WIDE-AREA DAMPING CONTROLLER VIA REINFORCEMENT LEARNING FOR POWER NETWORKS WITH WIND AND SOLAR FARMS

Muhammad Nadeem, MirSaleh Bahavarnia, and Ahmad F. Taha

***Abstract*— Accurately predicting power system behavior is becoming more complex with the increased penetration of uncertain wind and solar-based renewable resources. Hence, there is a growing motivation to transition from model-based feedback control strategies to completely model-free counterparts. Reinforcement learning (RL) is a key methodology in designing a model-free controller. Various studies have been carried out to study voltage/frequency control strategies via RL. However, they usually consider a simplified power system model either by completely neglecting system differential equations (and thus only modeling the system via power balance equations) or considering simplified generator models. Furthermore, damping system-wide oscillations after large disturbances are usually ignored in the controller design. In contrast, we propose an RL-based wide-area damping controller (WADC) for an advanced power system model with comprehensive higher-order generator dynamics, power electronics-based wind and solar models, and composite load dynamics. The presented controller sends control signals to synchronous generators, wind, and solar power plants to actively adjust their power and voltage setpoints—thereby providing damping to the system oscillations after large disturbances. Case studies demonstrate that the system's transient stability can be significantly improved after large disturbances.**

***Index Terms*— Reinforcement learning, solar and wind-based power system models, feedback control, adaptive control.**

I. INTRODUCTION AND MOTIVATION

MODEL-based stability analysis and state/output feedback control have been the mainstream approaches to studying power system dynamics [1], [2]. These models, based on complex differential algebraic equations (DAE) offer invaluable insights into grid behavior, allowing for the design of sophisticated controllers that can effectively regulate voltage, frequency, and other critical parameters. However, the effectiveness of model-based control is contingent upon the accuracy of these models, which are inherently limited by simplifications and assumptions.

As power systems become more uncertain and dynamic due to the integration of renewables and the proliferation of distributed energy resources, accurately modeling their behavior becomes an increasingly daunting task. This leads to control strategies that are inadequate for real-world challenges [3]. In response to these evolving issues, there is a growing motivation to transition from traditional model-based feedback control to completely model-free control approaches [4]–[6]. This has also been highly encouraged by the recent developments in the wide-area monitoring systems technologies, such as phasor measurement units (PMUs) and robust state estimation algorithms [4].

Using measurements received from the PMUs, these estimation techniques can efficiently estimate all the states of the power system including the states of generators, renewables, loads, and network. This paves the way for completely model-free control techniques in the future power grid.

RL is one of the key techniques in designing model-free feedback control strategies. This is because RL is a self-learning technique in which the agent (or in our case the controller itself) learns the optimal control policy dynamically by merely continuously interacting with its environment or simulation while satisfying the objective function or goal. In this regard, many researchers have recently proposed various RL control algorithms for power systems. In [7], [8], an RL-based optimal frequency controller has been proposed to minimize frequency deviations and the required control action for a simplified second-order power system model. Similarly, in [9] a voltage control strategy has been proposed using deep RL. The RL algorithm is trained to minimize the voltage deviation on all buses by actively adjusting generator bus voltages. However, system dynamic equations (i.e., generator model and power electronics-based model of renewables) are not considered; the corresponding power system is only modeled via power balance equations between generators and the network.

In [10]–[12] a multi-area automatic generation control (AGC) has been designed for frequency regulation using various RL-based techniques such as a deep deterministic policy gradient (DDPG), integral RL (actor-critic), and Q-learning based approaches. Similarly, in [13], [14] various reactive power control schemes have been proposed to improve system voltages via different RL-based techniques. A comprehensive review of the application of RL in the control and stability of power systems can be found in [3], [15].

All of the techniques in the literature usually consider simplified power system models (modeling lower order generator models), neglect power system algebraic constraints (current/power balance equations), and do not provide a control mechanism for (power-electronics)-based wind/solar dynamic models. Furthermore, most of the studies usually focus only on minimizing voltage/frequency deviations and do not consider damping system-wide low-frequency oscillations (LFOs) and ultra-low frequency oscillations (ULFOs) in their architecture. Damping LFOs and ULFOs are critical in the power grid as they can lead to system instability and limit power transfer capability. Furthermore, considering power system algebraic

The authors are with the Civil and Environmental Engineering Department, Vanderbilt University, 2201 West End Ave, Nashville, TN 37235, USA. Email addresses: muhammad.nadeem@vanderbilt.edu, mirsaleh.bahavarnia@vanderbilt.edu, ahmad.taha@vanderbilt.edu. This work is supported by the National Science Foundation under Grants 2152450 and 2151571.

constraints while designing a feedback control strategy is crucial because these constraints model the network dynamics. Thus without considering these constraints, the power system models cannot capture the effects of various load dynamics (such as constant impedance or motor-based loads) and topological changes (such as the tripping of transmission lines, etc) [2].

Some recent work has been carried out to design various centralized and decentralized RL-based damping controllers. In particular, the study [16] designs a wide-area damping controller based on a reduced system model. In [17] a dense centralized damping controller has been proposed based on a policy iteration algorithm. Similarly, in [18] a decentralized off-policy iteration-based damping controller has been designed. However, in all these studies again a simplified power system model is considered (considering lower order generator dynamics and neglecting system algebraic constraints) and no feedback control mechanism for renewables such as solar and wind-based power plants has been designed. To that end, we design an RL-based wide-area damping controller (RL-WADC) for an interconnected power system model with comprehensive 9^{th}-order synchronous generator dynamics, detailed (power-electronics)-based wind and solar dynamics models, and composite load dynamics consisting of constant power, constant impedance, and motor loads. This brings the control problem closer to being more applicable and realistic. The technical contributions of the paper are as follows:

- We present a completely model-free approach to designing a wide-area damping controller for an advanced interconnected power system model. The presented approach considers: higher-order synchronous generators dynamics, detailed (power electronic)-based models of wind/solar-based power plants, models of the system algebraic constraints, and various load dynamics (constant power/impedance and motor-based loads) in its control architecture (Section II).
- The presented approach acts as a secondary control loop and sends additional control signals not only to the conventional power plants but also to the renewable resources (solar and wind-based power plants) so that they can participate in improving system transient stability after large disturbances in load/renewables (Sections III and IV).
- We test the performance of the presented RL-WADC under various load and renewable uncertainties on a modified IEEE 9-bus system—which is one of the widely used test systems in control/stability studies of power systems. The advantages of presented RL-WADC are also demonstrated by comparing the system dynamic response with conventional/primary controllers of the power system and with RL-WADC acting on top of them (Section V).

Paper Notation. All matrices and vectors are denoted using boldface type. Sets are represented using calligraphic font, such as $\mathcal{R}, \mathcal{N}$, etc. The symbol $\mathbb{R}^a$ represents the set of column vectors with 'a' elements. Similarly, $\mathbb{R}^{a\times b}$ denotes the set of real matrices with dimensions 'a' by 'b'. The symbol $\boldsymbol{O}$ represents a zero matrix, $\boldsymbol{I}$ denotes an identity matrix of appropriate size, $\cup$ denotes the union between two sets while the symbol $\otimes$ represents the Kronecker product. In addition, $\mathbb{S}_{++}^{a\times b}$ signifies the set of positive definite matrices of size 'a' by 'b'. The notation $\text{vec}(\boldsymbol{M})$ represents the vectorization of matrix $\boldsymbol{M}$ which is computed by stacking the columns of $\boldsymbol{M}$ on top of each other while $\hat{\boldsymbol{A}}$ denotes the half vectorization of a symmetric matrix $\boldsymbol{A}$ with off-diagonal element taken as $2A_{ij}$. Similarly the notation $\bar{\boldsymbol{B}}$ represent the half vectorization plus the off-diagonal element of matrix $\boldsymbol{B}$, such as for $\boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\bar{\boldsymbol{B}} = [1, 3, 4]^\top + [0, 2, 4]^\top$. Furthermore, the units for all the quantities are in per unit (p.u) unless otherwise specified, also, in some cases for brevity, the time stamp for some of the vectors is removed such as $\boldsymbol{u}_G(t)$ is written just as $\boldsymbol{u}_G$ vice versa.

II. Interconnected Power System Model with Wind, Solar Plants, and Composite Loads

We model a power network with R solar farms, G conventional power plants, W wind-based power plants, and L_z, L_p, L_k, number of constant power, constant impedance, and motor-based loads, respectively. The complete electrical grid is represented via graph with $\mathcal{E}$ set of transmission lines, and $\mathcal{N} = \{1, ..., N\}$ set of nodes or buses. The set of buses $\mathcal{N}$ contains; $\mathcal{G} = \{1, ..., G\}$ set of buses with synchronous generators, $\mathcal{R} = \{1, ..., R\}$ set of buses with solar farms, $\mathcal{W} = \{1, ..., W\}$ set of buses connected to a wind-based power plant, $\mathcal{L}$ set of buses collecting L_z, L_p, and L_k loads, and $\mathcal{U}$ set of non-unit buses (meaning buses not connected to any elements), such that $\mathcal{N} = \mathcal{R} \cup \mathcal{W} \cup \mathcal{G} \cup \mathcal{U} \cup \mathcal{L}$.

The overall power network is described via nonlinear differential-algebraic equations (NDAEs) as follows:

$$\textit{differential equations:} \quad \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}_a, \boldsymbol{x}_d, \boldsymbol{u}) \tag{1a}$$

$$\textit{algebraic equations:} \quad \boldsymbol{0} = \boldsymbol{h}(\boldsymbol{x}_a, \boldsymbol{x}_d, \boldsymbol{u}). \tag{1b}$$

The differential equations model the detailed dynamics of solar plants, wind-based generators, conventional power plants, and composite loads while the algebraic equations describe the algebraic constraints in the power network which are the power/current balance equations. In (1) $\boldsymbol{x}_a \in \mathbb{R}^{n_a}$ denotes algebraic states, $\boldsymbol{x}_d \in \mathbb{R}^{n_d}$ represents dynamic states and it lumps solar plants, wind-based plant, loads, and conventional power plant state variables, and $\boldsymbol{u} \in \mathbb{R}^{n_u}$ denotes the power systems control inputs. The more detailed explanations about each of these variables can be found in [1], [19], [20].

That being said, for control purposes we approximate the overall power system model in the following linear DAE format:

$$\boldsymbol{E}\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \tag{2}$$

where $\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{x}_d^\top & \boldsymbol{x}_a^\top \end{bmatrix}^\top \in \mathbb{R}^n$ is the overall state vector and $\boldsymbol{E}$ is a singular matrix encoding algebraic constraints with rows of zeros.

III. Preliminaries and Problem Formulation

Here we discuss the preliminaries and the problem formulation for designing a state feedback controller for the

interconnected power system model described in (2). The primary objective is to design a feedback control law that minimizes a certain performance criterion or cost function. To that end, let us consider the closed-loop dynamics of (2) along with its performance criterion vector $\boldsymbol{z} \in \mathbb{R}^n$ as follows:

$$\boldsymbol{E}\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}_C \tag{3a}$$

$$\boldsymbol{z} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}_C \tag{3b}$$

where the control policy $\boldsymbol{u}_C$ for every k^{th} dispatch period is defined as

$$\boxed{\textbf{Control policy:} \quad \boldsymbol{u}_C(t) = \boldsymbol{u}_0^k - \boldsymbol{K}\left(\boldsymbol{x}(t) - \boldsymbol{x}^k\right)} \tag{4}$$

in which $\boldsymbol{x}^k$ is the state vector before fault/disturbance, $\boldsymbol{u}_0^k$ is the initial set point of the input $\boldsymbol{u}$ (which is usually determined by running power flow for every dispatch period), and $\boldsymbol{K} \in \mathbb{R}^{n_u \times n}$ is the feedback controller gain matrix. The main idea in the given control law is to design a gain matrix $\boldsymbol{K}$ such that the power system converges back to its equilibrium and its transient stability can be improved after a large disturbance. To do that, one first needs to design the perturbed closed-loop dynamics whose states are then driven asymptotically to zero while minimizing the performance criterion associated with $\boldsymbol{z}$. With that in mind, let us assume the power system model (2) is disturbed by an external uncertainty (which can be because of disturbance in load demand or renewable, etc.) and thus the system moves from its initial equilibrium status to a new status $\boldsymbol{x}'$. Then, the perturbed closed loop dynamics and its associated performance criterion vector $\boldsymbol{z}$ can be written as follows:

$$\boldsymbol{E}\Delta\dot{\boldsymbol{x}} = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{K})\Delta\boldsymbol{x} \tag{5a}$$

$$\Delta\boldsymbol{z} = (\boldsymbol{C} - \boldsymbol{D}\boldsymbol{K})\Delta\boldsymbol{x} \tag{5b}$$

where $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}'$. The main objective of the matrix $\boldsymbol{K}$ is to drive $\Delta\boldsymbol{x}$ asymptotically to zero which in other words means that the power system (2) will converge to the new steady-state after large disturbance in system dynamics.

With little abuse of notation, for the sake of simplicity, from now on we drop the Δ notation from $\Delta\boldsymbol{x}$, and $\Delta\boldsymbol{z}$ and simply use $\boldsymbol{x}$ and $\boldsymbol{z}$ instead, respectively. To that end, notice in (2) that conventional power plants, solar plants, wind power plants, and motor loads are dynamic systems while loads, non-unit buses, and other interconnections are considered static systems. Thus, the algebraic variables $\boldsymbol{x}_a$ in the power system model (the voltage and current phasors) are redundant states and thus can be eliminated while designing a feedback controller [2]. This can be done by finding an explicit equation for the algebraic variables using the algebraic constraint model and plugging it back into the system dynamical model as follows: Considering

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{dd} & \boldsymbol{A}_{da} \\ \boldsymbol{A}_{ad} & \boldsymbol{A}_{aa} \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_d^\top & \boldsymbol{x}_a^\top \end{bmatrix}^\top, \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_d^\top & \boldsymbol{B}_a^\top \end{bmatrix}^\top$$

one can write the DAE system (2) into the separate dynamic and algebraic model as follows:

$$\dot{\boldsymbol{x}}_d = \boldsymbol{A}_{dd}\boldsymbol{x}_d + \boldsymbol{A}_{da}\boldsymbol{x}_a + \boldsymbol{B}_{ud}\boldsymbol{u} \tag{6a}$$

$$\boldsymbol{0} = \boldsymbol{A}_{ad}\boldsymbol{x}_d + \boldsymbol{A}_{aa}\boldsymbol{x}_a + \boldsymbol{B}_{ua}\boldsymbol{u}. \tag{6b}$$

Now assuming $\boldsymbol{A}_{aa}$ to be invertible which is a standard assumption in power systems [2], [20].Then an explicit equation for the algebraic variable $\boldsymbol{x}_a$ from (6b) can be written as:

$$\boldsymbol{x}_a = {\boldsymbol{A}_{aa}}^{-1}(-\boldsymbol{A}_{ad}\boldsymbol{x}_d - \boldsymbol{B}_{ua}\boldsymbol{u}). \tag{7}$$

Finally, based on this, the closed-loop model (5) can be rewritten in equivalent ODE format as follows:

$$\dot{\boldsymbol{x}}_d(t) = (\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{B}}\boldsymbol{K}_d)\boldsymbol{x}_d(t) \tag{8a}$$

$$\boldsymbol{z}_1(t) = (\tilde{\boldsymbol{C}} - \tilde{\boldsymbol{D}}\boldsymbol{K}_d)\boldsymbol{x}_d(t) \tag{8b}$$

where $\boldsymbol{K}_d \in \mathbb{R}^{n_u \times n_d}$ and the rest of the system matrices are defined as:

$$\tilde{\boldsymbol{A}} = \boldsymbol{A}_{dd} - \boldsymbol{A}_{da}\boldsymbol{A}_{aa}^{-1}\boldsymbol{A}_{ad}, \quad \tilde{\boldsymbol{B}} = \boldsymbol{B}_d - \boldsymbol{A}_{da}\boldsymbol{A}_{aa}^{-1}\boldsymbol{B}_a$$

$$\tilde{\boldsymbol{D}} = \boldsymbol{B}_d - \boldsymbol{C}_a\boldsymbol{A}_{aa}^{-1}\boldsymbol{B}_a, \quad \tilde{\boldsymbol{C}} = \boldsymbol{C}_d - \boldsymbol{C}_a\boldsymbol{A}_{aa}^{-1}\boldsymbol{A}_{ad}.$$

Notice that the computed gain $\boldsymbol{K}_d$ from ODE (8) can be plugged back into the complete interconnected power system as $\boldsymbol{K} = [\boldsymbol{K}_d \quad \boldsymbol{O}]$. Now, by just using the knowledge of the system matrices $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$, [21] has proposed an iterative algorithm to design a linear state feedback controller described as follows:

Theorem 1 ([21]): Let $\boldsymbol{K}_{d_0}$ be the initial stabilizing controller gain matrix such that $\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{B}}\boldsymbol{K}_{d_0}$ is Hurwitz and $\boldsymbol{P}_k \in \mathbb{S}_{++}^{n_d \times n_d}$, $k = 0, 1, \ldots$ be the unique solution to the following Lyapunov equation (which is linear in $\boldsymbol{P}_k$)

$${\boldsymbol{A}_c}^\top\boldsymbol{P}_k + \boldsymbol{P}_k\boldsymbol{A}_c + \boldsymbol{Q} + \boldsymbol{K}_{d_k}^\top\boldsymbol{R}\boldsymbol{K}_{d_k} - \boldsymbol{N}\boldsymbol{K}_{d_k} - (\boldsymbol{N}\boldsymbol{K}_{d_k})^\top = \boldsymbol{0} \tag{9}$$

with $\boldsymbol{K}_{d_k}$ being updated recursively at each iteration k using

$$\boldsymbol{K}_{d_{k+1}} = \boldsymbol{R}^{-1}(\tilde{\boldsymbol{B}}^\top\boldsymbol{P}_k + \boldsymbol{N}^\top) \tag{10}$$

where $\boldsymbol{A}_c = \tilde{\boldsymbol{A}} - \tilde{\boldsymbol{B}}\boldsymbol{K}_{d_k}$, $\boldsymbol{Q} = \tilde{\boldsymbol{C}}^\top\tilde{\boldsymbol{C}}$, $\boldsymbol{R} = \tilde{\boldsymbol{D}}^\top\tilde{\boldsymbol{D}}$, and $\boldsymbol{N} = \tilde{\boldsymbol{C}}^\top\tilde{\boldsymbol{D}}$. Then, $\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{B}}\boldsymbol{K}_{d_k}$ is Hurwitz, and the solution of $\boldsymbol{P}_k$ as $k \to \infty$ converges to the optimal solution of the well-known algebraic Riccati equation (ARE) given as:

$${\boldsymbol{A}_c}^\top\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}_c - (\boldsymbol{P}\tilde{\boldsymbol{B}} + \boldsymbol{N})\boldsymbol{R}^{-1}(\tilde{\boldsymbol{B}}^\top\boldsymbol{P} + \boldsymbol{N}^\top) + \boldsymbol{Q} = \boldsymbol{0} \tag{11}$$

The above technique requires the complete knowledge of the system matrices $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}$ which is unrealistic as accurate models might not be available or be uncertain specifically in the case of power systems with high penetration of renewable energy resources. Then, to design a similar feedback gain $\boldsymbol{K}_d$ without the knowledge of the system matrices and just using the data (the system state trajectories which can be computed via PMU data, for example), we present the following problem statement:

Problem 1: Given the state trajectories $\boldsymbol{x}_d(t)$ of the LODE system (8a) and its performance metric (8b), design feedback controller $\boldsymbol{K}_d$ without utilizing the knowledge of system matrices.

IV. RL-based Feedback Controller Design

Primarily based on the theory in [22], here we present model-free RL-based centralized feedback controller design. The presented approach computes the feedback controller gain matrix $\boldsymbol{K}_d$ purely based on the continuous interaction with the power system in real-time and does not require the information of system matrices. Thus, the presented technique has a major advantage over model-based feedback controller design as it can adaptively tune the parameters of its gain matrix based on the state information received.

To that end, we rewrite Eq. (8a) as

$$\dot{\boldsymbol{x}}_d(t) = \boldsymbol{A}_c\boldsymbol{x}_d(t) + \tilde{\boldsymbol{B}}(\boldsymbol{K}_{d_k}\boldsymbol{x}_d(t) + \boldsymbol{u}_d(t)) \tag{12}$$

where $\boldsymbol{u}_d = -\boldsymbol{K}_{d_k}\boldsymbol{x}_d$. The objective of the controller is to minimize the energy of the performance criterion vector $\boldsymbol{z}_1$ (8b) which can be written as

$$\begin{aligned}\int_0^\infty \boldsymbol{z}_1^\top \boldsymbol{z}_1 dt &= \int_0^\infty (\tilde{\boldsymbol{C}}\boldsymbol{x}_d - \tilde{\boldsymbol{D}}\boldsymbol{u}_d)^\top(\tilde{\boldsymbol{C}}\boldsymbol{x}_d - \tilde{\boldsymbol{D}}\boldsymbol{u}_d)dt \\ &= \int_0^\infty \boldsymbol{x}_d^\top\boldsymbol{Q}\boldsymbol{x}_d + \boldsymbol{u}_d^\top\boldsymbol{R}\boldsymbol{u}_d - 2\boldsymbol{x}_d^\top\boldsymbol{N}\boldsymbol{u}_d\, dt\end{aligned}$$

with $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{N}$ are given as in Theorem 1. Then, let us consider the optimal cost-to-go as $V(\boldsymbol{x}_d) = \boldsymbol{x}_d^\top\boldsymbol{P}_k\boldsymbol{x}_d$ where $\boldsymbol{P}_k \in \mathbb{S}_{++}^{n_d\times n_d}$. Notice that $V(\boldsymbol{x}_d)$ is a Lyapunov function and in the case of a model-based approach, $\boldsymbol{P}_k$ can be computed via a recursive approach given in Theorem 1 which converges to the solution of the ARE.

With that in mind, taking the derivative of $V(\boldsymbol{x}_d)$ along the trajectories of (12), we then get

$$\dot{V} = \boldsymbol{x}_d^\top(\boldsymbol{A}_c^\top\boldsymbol{P}_k + \boldsymbol{P}_k\boldsymbol{A}_c)\boldsymbol{x}_d + 2(\boldsymbol{u}_d + \boldsymbol{K}_{d_k}\boldsymbol{x}_d)^\top\tilde{\boldsymbol{B}}^\top\boldsymbol{P}_k\boldsymbol{x}_d \tag{13}$$

From (9) let

$$\boldsymbol{X} = \boldsymbol{Q} + \boldsymbol{K}_{d_k}^\top\tilde{\boldsymbol{R}}\boldsymbol{K}_{d_k} - \boldsymbol{N}\boldsymbol{K}_{d_k} - (\boldsymbol{N}\boldsymbol{K}_{d_k})^\top. \tag{14}$$

Then, using (10) we can rewrite (13) as

$$\dot{V} = -\boldsymbol{x}_d^\top\boldsymbol{X}\boldsymbol{x}_d + 2(\boldsymbol{u}_d + \boldsymbol{K}_{d_k}\boldsymbol{x}_d)^\top(\boldsymbol{R}\boldsymbol{K}_{d_{k+1}} - \boldsymbol{N}^\top)\boldsymbol{x}_d. \tag{15}$$

With that in mind, taking integral on both sides of (15), we then have

$$\begin{aligned}\boldsymbol{x}_d^\top\boldsymbol{P}_k\boldsymbol{x}_d\Big|_t^{t+\delta t} &= \int_t^{t+\delta t} 2(\boldsymbol{u}_d + \boldsymbol{K}_{d_k}\boldsymbol{x}_d)^\top(\boldsymbol{R}\boldsymbol{K}_{d_{k+1}} - \boldsymbol{N}^\top)\boldsymbol{x}_d d\tau \\ &\quad - \int_t^{t+\delta t} \boldsymbol{x}_d^\top\boldsymbol{X}\boldsymbol{x}_d d\tau.\end{aligned} \tag{16}$$

Notice that Eq. (16) does not depend on any system matrices of the power system model. Then, if we can solve (16) for $\boldsymbol{P}_k, \boldsymbol{K}_{d_{k+1}}$ we have a solution to the posed Problem 1.

To that end, let us define matrices $\boldsymbol{\delta}_{x_dx_d} \in \mathbb{R}^{s\times\frac{1}{2}n_d(n_d+1)}$, $\boldsymbol{\Upsilon}_{x_dx_d} \in \mathbb{R}^{s\times n_d^2}$, $\boldsymbol{\Upsilon}_{x_du} \in \mathbb{R}^{s\times n_dn_u}$, and $\hat{\boldsymbol{P}} \in \mathbb{R}^{\frac{1}{2}n_d(n_d+1)}$ as follows:

$$\begin{aligned}\boldsymbol{\delta}_{x_dx_d} &= \left[\bar{\boldsymbol{x}}_{\boldsymbol{d}}\Big|_{t_0}^{t_1},\ \bar{\boldsymbol{x}}_{\boldsymbol{d}}\Big|_{t_1}^{t_2},\ \bar{\boldsymbol{x}}_{\boldsymbol{d}}\Big|_{t_2}^{t_3},\ \cdots,\ \bar{\boldsymbol{x}}_{\boldsymbol{d}}\Big|_{t_{s-1}}^{t_s}\right]^\top \\ \boldsymbol{\Upsilon}_{x_dx_d} &= \left[\int_{t_0}^{t_1}\boldsymbol{x}_{\boldsymbol{d}}\otimes\boldsymbol{x}_{\boldsymbol{d}}d\tau,\ \cdots,\ \int_{t_{s-1}}^{t_s}\boldsymbol{x}_{\boldsymbol{d}}\otimes\boldsymbol{x}_{\boldsymbol{d}}d\tau\right]^\top \\ \boldsymbol{\Upsilon}_{x_du} &= \left[\int_{t_0}^{t_1}\boldsymbol{x}_{\boldsymbol{d}}\otimes\boldsymbol{u}_{\boldsymbol{d}}d\tau,\ \cdots,\ \int_{t_{s-1}}^{t_s}\boldsymbol{x}_{\boldsymbol{d}}\otimes\boldsymbol{u}_{\boldsymbol{d}}d\tau\right]^\top \\ \hat{\boldsymbol{P}} &= \left[P_{11},\cdots,2P_{1n_d},P_{22},\cdots,2P_{n_d-1n_d},P_{n_dn_d}\right]^\top\end{aligned} \tag{17}$$

where $s > 0$ are the total number of data samples collected, $\hat{\boldsymbol{P}}$ denotes the half vectorization of symmetric matrix $\boldsymbol{P}$ with off-diagonal element taken as $2P_{ij}$, and vector $\bar{\boldsymbol{x}}_d \in \mathbb{R}^{\frac{1}{2}n_d(n_d+1)}$ given as:

$$\bar{\boldsymbol{x}}_d = \left[\bar{x}_{d,1}^2,\cdots,\bar{x}_{d,1}\bar{x}_{d,n_d},\bar{x}_{d,2}^2,\cdots,\bar{x}_{d,n_d-1}\bar{x}_{d,n_d},\bar{x}_{d,n_d}^2\right]^\top$$

Then, (16) can equivalently be written as

$$\boldsymbol{\Xi}_k\begin{bmatrix}\hat{\boldsymbol{P}}_k\\ \mathrm{vec}(\boldsymbol{K}_{d_{k+1}})\end{bmatrix} = \boldsymbol{\Psi}_k \tag{18}$$

where $\boldsymbol{\Xi}_k \in \mathbb{R}^{s\times(\frac{1}{2}n_d(n_d+1)+n_dn_u)}$ and $\Psi_k \in \mathbb{R}^s$ are given as

$$\boldsymbol{\Xi}_k = \left[\boldsymbol{\delta}_{x_dx_d},\ -2\boldsymbol{\Upsilon}_{x_dx_d}(\boldsymbol{I}_{n_d}\otimes\boldsymbol{K}_{d_k}^\top\boldsymbol{R}) - 2\boldsymbol{\Upsilon}_{x_du}(\boldsymbol{I}_{n_d}\otimes\boldsymbol{R})\right]$$

$$\begin{aligned}\boldsymbol{\Psi}_k = &-2\boldsymbol{\Upsilon}_{x_du}\mathrm{vec}(\boldsymbol{N}) - 2\boldsymbol{\Upsilon}_{x_dx_d}\mathrm{vec}(\boldsymbol{K}_{d_k}^\top\boldsymbol{N}^\top) \\ &- \boldsymbol{\Upsilon}_{x_dx_d}\mathrm{vec}(\boldsymbol{X}).\end{aligned}$$

To that end, Eq. (18) can efficiently be solved in least-squares sense for the variables $\hat{\boldsymbol{P}}_k$, $\mathrm{vec}(\boldsymbol{K}_{d_{k+1}})$ as follows:

$$\begin{bmatrix}\hat{\boldsymbol{P}}_k\\ \mathrm{vec}(\boldsymbol{K}_{d_{k+1}})\end{bmatrix} = (\boldsymbol{\Xi}_k^\top\boldsymbol{\Xi}_k)^{-1}\boldsymbol{\Xi}_k^\top\boldsymbol{\Psi}_k. \tag{19}$$

If (19) has a solution, then the iterative model-based procedure to compute the feedback gain $\boldsymbol{K}_d$ given by (9) and (10) in Theorem 1 can be replaced by model-free approach in (19). With that in mind, we now present a key assumption on the constructed data matrices $\boldsymbol{\Upsilon}_{x_dx_d}, \boldsymbol{\Upsilon}_{x_du}$ which if gets satisfied guarantees the convergence of (19) to the solution of the corresponding ARE.

Assumption 1: The rank of matrix $\left[\boldsymbol{\Upsilon}_{x_dx_d},\ \boldsymbol{\Upsilon}_{x_du}\right]$ is equal to $(\frac{1}{2}n_d(n_d+1) + n_dn_u)$.

Notice that Assumption 1 is satisfied by collecting sufficiently many samples and by adding exploration noise in the control inputs as explained in Section V.

Theorem 2: Given Assumption 1 is satisfied then the recursive solution of (19) as $k \to \infty$ converges to the optimal solution given by the corresponding ARE.

Proof: First, we need to prove that (19) has a unique solution which means we have to show that for

$$\boldsymbol{\Xi}_k\begin{bmatrix}\hat{\boldsymbol{P}}_k\\ \mathrm{vec}(\boldsymbol{K}_{d_{k+1}})\end{bmatrix} = \boldsymbol{O} \tag{20}$$

$\hat{\boldsymbol{P}}_k$ and $\mathrm{vec}(\boldsymbol{K}_{d_{k+1}})$ only has trivial solution, i.e., $\hat{\boldsymbol{P}}_k = \boldsymbol{O}$, $\mathrm{vec}(\boldsymbol{K}_{d_{k+1}}) = \boldsymbol{O}$. This can be proved by contradiction. Let us assume $\boldsymbol{W} = \begin{bmatrix}\boldsymbol{Y}\ \boldsymbol{Z}\end{bmatrix}^\top$ where $\boldsymbol{Y} \in \mathbb{R}^{n_d\times n_d}$ and $\boldsymbol{Z} \in \mathbb{R}^{n_u\times n_d}$, is a non-zero solution to (20). Now, for $\boldsymbol{\Upsilon}_{\bar{x}_d} \in \mathbb{R}^{s\times\frac{1}{2}n_d(n_d+1)}$ Eq. (18) can be expanded and rewritten as follows:

$$\begin{aligned}\boldsymbol{\Xi}_k\begin{bmatrix}\hat{\boldsymbol{P}}_k\\ \mathrm{vec}(\boldsymbol{K}_{d_{k+1}})\end{bmatrix} &= -2\boldsymbol{\Upsilon}_{x_du}\mathrm{vec}(\boldsymbol{N}) - 2\boldsymbol{\Upsilon}_{\bar{x}_d}(\bar{\boldsymbol{N}}) - \boldsymbol{\Upsilon}_{\bar{x}_d}(\hat{\boldsymbol{X}}) \\ &= \left[\boldsymbol{\Upsilon}_{\bar{x}_d},\ 2\boldsymbol{\Upsilon}_{x_du}\right]\left[(\hat{\boldsymbol{X}} - 2\bar{\boldsymbol{N}})^\top,\ \mathrm{vec}(\boldsymbol{N})^\top\right]^\top\end{aligned}$$

or by (20) we have

$$\left[\boldsymbol{\Upsilon}_{\bar{x}_d},\ 2\boldsymbol{\Upsilon}_{x_du}\right]\left[(\hat{\boldsymbol{X}} - 2\bar{\boldsymbol{N}})^\top,\ \mathrm{vec}(\boldsymbol{N})^\top\right]^\top = \boldsymbol{O}. \tag{21}$$

By Assumption 1 we know that $\left[\boldsymbol{\Upsilon}_{\bar{x}_d},\ 2\boldsymbol{\Upsilon}_{x_du}\right]$ has a full-column rank, which means that (21) only has trivial solution thus implying $\boldsymbol{N} = \boldsymbol{O}$ and $\boldsymbol{X} = \boldsymbol{O}$. Now from (9) with $\boldsymbol{X} = \boldsymbol{O}$ we have

$$\boldsymbol{A}_c^\top\boldsymbol{Y} + \boldsymbol{Y}\boldsymbol{A}_c = \boldsymbol{O}. \tag{22}$$

Since $\tilde{\boldsymbol{A}}_c$ is Hurwitz, then the only solution for (22) is $\boldsymbol{Y} = \boldsymbol{O}$. Then, from (10) with $\boldsymbol{Y} = 0$ and $\boldsymbol{N} = \boldsymbol{O}$ we have $\boldsymbol{Z} = \boldsymbol{O}$, thus $\boldsymbol{W} = \boldsymbol{O}$ which contradicts with our assumption that $\boldsymbol{W} \neq \boldsymbol{O}$. Thus, a unique solution exists for Eq. (19).

With that in mind, since $\hat{\boldsymbol{P}}_k$ and $\mathrm{vec}(\boldsymbol{K}_{d_{k+1}})$ can uniquely be determined, then the policy iteration in (19) is equivalent to (9) and (10) and thus by Theorem 1 the convergence to the ARE solution is guaranteed. This completes the proof. ■

Computing controller gain matrix $\boldsymbol{K}_d$ by solving (19) gives us a model-free approach to designing an optimal feedback controller. The presented approach does not require any information regarding the power system model and thus is an

Algorithm 1: RL-WADC for Interconnected Power Systems

1 ***Save data***: Simulate power system model (8) with control input $\boldsymbol{u}(t) = \boldsymbol{u}_0 + \boldsymbol{e}(t)$ and construct the data matrices given (17) until Assumption 1 is satisfied.

2 ***Compute $\boldsymbol{K}_d$ iteratively***: Starting with initial stabilizing controller $\boldsymbol{K}_{d_0}$ compute $\boldsymbol{K}_{d_k}$ recursively for $k = 0, 1, ...$ via solving the following equation till $|\hat{\boldsymbol{P}}_k - \hat{\boldsymbol{P}}_{k-1}| < \epsilon$ where ϵ is a user-defined threshold

$$\begin{bmatrix} \hat{\boldsymbol{P}}_k \\ \text{vec}(\boldsymbol{K}_{d_{k+1}}) \end{bmatrix} = (\boldsymbol{\Xi}_k^\top \boldsymbol{\Xi}_k)^{-1} \boldsymbol{\Xi}_k^\top \boldsymbol{\Psi}_k.$$

3 ***Apply the controller***: Remove the exploration noise $\boldsymbol{e}(t)$ from the control input, design $\boldsymbol{K}$ as $\boldsymbol{K} = [\boldsymbol{K}_d \;\; \boldsymbol{O}]$, and apply the following control policy to the complete NDAE power system model

$$\boldsymbol{u}_C(t) = \boldsymbol{u}_0^k - \boldsymbol{K}\left(\boldsymbol{x}(t) - \boldsymbol{x}^k\right).$$

adaptive approach that can finely tune its controller gain matrix depending on the received data. This is especially useful for the power grid with uncertain renewable resources and dynamic loads. In the following section, we assess the performance of the presented RL-WADC through numerical case studies on an interconnected power system model with detailed renewable dynamics and various composite loads.

V. Case Studies

The presented RL-WADC has been tested on a modified IEEE 9-bus test network. This test system consists of one conventional steam power plant at Bus 1, a solar farm at Bus 2, a wind power plant at Bus 3, constant power and constant impedance loads at Buses 5 and 6, and a motor load at Bus 8. Further details about the parameters of solar/wind farms, conventional power plants, and dynamics of the test system used in this study can be found in [1], [19], [20].

All the simulations are carried out on a personal computer with 64GB RAM and an Intel $i9 - 11900K$ processor. The power system NDAE dynamics have been modeled in MATLAB R2021a and are simulated using `ode15s` MATLAB DAEs solver. The initial conditions for the `ode15s` are computed using power flow studies in MATPOWER through function `runpf`. The volt-ampere base for the power system is chosen to be $S_b = 100$MVA while the frequency base is selected to be $w_b = 120\pi$rad/s. It is worth mentioning here that in the above test system, to meet the overall load demand in the steady state, on average 63% of the power is generated by the renewables (both solar and wind power plants), hence it can be considered as a renewable heavy power grid.

To design the RL-WADC based feedback controller gain matrix $\boldsymbol{K}_d$ the main requirement is to compute the data matrices $\boldsymbol{\delta}_{x_d x_d}$, $\boldsymbol{\Upsilon}_{x_d x_d}$, $\boldsymbol{\Upsilon}_{x_d u}$ as given in (17). To do that, the power system model is simulated with exploration noise in the initial control policy and the data matrices are saved until Assumption 1 is satisfied. Notice that, since the power system model is already stable, the initial control gain $\boldsymbol{K}_{d_0}$ is

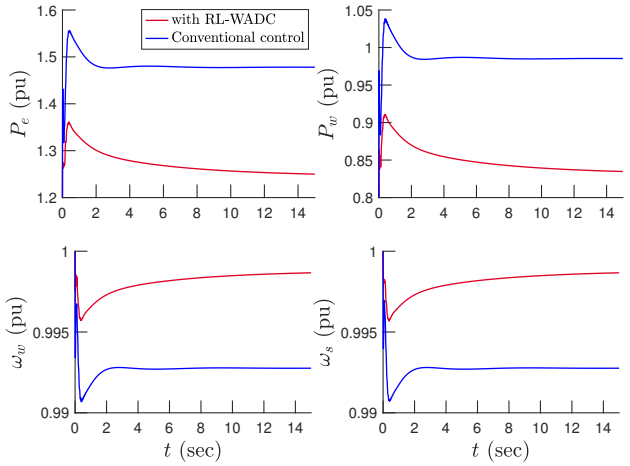

Fig. 1. Performance under abrupt load disturbance: Active power injected and relative speed of solar and wind power plant, respectively.

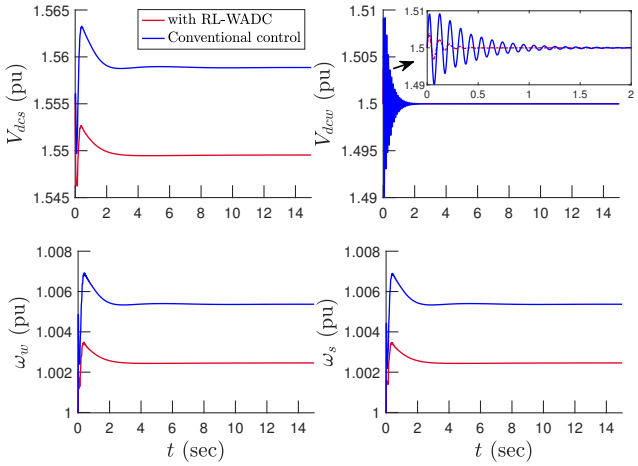

Fig. 2. Performance under renewable and load disturbance: DC link voltage and relative speed of both solar and wind power plant.

then chosen to be zero. The selection of the type of exploration noise is not trivial in RL-based and other machine learning-based techniques and various types of noises have been used in the literature such as sum of sinusoids, exponentially decreasing noise, and random bounded noise and vice versa. In this work, we have used a sum of sinusoids with different frequencies as exploration noise in the initial control policy as given in [22]. The overall implementation of the presented RL-WADC is summarized in Algorithm 1. Notice that in the case of parametric uncertainty the Algorithm 1 can simply be repeated to adjust gain matrix $\boldsymbol{K}$ accordingly.

A. Performance under transient conditions

Here we assess the performance of the presented RL-WADC under transient conditions created by various disturbances in renewables and load demand. To that end, we initially test the performance of RL-WADC by adding, right after $t > 0$, a random step increasing in load demand at Bus 6. This can be a result of a sudden generator disconnection or a transmission line trip, hence increasing the overall load demand of the power network. The system dynamic response with just system primary controllers and with RL-WADC acting on top of them is shown in Fig. 1. Notice that the

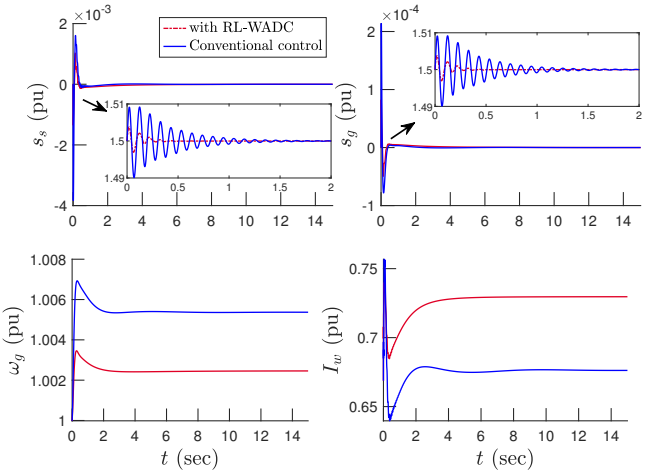

Fig. 3. Performance under renewable and load disturbance: solar plant slip (top left), generator slip (top right) and frequency (bottom left), and wind power plant current (bottom right).

primary/conventional controllers in the test power system include an automatic voltage regulator (AVR), governor, and power system stabilizer (PSS) for conventional power plants [20] and proportional-integral (PI) type controllers for both solar and wind power plants [19], [20]. The proposed RL-WADC acts on top of these primary controllers and sends additional control signals via input $\boldsymbol{u}$. The overall objective of the proposed RL-WADC is to improve system transient stability by adding damping to the system oscillation and bringing the system back close to its nominal/equilibrium value after a large disturbance.

To that end, we can see from Fig. 1 that without the proposed controller the power output from both solar and wind power plants have oscillations while with RL-WADC there is significant damping. Similarly, for the frequencies of both solar and wind power plants, we can see that without RL-WADC after disturbance their frequencies dip close to 0.99 (pu) while with the proposed controller the overall dip is around 0.994 (pu), thus improving system frequency nadir.

To further advocate the performance of the presented RL-based damping controller, here we add further disturbance this time by decreasing load demand by 20% and solar irradiance by 10% on the solar power plant. The results are presented in Figs. 2 and 3. We can see from Fig. 2 that with the proposed controller after disturbance there is significant damping in the slip of synchronous generator, wind, and solar power plant showing significant improvements in LFOs and ULFOs. Similarly, from Fig. 3 we can see that there is a significant damping in the system oscillations thus the overall transient stability after the large disturbance has been improved.

References

[1] M. Nadeem, M. Bahavarnia, and A. F. Taha, "On wide-area control of solar-integrated dae models of power grids," in *American Control Conference (ACC)*. IEEE, 2023, pp. 4495–4500.

[2] T. Sadamoto, A. Chakrabortty, T. Ishizaki, and J.-i. Imura, "Dynamic modeling, stability, and control of power systems with distributed energy resources: Handling faults using two control methods in tandem," *IEEE Control Systems Magazine*, vol. 39, no. 2, pp. 34–65, 2019.

[3] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Reinforcement learning for selective key applications in power systems: Recent advances and future challenges," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2935–2958, 2022.

[4] A. Chakrabortty, "Wide-area control of power systems: Employing data-driven, hierarchical reinforcement learning," *IEEE Electrification Magazine*, vol. 9, no. 1, pp. 45–52, 2021.

[5] A. F. Dizche, A. Chakrabortty, and A. Duel-Hallen, "Sparse wide-area control of power systems using data-driven reinforcement learning," in *2019 American Control Conference (ACC)*, 2019, pp. 2867–2872.

[6] T. Yu, B. Zhou, K. W. Chan, L. Chen, and B. Yang, "Stochastic optimal relaxed automatic generation control in non-markov environment based on multi-step $q(\lambda)$ learning," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1272–1282, 2011.

[7] 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, 2022.

[8] J. Feng, W. Cui, J. Cortés, and Y. Shi, "Online event-triggered switching for frequency control in power grids with variable inertia," *arXiv preprint arXiv:2408.15436*, 2024.

[9] S. Wang, J. Duan, D. Shi, C. Xu, H. Li, R. Diao, and Z. Wang, "A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4644–4654, 2020.

[10] Z. Yan and Y. Xu, "A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4599–4608, 2020.

[11] H. Wang, Z. Lei, X. Zhang, J. Peng, and H. Jiang, "Multiobjective reinforcement learning-based intelligent approach for optimization of activation rules in automatic generation control," *IEEE access*, vol. 7, pp. 17 480–17 492, 2019.

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

[13] Y. Shi, G. Qu, S. Low, A. Anandkumar, and A. Wierman, "Stability constrained reinforcement learning for real-time voltage control," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 2715–2721.

[14] L. Yin, C. Zhang, Y. Wang, F. Gao, J. Yu, and L. Cheng, "Emotional deep learning programming controller for automatic voltage control of power systems," *IEEE Access*, vol. 9, pp. 31 880–31 891, 2021.

[15] M. Glavic, "(deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annual Reviews in Control*, vol. 48, pp. 22–35, 2019.

[16] S. Mukherjee, A. Chakrabortty, H. Bai, A. Darvishi, and B. Fardanesh, "Scalable designs for reinforcement learning-based wide-area damping control," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2389–2401, 2021.

[17] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, 2009.

[18] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 10, pp. 693–697, 2012.

[19] H. N. V. Pico and V. Gevorgian, "Blackstart capability and survivability of wind turbines with fully rated converters," *IEEE Transactions on Energy Conversion*, vol. 37, no. 4, pp. 2482–2497, 2022.

[20] M. Nadeem, M. Bahavarnia, and A. F. Taha, "Robust feedback control of power systems with solar plants and composite loads," *IEEE Transactions on Power Systems*, vol. 39, no. 3, pp. 4949–4962, 2024.

[21] W. Arnold and A. Laub, "Generalized eigenproblem algorithms and software for algebraic riccati equations," *Proceedings of the IEEE*, vol. 72, no. 12, pp. 1746–1754, 1984.

[22] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.