restatement of the small-signal model of droop-controlled microgrids. Subsequently, considering parameter uncertainties, an uncertainty model of the microgrid is constructed, and the robust D-stability is introduced.

### A. Small-signal model of droop-controlled microgrid

A typical droop-controlled microgrid is shown in Fig. 1. Inverters and loads are connected to buses, which are further interconnected through distribution lines to form a simple and representative microgrid. This paper adopts a fifth-order electromagnetic model as the small-signal representation of the microgrid, which reported in many papers [24], [25] as a suitable model for the stability analysis of the droop-controlled microgrids.

The model includes the following states: the dynamic states of the inverters, namely the phase angle  $\delta_i$ , frequency  $\omega_i$ , and voltage  $V_i$ ; the dynamic states of the lines, namely the current components  $I_{D,ik}$  and  $I_{Q,ik}$  in the  $dq$  reference frame; and the dynamic states of the loads, namely the injected current components  $I_{D,i}$  and  $I_{Q,i}$  in the  $dq$  reference frame. The system dynamics can be expressed as follows:

$$\dot{\delta}_i(t) = \int (\omega_i - \omega_{com}) dt \quad (1)$$

$$\dot{\omega}_i = \omega_c (-\omega_i + \omega_n - k_{p,i} P_i) \quad (2)$$

$$\dot{V}_i = \omega_c (-V_i + V_n - k_{q,i} Q_i) \quad (3)$$

$$P_i = 1.5 (I_{d,i} V_{d,i} + I_{q,i} V_{q,i}) \quad (4)$$

$$Q_i = 1.5 (I_{d,i} V_{q,i} - I_{q,i} V_{d,i}) \quad (5)$$

$$L_{ik} \dot{I}_{D,ik} = -R_{ik} I_{D,ik} + \omega_{com} L_{ik} I_{Q,ik} + V_{D,i} - V_{D,k} \quad (6)$$

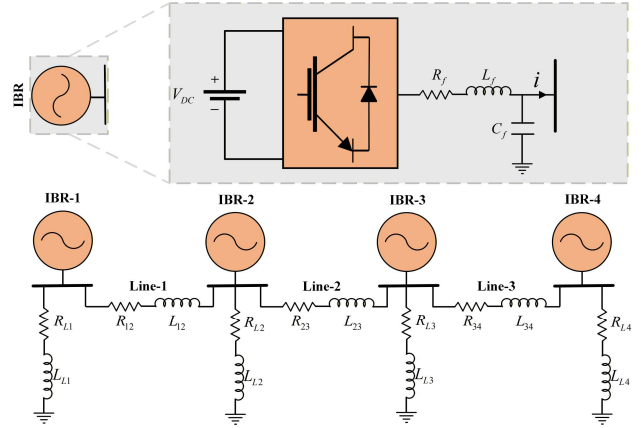


Fig. 1. A microgrid with 4 IBRs and power inverter structure of each IBR.

$$L_{ik} \dot{I}_{Q,ik} = -R_{ik} I_{Q,ik} - \omega_{com} L_{ik} I_{D,ik} + V_{Q,i} - V_{Q,k} \quad (7)$$

$$L_{Li} \dot{I}_{D,Li} = -R_{Li} I_{D,Li} + \omega_{com} L_{Li} I_{Q,Li} + V_{D,i} \quad (8)$$

$$L_{Li} \dot{I}_{Q,Li} = -R_{Li} I_{Q,Li} - \omega_{com} L_{Li} I_{D,Li} + V_{Q,i} \quad (9)$$

$$I_{D,i} = I_{D,ik} + I_{D,Li} \quad (10)$$

$$I_{Q,i} = I_{Q,ik} + I_{Q,Li} \quad (11)$$

By linearizing the above dynamic equations around the steady-state operating point, the small-signal model of the inverter-based microgrid can be derived. The steady-state operating point of the system can be obtained through power flow calculations. The resulting linearized model is expressed as:

$$\Delta \dot{\mathbf{x}}_{sys} = \mathbf{A}_{sys} \Delta \mathbf{x}_{sys} \quad (12)$$

Further details regarding this model can be found in [23], [25].

### B. Uncertainty Modeling and Robust D-stability

In practical power systems, parameter uncertainties can arise from various sources. Beyond uncertainties caused by manufacturing and measurement errors in device parameters, and environmental or temporal variations in network parameters, changes in controller parameters due to cyberattacks can also be regarded as a form of uncertainty. These parameter uncertainties lead to discrepancies between the employed model and the actual system, which ultimately affect the results of small-signal stability analysis by altering the system's state matrix  $\mathbf{A}_{sys}$ , as expressed by the following equation:

$$\Delta \dot{\mathbf{x}}_{sys} = (\mathbf{A}_{sys} + \Delta \mathbf{A}(\mathbf{q})) \Delta \mathbf{x}_{sys} \quad (13)$$

where  $\Delta \mathbf{A}(\mathbf{q})$  represents the variation in the state matrix  $\mathbf{A}_{sys}$  caused by the uncertainty vector  $\mathbf{q}$ . It is assumed that the uncertain parameter vector  $\mathbf{q}$  takes values within a hyper-rectangular parameter space  $\mathcal{Q}$ , that is,  $\mathbf{q} \in \mathcal{Q}$  where:

$$\mathcal{Q} = \left\{ \mathbf{q} \mid q_i \in [q_i^-, q_i^+], i = 1 \dots m \right\} \quad (14)$$

The system's stability is determined by the set of state matrices defined as  $\mathcal{A}(\mathcal{Q}) = \{ \mathbf{A} + \Delta \mathbf{A}(\mathbf{q}) \mid \mathbf{q} \in \mathcal{Q} \}$ . As the parameter vector  $\mathbf{q}$  varies continuously within the parameter space  $\mathcal{Q}$ , the eigenvalues of the state matrix also move continuously in the complex plane.

The ability of a system to maintain small-signal stability varies under different operating conditions. In some cases, even

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

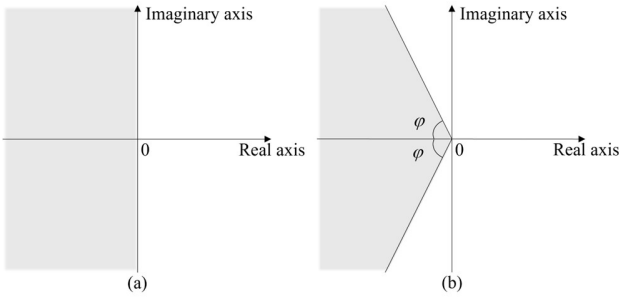


Fig. 2. (a) Conventional Stability Region. (b) D-stability Region.

slight changes in system parameters can lead to performance degradation or even a loss of stability, highlighting the need for strong robust stability in power systems. Robust stability refers to the capability of a control system to remain stable under a certain range of parametric uncertainties. For microgrid systems, it is essential to ensure stable operation under various uncertain parameters, making it necessary to account for robust stability when analyzing their small-signal stability.

Traditional stability criteria for microgrids are typically based on Lyapunov's method, which defines the imaginary axis as the stability boundary. As long as all eigenvalues of the system lie to the left of the imaginary axis, the system is considered small-signal stable, as illustrated in Fig. 2(a). However, in microgrid systems, to ensure rapid damping of oscillations, it is often necessary not only for the eigenvalues to lie in the stable region, but also to exhibit a certain minimum damping ratio. This imposes stricter requirements on the distribution of the system's eigenvalues. Therefore, the region D is defined as follows:

$$D = \{\lambda \in \mathbb{C} : |\arg(-\lambda)| \leq \varphi\}, \quad 0 < \varphi < \pi/2 \quad (15)$$

In other words, D is a sector symmetric with respect to the negative real axis, having an opening angle  $2\varphi$ , as shown in Fig. 2(b). Region D effectively introduces a minimum damping ratio cone in the left-half plane, ensuring a minimum damping margin for the oscillating mode. A system is considered D-stability if all of its eigenvalues  $\lambda_i$  lie within the region D. Compared to traditional stability, D-stability eliminates the weakly damped region near the imaginary axis in the complex plane, resulting in stronger dynamic stability for the system. Accordingly, the objective of this study is to ensure that the droop-controlled microgrid maintains D-stability under various parametric uncertainties—i.e., the system maintains robust D-stability.

### III. D-STABILITY ANALYSIS BASED ON PARAMETERS

This section first categorizes the various parameters of the microgrid, then highlights the importance of the minimum damping ratio for robust D-stability, and finally analyzes how different parameters affect the D-stability of the microgrid.

#### A. Parameter Classification and Damping Ratio

First, the parameter set of the microgrid is divided into two groups. The uncertainty parameter set  $B_{unc}$  includes uncertainty parameters whose variations can affect system stability.

The controllable parameter set  $B_{con}$  consists of controllable parameters that can be finely adjusted to stabilize the system. A system incorporating both uncertain and controllable parameters can be expressed as follows:

$$\Delta \dot{x}_{sys} = A_{act} \Delta x_{sys} \quad (16)$$

$$A_{act} = f(u_{nom} + u_{unc} + u_{con}) \quad (17)$$

where  $A_{act}$  is the actual system state matrix,  $u_{nom}$  denotes the nominal parameters,  $u_{unc}$  represents the uncertain parameters, and  $u_{con}$  represents the controllable parameters. The nominal, uncertain, and controllable parameters together determine the actual parameters of the system, and consequently, the state matrix  $A_{act}$ , which affects the system's robust D-stability.

Once the system state matrix  $A$  is obtained, the eigenvalues can be computed using the following equation:

$$|\lambda_i I - A| = 0 \quad (18)$$

where  $I$  is the identity matrix, and  $\lambda_i$  can be computed using the following expression:

$$\lambda_i = \alpha_i + j\beta_i \quad (19)$$

Each eigenvalue of the state matrix corresponds to a specific oscillation mode. When  $\alpha_i < 0$ , the mode is decaying, and its rate of decay can be characterized by the damping ratio  $\zeta_i$ :

$$\zeta_i = -\alpha_i / \sqrt{\alpha_i^2 + \beta_i^2} \quad (20)$$

According to Hurwitz stability, if the system is asymptotically stable, then the following condition holds:

$$\text{Re}(\lambda_i) < 0, \quad \forall i \quad (21)$$

that is, the system is small-signal stable when all real parts  $\alpha$  are negative. Combining equation (20) and (21), the system is small-signal stable if and only if all damping ratios satisfy  $\zeta_i \in (0, 1]$ . Moreover, a larger indicates a faster decay rate of the corresponding oscillation mode, meaning that the associated state response component approaches zero more quickly. Therefore, compared to the simple criterion of  $\text{Re}(\lambda_i) < 0$ , the damping ratio serves as a more comprehensive indicator of system stability, response speed, and oscillation severity. It not only determines whether the system is small-signal stable, but also reflects the combined characteristics of decay rate and oscillation frequency. Hence, this paper adopts the minimum damping ratio to characterize the robust D-stability of the system, as shown below:

$$0 < \underline{\zeta} \leq \zeta_{\min} \leq \bar{\zeta} < 1 \quad (22)$$

$$|\zeta_{\min} - \zeta^*| \leq o \quad (23)$$

The value  $\zeta^* = (\bar{\zeta} + \underline{\zeta})/2$  denotes the optimal minimum damping ratio, and  $o$  indicates a small positive margin approaching zero.  $\zeta_{\min}$  represents the weakest oscillation damping capability among all modes of the system. Increasing  $\zeta_{\min}$  is generally considered beneficial for enhancing the dynamic performance of the system's most weakly damped components. However, an excessively high damping ratio may lead to overly sluggish system responses. Although the system may exhibit almost no oscillations, the slow response may not meet certain fast dynamic performance requirements. Therefore, when  $\zeta_{\min}$  satisfies equation (22), the system is considered D-stability; when  $\zeta_{\min}$  satisfies equation (23), it indicates that the minimum

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

TABLE 1  
PARAMETERS OF A MICROGRID SYSTEM WITH 4 IBRS

| Parameters        | Values                |                       |                       |                       |
|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| IBR(No.)          | 1                     | 2                     | 3                     | 4                     |
| $k_{p,i}$ (rad/W) | $7 \times 10^{-5}$    | $7 \times 10^{-5}$    | $7 \times 10^{-5}$    | $7 \times 10^{-5}$    |
| $k_{q,i}$ (V/Var) | $1.12 \times 10^{-4}$ | $1.12 \times 10^{-4}$ | $1.12 \times 10^{-4}$ | $1.12 \times 10^{-4}$ |
| $P_i$ (kW)        | [0,20]                | [0,20]                | [0,20]                | [0,20]                |
| $Q_i$ (kVar)      | [-20,20]              | [-20,20]              | [-20,20]              | [-20,20]              |
| Line(No.)         | 1                     | 2                     | 3                     |                       |
| R ( $\Omega$ )    | 0.16                  | 0.32                  | 0.1                   |                       |
| L(mH)             | 0.5                   | 1                     | 0.3                   |                       |
| Load(No.)         | 1                     | 2                     | 3                     | 4                     |
| R ( $\Omega$ )    | 5.6                   | 14.1                  | 11.2                  | 8.3                   |
| L(mH)             | 8.9                   | 45                    | 17.8                  | 8.8                   |
| Reference         | $f^* = 50$ Hz         | $V^* = 311.1$ V       | $\tau = 27$ ms        |                       |

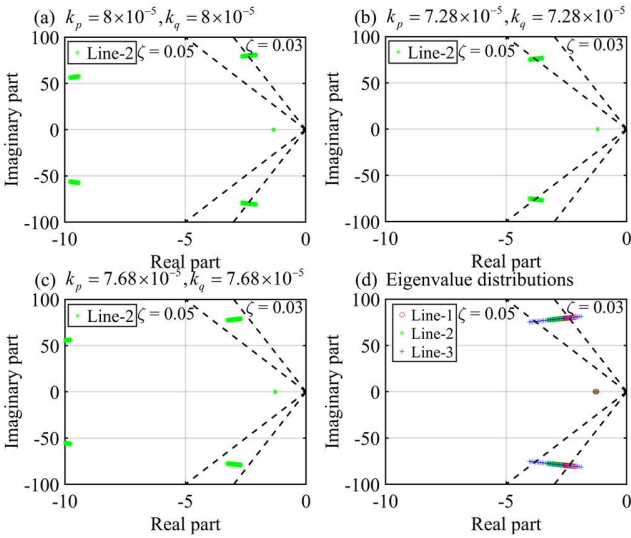


Fig. 3. The trace of system eigenvalues under varying line parameters with different droop coefficient combinations.

damping ratio lies near the optimal damping ratio, suggesting that the system possesses good robust D-stability.

#### B. Impact of Parameters on D-stability of Microgrid

As shown in equation (13), the uncertain parameters affect the distribution of the system's eigenvalues by influencing the state matrix  $A$ . This implies that under certain combinations of uncertain parameters, the minimum damping ratio of the system may fall outside the specified upper and lower bounds, that is:

$$\exists \mathbf{q}_i \in Q, \zeta_{\min}(\mathbf{q}_i) < \underline{\zeta} \text{ or } \zeta_{\min}(\mathbf{q}_i) > \bar{\zeta} \quad (24)$$

The following section uses a 4-IBR microgrid as an example to demonstrate how uncertain parameters affect the system's minimum damping ratio under different droop parameter combinations. The detailed parameters of the 4-IBR microgrid are provided in Table 1. The eigenvalue trajectories of the system's small-signal model under variations in line parameters are illustrated in Fig. 3. Here, the upper and lower bounds of the

minimum damping ratio are preset to 0.05 and 0.03. Figures 3(a)–(c) show how the system eigenvalues shift as the impedance of Line 2 varies within the range [0.8,1.2] under different droop coefficient settings. As observed from Figs. 3(a) and 3(b), if appropriate droop parameters are not selected, variations in the parameters of Line 2 alone can cause the system's minimum damping ratio to fall outside the specified bounds. As shown in Fig. 3(c), when the droop coefficients are set to  $k_p = 7.68 \times 10^{-5}$ ,  $k_q = 7.68 \times 10^{-5}$ , the system's minimum damping ratio remains within the predefined bounds regardless of the variations in Line 1 parameters. This demonstrates that properly tuning the droop parameters can enhance the system's robust D-stability. In addition, Fig. 3(d) shows that modes with the minimum damping ratio exhibit varying sensitivities to line parameters. As a result, when more uncertainty parameters are introduced, predicting and ensuring the minimum damping ratio becomes increasingly challenging.

#### IV. ROBUST CONTROL STRATEGY DESIGN UNDER PARAMETER UNCERTAINTY

In this section, we respectively provide a brief introduction to MDP and AMDP. Then, the robust optimization problem of controllable parameters, which considers system parameter uncertainties in inverter-based microgrids, is reformulated under the frameworks of MDP and AMDP.

##### A. MDP and WCUPB

During the training phase, the agent observes the state of the environment and selects corresponding actions based on the current state. It then receives a reward signal from the environment and updates its policy accordingly. The goal of the agent is to maximize the expected cumulative reward over time. To simplify analysis, it typically assumes the environment follows a Markov property—i.e., the next state depends only on the current state and action, independent of the past. Based on this assumption, the decision-making process in DRL can be modeled as an MDP. The MDP is defined as a tuple  $(S, A, P, R, \gamma)$ , which describes the interaction between the agent and the environment. Here,  $S$  denotes the state space, representing all possible system states, and  $A$  denotes the action space, representing all actions the agent can take. The state transition probability is defined by the function  $P: S \times A \times S \rightarrow [0,1]$ , which gives the probability of transitioning from state  $s_t \in S$  to state  $s_{t+1} \in S$  under action  $a_t \in A$ . The reward function  $R: S \times A \rightarrow \mathbb{R}$  represents the expected reward received when taking action  $a$  in state  $s$ . The discount factor  $\gamma \in [0,1]$  reflects the relative importance of future rewards compared to immediate rewards.

In inverter-based microgrids, WCUPB is defined as follows:

$$\max_{u_{unc} \in B_{unc}} \|\zeta_{\min}(A_{unc}) - \zeta^*\| \quad (25)$$

$$s.t. \zeta_{\min}(A_{unc}) < \underline{\zeta} \text{ or } \zeta_{\min}(A_{unc}) > \bar{\zeta} \quad (26)$$

$$u_{unc} \leq u_{unc} \leq \bar{u}_{unc} \quad (27)$$

where  $A_{unc} = f(u_{nom} + u_{unc})$  represents the state matrix of the system after undergoing the WCUPB.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

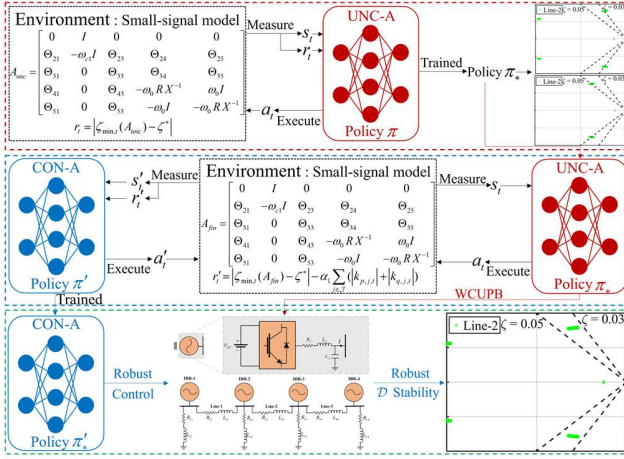


Fig. 4. Framework and Implementation Process of the Proposed Adversarial Training Scheme

The WCUPB problem is formulated as an MDP, with its tuple defined as follows:

1) *State Space*: In the WCUPB problem, the uncertain parameter agent (Unc-A) receives a deterministic state from the environment. At time step  $t$ , the observed state is represented as:

$$s_t = \{\zeta_{\min,t}(A_{unc}), \varepsilon_t\} \in S \quad (28)$$

where  $\varepsilon_t = \zeta_{\min,t}(A_{unc}) - \zeta^*$  represents the difference between the system's minimum damping ratio and the optimal minimum damping ratio.

2) *Action Space*: In this study, we select the deviation of line impedance parameters in the microgrid system as the actions of Unc-A. In inverter-based microgrids, line parameters are determined by cable type and length, which are subject to variations in practical applications due to factors such as temperature, material geometry, and manufacturing tolerances. Modeling disturbances in these parameters is of practical significance. Therefore, the action space of Unc-A is defined as:

$$a_t = \{\Delta L_{ik,t}, \Delta R_{ik,t}\} \in A, \quad (i, k) \in L \quad (29)$$

where  $L$  denotes the set of lines, and the ranges of  $\Delta R_{ik,t}$  and  $\Delta L_{ik,t}$  determine the strength of the uncertain parameter behavior and subsequent adversarial intensity.

The actual parameters of the line impedance are calculated as follows:

$$R_{ik,t} = R_{ik,t}^0 + \Delta R_{ik,t}, \quad \forall (i, k) \in L \quad (30)$$

$$L_{ik,t} = L_{ik,t}^0 + \Delta L_{ik,t}, \quad \forall (i, k) \in L \quad (31)$$

where  $R_{ik,t}$  and  $L_{ik,t}$  denote the actual resistance and inductance of line  $i$ , and  $R_{ik,t}^0$  and  $L_{ik,t}^0$  are their corresponding nominal values.

3) *State Transition Probability*: The state transition probability of the system is defined as follows:

$$s_{t+1} = P(s_t, a_t) \quad (32)$$

The next state of the system  $s_{t+1}$  is jointly determined by the current state  $s_t$  and the action  $a_t$  taken by the Unc-A.

4) *Reward Function*: The reward function is defined based on the objective of the Unc-A. Since the goal in this problem is

#### Algorithm 1 Adversarial Training based on TD3

- 1: Import trained uncertainty agent Unc-A with policy  $\pi_\phi$ .
- 2: Initialize agent Con-A's two critic networks and policy network with  $\theta'_1, \theta'_2$  and  $\phi'$ . Initialize target network with weight  $\bar{\theta}'_1 \leftarrow \theta'_1, \bar{\theta}'_2 \leftarrow \theta'_2, \bar{\phi}' \leftarrow \phi'$ .
- 3: Initialize replay buffer  $D$  and configure TD3 algorithm according to Table II.
- 4: **For** episode=1 to  $M$  **do**
- 5: Initialize state  $s'_1$ .
- 6: **For**  $t=1$  to  $T$  **do**
- 7: Obtain the current state  $s'_t$ . Unc-A generates action  $a_t$  based on  $\pi_\phi$ . Con-A generates a noisy action  $a'_t$  based on the policy  $\pi_{\phi'}$  then apply clipping to action  $a'_t$ .
- 8: Execute actions  $a_t$  and  $a'_t$ , compute the reward  $r'_t$ , and observe the next state  $s'_{t+1}$ .
- 9: Store the transition  $(s'_t, a_t, a'_t, r'_t, s'_{t+1})$  into the replay buffer  $D$ .
- 10: **If** it is the update step, **then** randomly sample a mini-batch of transitions  $(s'_k, a_k, a'_k, r'_k, s'_{k+1})$  from the replay buffer  $D$ . Compute the target action  $\bar{a}'_t$  and the target  $\bar{Q}$ -value. Update the critic networks' weights  $\theta'_1, \theta'_2$ .
- 11: **If**  $t \bmod N_{delay} = 0$  **then** update the actor networks' weights.
- 12: Soft update the target network parameters.
- 13: **end for**
- 14: **end for**

to identify the worst-case scenario for the minimum damping ratio, the reward function is given as:

$$r_t = |\zeta_{\min,t}(A_{unc}) - \zeta^*| \quad (33)$$

#### B. AMDP and robust control

To ensure that the control policy of the controllable parameter agent (Con-A) remains robust under system parameter uncertainties, the optimization problem is defined as follows:

$$\min_{u_{con} \in B_{con}} \|\zeta_{\min}(A_{fin}) - \zeta^*\|_1 + |u_{con}| \quad (34)$$

$$s.t. \underline{\zeta} < \zeta_{\min}(A_{fin}) < \bar{\zeta} \quad (35)$$

$$\underline{u}_{con} \leq u_{con} \leq \bar{u}_{con} \quad (36)$$

The objective of this optimization problem is to ensure that, under the WCUPB imposed by Unc-A, the control policy of agent Con-A keeps the system's minimum damping ratio as close as possible to the desired optimal value within the predefined bounds.

AMDP is also composed of a tuple  $(S', A, A', P', R', \gamma')$ . However, unlike MDP, AMDP involves two agents: Unc-A and Con-A, who have opposing goals. Each agent takes different actions to achieve their own goals. Therefore, the definition of the tuple in AMDP differs accordingly and is given as follows:

1) *State space*: Similar to MDP, the state information of interest in AMDP is still the system's minimum damping ratio and its deviation from the predefined optimal minimum damping ratio. The only difference is that in this process, the

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

minimum damping ratio is calculated based on the state matrix  $A_{fin}$ , rather than  $A_{unc}$ .

TABLE 2  
HYPERPARAMETERS OF TD3 ALGORITHM

| Parameters                             | Values             |
|----------------------------------------|--------------------|
| Experience buffer length               | $1 \times 10^6$    |
| Discount factor                        | 0.99               |
| Minibatch size                         | 256                |
| Critic /Actor learning rate            | $1 \times 10^{-3}$ |
| Optimizer                              | Adam               |
| Target smooth noise standard deviation | 0.2                |
| Policy updating frequency              | 2                  |
| Target smooth factor                   | $5 \times 10^{-3}$ |
| Exploration noise standard deviation   | 0.1                |
| Target-policy noise clipping bound     | [-0.5,0.5]         |
| Reward-function weight                 | 0.2                |

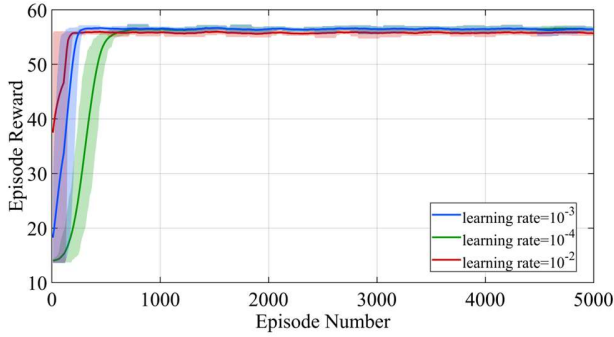


Fig. 5. The average reward and episode reward during the training stage of the Unc-A.

$$s_t' = \{\zeta_{\min,t}(A_{fin}), \varepsilon_t'\} \in S' \quad (37)$$

where  $\varepsilon_t' = \zeta_{\min,t}(A_{fin}) - \zeta^*$

2) *Action space*: For the Con-A, its action space is defined as  $A'$ , which includes all parameters in the controllable parameter set  $B_{con}$ . In this paper, based on the inverter-based microgrid model, these controllable parameters refer specifically to the droop coefficients of the inverters. Therefore, the action space is defined as:

$$a_t' = \{k_{p,j,t}, k_{q,j,t}\} \in A', \quad j \in J \quad (38)$$

where  $k_{p,j,t}$  and  $k_{q,j,t}$  respectively represent the active and reactive power droop coefficients of inverter  $j$ .

The action space of Unc-A remains the same as the action space  $A$  in the MDP.

3) *State Transition Probability*: In the AMDP framework, the state transition probability of the system is given by:

$$s_{t+1}' = P(s_t', a_t', a_t) \quad (39)$$

the next state  $s_{t+1}'$  of the system is jointly determined by the current state  $s_t'$  and the actions  $a_t'$ ,  $a_t$  taken by the two agents.

4) *Reward function*: Con-A not only restore the system damping to the optimal range under the worst-case environment created by Unc-A, but also minimize the control cost as much as possible. Therefore, the reward can be defined as:

$$r_t' = R'(s_t', a_t', a_t, s_{t+1}') = \alpha_1 r_{1,t}' + r_{2,t}' \quad (40)$$

$$r_{1,t}' = -\sum_{j \in J} (|k_{p,j,t}| + |k_{q,j,t}|) \quad (41)$$

$$r_{2,t}' = -|\zeta_{\min,t}(A_{fin}) - \zeta^*| \quad (42)$$

The reward function consists of two components,  $r_{1,t}'$  and  $r_{2,t}'$ . The first component, as shown in equation (41), penalizes excessive variations of droop coefficients to minimize the overall control cost. The second component, as shown in equation (42), encourages the improvement of the system's minimum damping ratio toward the desired target, thereby enhancing the damping margin and dynamic stability. The hyperparameter  $\alpha_1$  serves as the weighting factor for  $r_{1,t}'$ , used to adjust its contribution to the total reward, indicating the degree of emphasis placed on it.

### C. Solution Approach for MDP and AMDP in WCUPB and robust control Design

Given the non-convexity of damping ratio calculations and the complexity of adversarial problems, this study adopts the TD3 algorithm to train DNNs. A brief overview of the TD3 algorithm is provided below.

TD3 is a DRL algorithm for continuous action spaces. It is built upon the DDPG framework and incorporates the idea of Double Q-learning to alleviate the overestimation issue in action-value function approximation [26]. TD3 maintains two independently initialized Q-networks,  $Q_{\theta_1}(s, a)$  and  $Q_{\theta_2}(s, a)$ , and uses the smaller of the two Q-values when updating the target to suppress overestimation bias:

$$y_t = r_t + \gamma \min[Q_{\theta_1}(s_{t+1}, \tilde{a}_{t+1}), Q_{\theta_2}(s_{t+1}, \tilde{a}_{t+1})] \quad (43)$$

where  $\tilde{a}_{t+1}$  is a target action with added noise, generated by the target policy network and smoothed by noise:

$$\tilde{a}_{t+1} = \pi_{\theta_{t+1}}(s_{t+1}) + \text{noise} \sim \text{clip}(N(0, \sigma), -c, c) \quad (44)$$

In continuous action spaces, Q-values for similar actions in the same state tend to be close. Therefore, Gaussian noise  $\delta$  is added to the action  $\tilde{a}_{t+1}$  to smooth the updates of the critic network. For the stability of training the critic network, the noise is clipped by a constant  $c$ , keeping  $\tilde{a}_{t+1}$  close to the output of the policy network.

The total loss function of the critic is composed of the sum of the mean squared error (MSE) loss functions of the two critic networks:

$$\text{loss} = \text{MSE}(Q_{\theta_1}(s_t, a_t), y_t) + \text{MSE}(Q_{\theta_2}(s_t, a_t), y_t) \quad (45)$$

where  $\theta$  are updated using the gradient descent algorithm.

The policy network is updated by maximizing the Q-value:

$$\phi_{t+1} \leftarrow \phi_t + \alpha_2 \cdot N^{-1} \sum \nabla_a Q_{\theta_1}(s_t, a_t) \Big|_{a_t = \pi_{\phi_t}(s_t)} \nabla_{\phi_t} \pi_{\phi_t}(s_t) \quad (46)$$

where  $\alpha_2$  denotes the learning rate of the policy network, and the action  $a_t$  in the critic network  $Q_{\theta_1}(s_t, a_t)$  is obtained from  $a_t = \pi_{\phi_t}(s_t)$ . To improve training stability, TD3 introduces a delayed policy update mechanism, meaning that for each update of the Q-networks, the policy network is updated only once. In this study, the soft update of the target networks is performed synchronously with the delayed actor updates, i.e., every  $N_{\text{delay}}$  steps, to ensure training stability. The soft update mechanism is mathematically described as:

$$\theta' \leftarrow \tau \theta + (1 - \tau) \theta' \quad (47)$$

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

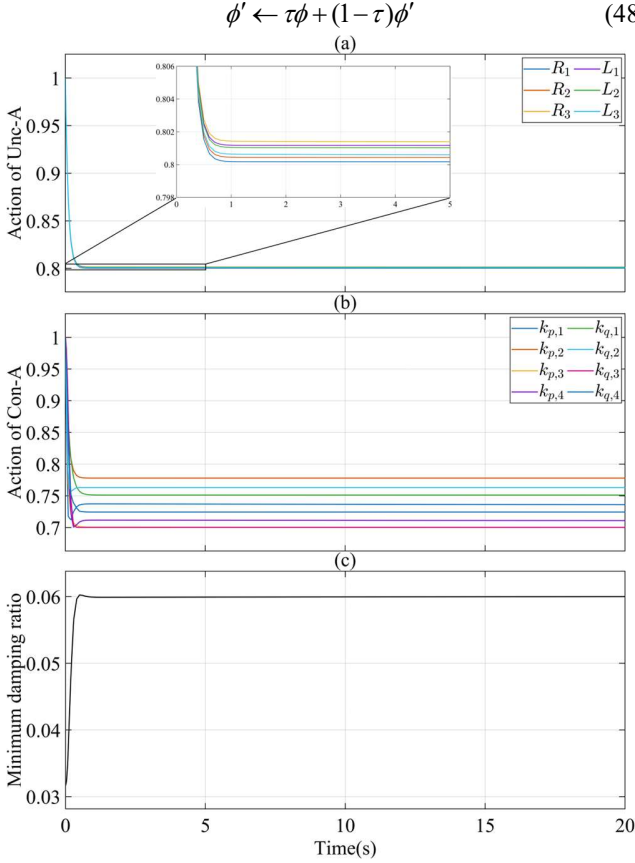


Fig. 6. (a) Variation of Unc-A's action. (b) Variation of Con-A's action. (c) Variation of the Minimum Damping Ratio of the 4-IBR microgrid under the actions of Unc-A and Con-A.

In the standard AMDP framework, Con-A maximizes the objective  $J(\pi', \pi)$  through its policy  $\pi'$ , while Unc-A minimizes the same objective through its policy  $\pi$ . Together, they form a min-max adversarial optimization problem, as shown below:

$$\pi_*(a_t | s_t) = \arg \max_{\pi} \min_{\pi'} J(\pi', \pi) \quad (49)$$

$$\pi_*(a_t | s_t) = \arg \min_{\pi} \max_{\pi'} J(\pi', \pi) \quad (50)$$

Although adversarial reinforcement learning methods based on jointly training the policies  $\pi'$  and  $\pi$  are theoretically effective, in practice the training process may suffer from issues such as low sample efficiency, unstable convergence, and high computational cost. Therefore, in this study, we first train agent Unc-A under the MDP framework. The well-trained agent Unc-A is then deployed in the AMDP framework, where its policy  $\pi$  is fixed during the training of agent Con-A's policy  $\pi'$ . The adversarial training framework is illustrated in Fig. 4, and the specific implementation of the TD3 algorithm within this framework is detailed in Algorithm 1.

## V. RESULTS

### A. Experimental Setup

As described in Section IV-C, the TD3 algorithm is employed in this study to train the Unc-A and Con-A for solving

the MDP and AMDP. All DNNs in this study are initialized with random weights and biases. The framework includes one

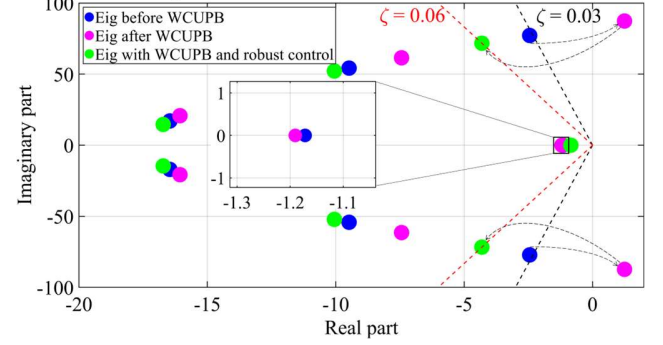


Fig. 7. Eigenvalue distributions of the 4-IBR microgrid system before WCUPB, after WCUPB, and with both WCUPB and robust control.

actor network and two independently initialized critic networks with identical structures. Each critic network takes the system state and action as inputs, which are concatenated at the first hidden layer. The critic networks contain two hidden layers with 100 and 50 neurons, respectively, both using ReLU activation functions. The output layer is a single neuron representing the Q-value of the state-action pair. The actor network maps the environment state to continuous actions through two hidden layers, which are configured identically to those in the critic networks. A Tanh activation function is applied at the output to generate actions within the range  $[-1, 1]$ , which are then linearly scaled back to the actual action range. A scaling layer at the output of the actor is used to ensure that the actions conform to the defined bounds of the action space.

The hyperparameters of the TD3 algorithm are summarized in Table 2. The parameter settings were primarily based on the typical values reported in [21], [26] and were further fine-tuned through experimental validation. Taking the learning rate as an example, Fig. 5 illustrates the episode rewards and their 200-step moving averages under different learning-rate configurations during training. As shown in the figure, an excessively large learning rate leads to oscillatory convergence and lower reward performance, whereas an overly small learning rate results in delayed convergence. In contrast, a learning rate of  $1 \times 10^{-3}$  achieves a favorable balance between convergence speed and stability.

### B. Case1: Worst-Case Uncertain-Parametric Behavior

To evaluate the performance of the proposed method, we first conduct tests on a 4-IBR microgrid, whose topology is shown in Fig. 1. Given that the line parameters in microgrid systems are subject to measurement errors, temperature variations, and model simplifications, this paper selects line parameters as representative uncertain variables for the study. Specifically, lines 1, 2, and 3 are all considered to contain uncertainties, forming the uncertain parameter set  $B_{unc} = \{R_{12}, L_{12}, R_{23}, L_{23}, R_{34}, L_{34}\}$ . The range of each uncertain parameter in this study is set to  $[0.8, 1.2]$  times its nominal value. This configuration not only covers the typical engineering error range of  $\pm 10\%$ , but also allows the adversarial reinforcement learning to train robust policies within a broader space of modeling errors, thereby

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

enabling adaptation to a wider variety of real-world scenario variations.

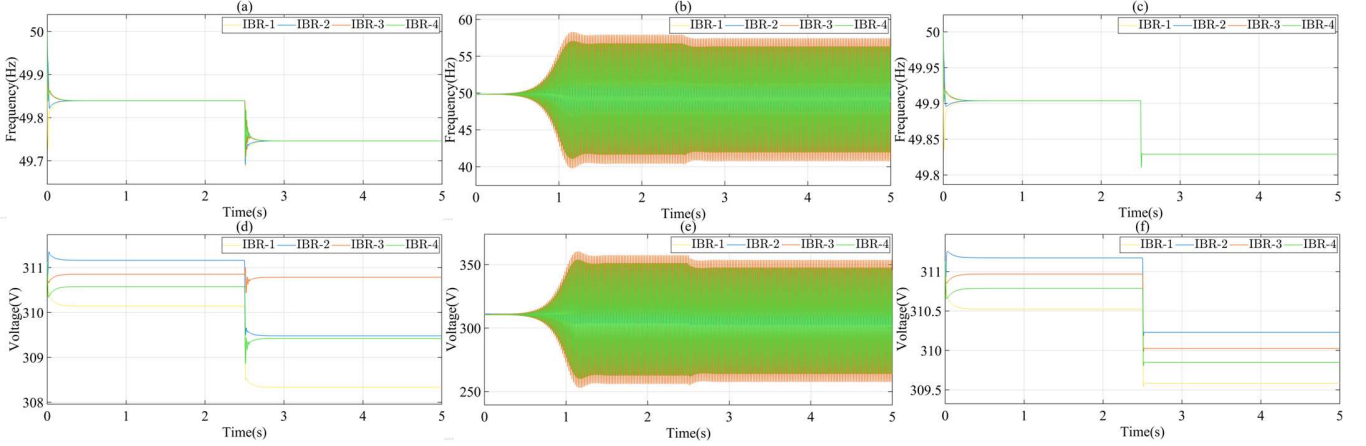


Fig. 8. Frequency and voltage trajectories of the 4-IBR microgrid system under the same disturbance: before WCUPB, after WCUPB, and with both WCUPB and robust control.

Fig. 6(a) illustrates how Unc-A's actions evolve over time, and the close alignment in their trends, rates, and final values suggests that the line parameters influence the eigenvalue associated with the system's minimum damping ratio in the same manner. As shown in Fig. 7, after the system is subjected to the WCUPB generated by the Unc-A, the minimum damping ratio decreases from 0.0317 to  $-0.0142$ . Furthermore, Figs. 8(b) and (e) illustrate the impact of the WCUPB on the system by showing the trajectories of both system frequency and voltage. These results indicate that the system can operate stably under the initial parameters; however, when subjected to a certain level of uncertain parameter disturbances, the system become unstable. It is worth noting that, because the DRL is a sampling-based data-driven search, the discovered WCUPB corresponds to the worst-case behavior within the explored uncertainty domain, rather than a mathematically global extremum. Nonetheless, the above results and analysis show that the identified scenarios by the DRL consistently drive the system toward its stability margin, effectively revealing the most critical combinations of uncertainty parameters. Therefore, the WCUPB obtained in this work should be interpreted as a practically realizable and policy-dependent worst case, providing a reliable basis for subsequent adversarial training.

### C. Case2: Robust Control Strategy

Building upon Case 1, Case 2 introduces the Con-A. Assuming that all droop coefficients are adjustable, the controllable parameter set is defined as  $B_{con} = \{k_{p,i}, k_{q,i} | i = 1, \dots, 4\}$  and the value range of each droop coefficient is uniformly set to  $[0.5, 1.5]$  times its initial value.

The Con-A is deployed alongside the previously trained Unc-A in the same environment for adversarial training. Similarly, after the reward fully converges, the neural network model of Con-A is extracted and deployed, allowing the system's control strategy to exhibit improved robustness under uncertainty. Fig. 6(b) illustrates how Con-A enhances the system's robust D-stability by adjusting the droop coefficients when responding to the WCUPB. In Fig. 6(b), the trajectories, rates, and final values of the droop parameters exhibit slight differences, yet

their overall trends remain generally consistent. Meanwhile, Fig. 6(c) clearly demonstrates a notable increase in the system's minimum damping ratio. This implies that although the sensitivities of individual droop parameters to this eigenvalue differ, their overall influence is greater than that of the line parameters. Ultimately, with the regulation of Con-A, the system's minimum damping ratio is improved to 0.06, which corresponds to the pre-defined target value.

Fig. 7 illustrates the changes in the eigenvalue distribution of the 4-IBR microgrid system before and after being subjected to the WCUPB generated by Unc-A, as well as after Con-A participates in the control. As shown in Fig. 7, although the slowest real mode shifts from  $-1.19$  to  $-0.844$  after robust control under parameter uncertainty, the damping of the oscillatory modes is improved overall. Especially, the system's minimum damping ratio is increased to 0.06, representing an 89.3% improvement compared with the original system. This not only signifies a substantial enhancement in the system's robust D-stability but also markedly strengthens its capability to attenuate disturbances. Taking the frequency as an example, compared with the original system under the same disturbance, implementing robust control reduces the steady-state frequency deviation from 0.094 to 0.074, shortens the stabilization time from 0.36 s to 0.08 s, and lowers the maximum frequency deviation from 0.155 to 0.094, as shown in Figs. 8(a) and 8(c). In addition, the improvement in voltage dynamics can also be observed in Figs. 8(d) and 8(f). Therefore, the proposed method significantly improves the 4-IBR microgrid's robust D-stability performance.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

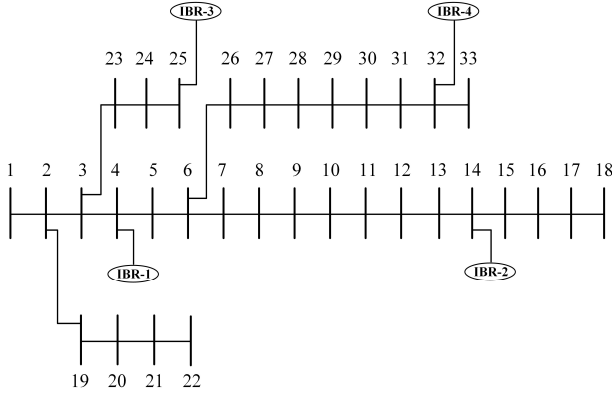


Fig. 9. The IEEE 33-bus system used for scalability test.

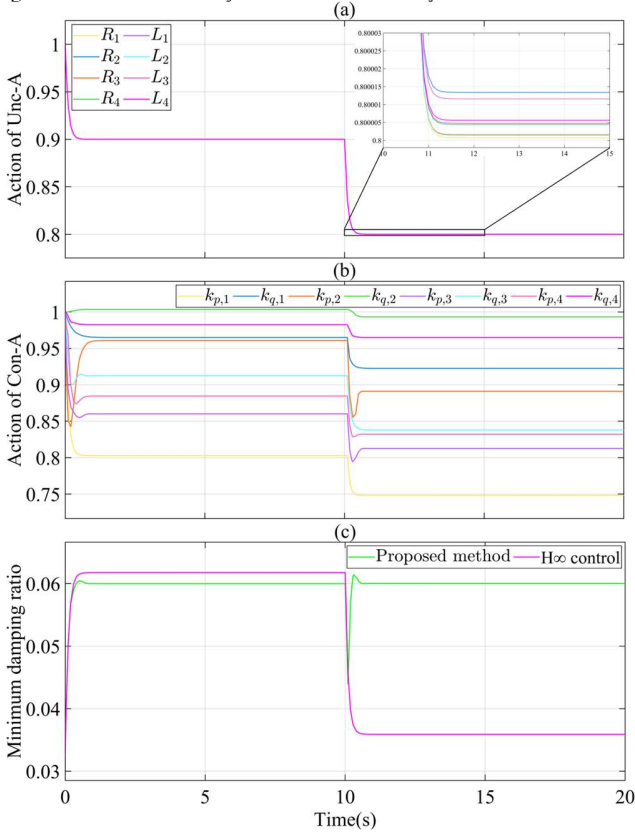


Fig. 10. (a) Variation of Unc-A's action. (b) Variation of Con-A's action. (c) Variation of the Minimum Damping Ratio of the modified IEEE 33-bus system under the proposed method and  $H_\infty$  control.

#### D. Case3: Scalability Test Based on the IEEE 33-bus System

In Case 3, the scalability of the proposed method is validated on the modified IEEE 33-bus system, whose topology is illustrated in Fig. 9. Although the system scale increases significantly, the dimensions of the state and action vectors used as inputs remain nearly the same as those of the previous system. The primary difference lies in the slightly higher action dimensionality of Unc-A. The training cost of the proposed method mainly stems from the two Q-networks and the actor network, with computational complexity on the order of  $O(n^2) \sim O(n^3)$ , where  $n$  denotes the number of network nodes. In the 4-IBR system, training of Unc-A takes approximately 0.8 hour, while

adversarial training of the Con-A requires about 1.1 hours. For the modified IEEE 33-bus system, Unc-A requires nearly 1.1 hours for training, and the Con-A takes roughly 1.3 hours. Overall, despite the increase in system scale, the training process can still be completed within a reasonable time frame and the online inference of the trained policy remains lightweight.

To demonstrate the adaptive robust control performance of Con-A under dynamically varying parameter uncertainty, the simulation assumes an initial uncertainty range of  $[0.9, 1.1]$  times the nominal value. Subsequently, at 10 seconds into the simulation, it is assumed that the factors contributing to parameter uncertainty intensify, thereby expanding the range to  $[0.8, 1.2]$  times the nominal value. The network impedance

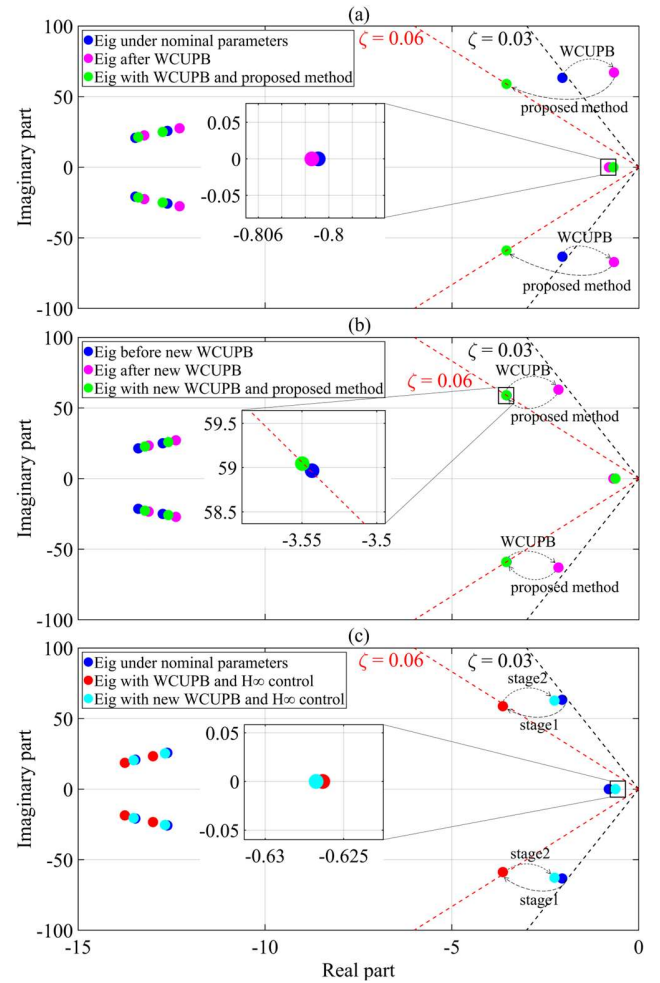


Fig. 11. (a) Eigenvalue distributions before 10s under the proposed method. (b) Eigenvalue distributions after 10s under the proposed method. (c) Eigenvalue distributions under  $H_\infty$  control with the same conditions.

parameters and load data in the system are derived from the standard 33-bus test case in MATPOWER. To facilitate computation and statistical analysis, the 32 distribution lines are grouped accordingly. Considering that the IEEE 33-bus system is a typical radial distribution network, the grouping is carried out based on its topological characteristics. Specifically, the system is divided into four electrical subtrees, in which the lines are topologically adjacent and have similar electrical distances.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

Therefore, it can be approximately considered that the parameter uncertainties of the lines within the same subtree are consistent. The detailed grouping is as follows: Group 1 includes lines 1–2, 2–3, 3–23, 2–19, 19–20, 20–21, 21–22, 23–24, and 24–25; Group 2 includes lines 3–4, 4–5, 5–6, 6–7, 7–8, 8–9, 9–10, and 10–11; Group 3 includes lines 11–12, 12–13, 13–14, 14–15, 15–16, 16–17, and 17–18; Group 4 includes lines 6–26, 26–27, 27–28, 28–29, 29–30, 30–31, 31–32, and 32–33.

Fig. 10(a)-(b) illustrate the action of Unc-A and Con-A. Additionally, Fig. 10(c) shows the evolution of the system's minimum damping ratio under the same WCUPB conditions for both the method proposed herein and a classical  $H_\infty$ -based robust control scheme. As the  $H_\infty$  approach is a mature and well-established framework, we do not elaborate on it further; interested readers are referred to [10], [27]. Since the eigenvalue corresponding to the system's minimum damping ratio is more sensitive to droop parameters than to line parameters, and the adjustable range of droop parameters is also larger, the system can theoretically achieve better small-signal damping margin and dynamic performance. As shown in Fig. 10, during the simulation, when the action range of Unc-A expands and introduces a new WCUPB to the system, Con-A responds to the observed change in system states by initiating a new round of droop coefficient adjustments to maintain the minimum damping ratio at 0.06. By contrast, while the  $H_\infty$ -based robust controller can deliver a prescribed robustness margin under a priori specified parameter uncertainties, its strategy of enveloping all uncertainties to guarantee robustness inevitably yields a conservative design. Moreover, as an ex-ante defense, once the actual uncertainties exceed the pre-assumed bounds, the system's robust stability deteriorates markedly—and may even collapse into instability—as illustrated in Fig. 10(c). Finally, Fig. 11 reports the distribution of the system's critical eigenvalues under various conditions, in agreement with the time-domain results shown in Fig. 10(c).

## VI. CONCLUSION

This paper developed an adversarial DRL framework for robust stability control of droop-controlled microgrids under worst-case parameter uncertainty. The motivation stems from the need to maintain adequate damping stability despite inevitable model mismatches and disturbances. By formulating the problem as an AMDP, the proposed approach jointly identifies the WCUPB and derives an adaptive robust droop control policy. Case studies on both a 4-IBR system and the modified IEEE 33-bus system demonstrate the method's effectiveness and scalability. The main advantage of this study lies in its ability to integrate quantitative stability analysis with data-driven learning, enabling closed-loop optimization of dynamic stability performance under uncertainty. However, the current framework is based on a linearized small-signal model and relies on offline training, which may limit its applicability to strongly nonlinear or time-varying scenarios. Moreover, the adversarial agent focuses on parameter perturbations only, and other sources of uncertainty—such as topology variation, communication delay, and measurement noise—are not yet incorporated.

Future work will focus on extending the framework toward nonlinear and time-varying microgrids, develop richer adversarial strategies encompassing cyber-physical uncertainties. Further research will also explore distributed multi-agent learning architectures for large-scale interconnected microgrids to further enhance adaptability and computational efficiency in practical applications.

## REFERENCES

- [1] M. J. B. Kabeyi and O. A. Olanrewaju, "Sustainable energy transition for renewable and low carbon grid electricity generation and supply," *Front. Energy Res.*, vol. 9, Art. no. 743114, 2022.
- [2] J. Rocabert, A. Luna, F. Blaabjerg and P. Rodríguez, "Control of Power Converters in AC Microgrids," *IEEE Transactions on Power Electronics*, vol. 27, no. 11, pp. 4734-4749, Nov. 2012.
- [3] D. E. Olivares, *et al.*, "Trends in Microgrid Control," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1905-1919, July 2014.
- [4] A. Kahrobaei and Y. A. -R. I. Mohamed, "Stability analysis and control of medium-voltage micro-grids with dynamic loads," 2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, Canada, 2013, pp. 1-5.
- [5] M. A. Hassan and M. A. Abido, "Optimal Design of Microgrids in Autonomous and Grid-Connected Modes Using Particle Swarm Optimization," *IEEE Transactions on Power Electronics*, vol. 26, no. 3, pp. 755-769, March 2011.
- [6] G. Raman, J. C. -H. Peng and H. H. Zeineldin, "Optimal Damping Recovery Scheme for Droop-Controlled Inverter-Based Microgrids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2805-2815, July 2020.
- [7] S. Chakraborty, S. Patel, G. Saraswat, A. Maqsood and M. V. Salapaka, "Seamless Transition of Critical Infrastructures Using Droop-Controlled Grid-Forming Inverters," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 2, pp. 1535-1546, Feb. 2024.
- [8] Q. Xu, Y. Xu, C. Zhang and P. Wang, "A Robust Droop-Based Autonomous Controller for Decentralized Power Sharing in DC Microgrid Considering Large-Signal Stability," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1483-1494, March 2020.
- [9] Q. -C. Zhong, "Power-Electronics-Enabled Autonomous Power Systems: Architecture and Technical Routes," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 7, pp. 5907-5918, July 2017.
- [10] M. Davari and Y. A. -R. I. Mohamed, "Robust Droop and DC-Bus Voltage Control for Effective Stabilization and Power Sharing in VSC Multiterminal DC Grids," *IEEE Transactions on Power Electronics*, vol. 33, no. 5, pp. 4373-4395, May 2018.
- [11] C. E. de Souza, K. A. Barbosa and A. Trofino, "Robust Filtering for Linear Systems with Convex-Bounded Uncertain Time-Varying Parameters," *IEEE Transactions on Automatic Control*, vol. 52, no. 6, pp. 1132-1138, June 2007.
- [12] H. Zhang, J. Zhou, Q. Sun, J. M. Guerrero and D. Ma, "Data-Driven Control for Interlinked AC/DC Microgrids Via Model-Free Adaptive Control and Dual-Droop Control," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 557-571, March 2017.
- [13] C. Yang, T. Zheng, M. Bu, P. Li and J. M. Guerrero, "Distributed Model-Free Adaptive Control Strategy for Hybrid AC/DC Microgrid With Event-Triggered Mechanism," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 8, pp. 9077-9086, Aug. 2024.
- [14] D. Glover and A. Dubey, "Learning Volt-VAR Droop Curves to Optimally Coordinate Photovoltaic (PV) Smart Inverters," *IEEE Transactions on Industry Applications*, vol. 61, no. 1, pp. 859-871, Jan.-Feb. 2025.
- [15] Y. Guo, B. C. Pal and R. A. Jabr, "Model-Free Optimal Control of Inverter for Dynamic Voltage Support," *IEEE Transactions on Power Systems*, vol. 38, no. 6, pp. 5860-5871, Nov. 2023.
- [16] S. -W. Lin, C. -C. Chu and C. -F. Tung, "Data-Driven Distributed Q-Learning Droop Control for Frequency Synchronization and Voltage Restoration in Isolated AC Micro-Grids," *IEEE Transactions on Industry Applications*, vol. 59, no. 6, pp. 7306-7317, Nov.-Dec. 2023.
- [17] R. Yan, Y. Wang, Y. Xu and J. Dai, "A Multiagent Quantum Deep Reinforcement Learning Method for Distributed Frequency Control of Islanded Microgrids," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 4, pp. 1622-1632, Dec. 2022.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

- [18] H. H. Goh, *et al.*, "An Assessment of Multistage Reward Function Design for Deep Reinforcement Learning-Based Microgrid Energy Management," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4300-4311, Nov. 2022.
- [19] Y. Huang, G. Li, C. Chen, Y. Bian, T. Qian and Z. Bie, "Resilient Distribution Networks by Microgrid Formation Using Deep Reinforcement Learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4918-4930, Nov. 2022.
- [20] L. Omnes, A. Marot and B. Donnot, "Adversarial Training for a Continuous Robustness Control Problem in Power Systems," 2021 IEEE Madrid PowerTech, Madrid, Spain, 2021, pp. 1-6.
- [21] Y. Wang and B. C. Pal, "Destabilizing Attack and Robust Defense for Inverter-Based Microgrids by Adversarial Deep Reinforcement Learning," *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4839-4850, Nov. 2023.
- [22] H. Liu and W. Wu, "Two-Stage Deep Reinforcement Learning for Inverter-Based Volt-VAR Control in Active Distribution Networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2037-2047, May 2021.
- [23] Y. Wang and B. C. Pal, "Least Effort Attack for Inverter-based Microgrids by Deep Reinforcement Learning," 2024 3rd International Conference on Power Systems and Electrical Technology (PSET), Tokyo, Japan, 2024, pp. 52-57.
- [24] S. d. J. Manrique Machado, S. A. O. da Silva, J. R. B. d. A. Monteiro and A. A. de Oliveira, "Network Modeling Influence on Small-Signal Reduced-Order Models of Inverter-Based AC Microgrids Considering Virtual Impedance," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 79-92, Jan. 2021.
- [25] A. Gorbunov, J. C. -H. Peng, J. W. Bialek and P. Vorobev, "Identification of Stability Regions in Inverter-Based Microgrids," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2613-2623, July 2022.
- [26] Z. Tu, W. Zhang and W. Liu, "Deep Reinforcement Learning Control for Pulsed Power Load Online Deployment in DC Shipboard Integrated Power System," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 3557-3567, July 2023.
- [27] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, "Uncertainty and robustness," in *Feedback Control Theory*. New York, NY, USA: Macmillan, 1992, ch. 4, pp.45-62.