Wide-Area Feedback Control for Renewables-Heavy Power Systems: A Comparative Study of Reinforcement Learning and Lyapunov-Based Design

Muhammad Nadeem[a], MirSaleh Bahavarnia[a], Ahmad F. Taha[a]

[a]*Vanderbilt University, Civil and Environmental Engineering Department, Nashville, 37235, TN, US*

Abstract

As renewable energy sources become more prevalent, accurately modeling power grid dynamics is becoming increasingly more complex. Concurrently, data acquisition and realtime system state monitoring are becoming more available for control centers. This motivates shifting from *model- and Lyapunov-based* feedback controller designs toward *model-free* ones. Reinforcement learning (RL) has emerged as a key tool for designing model-free controllers. Various studies have been carried out to study voltage/frequency control strategies via RL. However, usually a simplified system model is used neglecting detailed dynamics of solar, wind, and composite loads—and damping system-wide oscillations and modeling power flows are all usually ignored. To that end, we pose an optimal feedback control problem for a detailed renewables-heavy power system, defined by a set of nonlinear differential algebraic equations (NDAE). The control problem is solved using a completely model-free design via RL as well as using a model-based approach built upon the Lyapunov stability theory with guarantees. The paper in its essence seeks to explore whether data-driven feedback control should be used in power grids over its model-driven counterpart. Theoretical developments and thorough case studies are presented with an eye on this exploration. Finally, a detailed analysis is provided to delineate the strengths and weaknesses of both approaches for renewables-heavy grids.

Keywords: Reinforcement learning, solar and wind-based power plants, feedback control, Lyapunov stability.

1. Introduction and motivation

With the increased penetration of wind and solar-based energy resources, the overall transient stability of power systems is deteriorating. In particular, there is a significant increase in low and ultra-low frequency oscillations (LFOs and ULFOs) in the future power grids with high penetration of renewables. These LFOs and ULFOs if not properly damped can cause system-wide instabilities and blackouts [1]. State/output feedback controllers can play a crucial role in mitigating such oscillations and improving system transient stability after large disturbances. Based on realtime measurements, these feedback controllers can provide additional control signals to the power plants thus improving system robustness toward disturbances [2].

Traditionally, in the literature, model-based approaches are utilized to design feedback controllers. These models, based on advanced differential-algebraic equations (DAEs) provide precise, physics-based descriptions of system dynamics, allowing for the design of sophisticated feedback controllers that can effectively regulate frequency, improve LFOs/ULFOs, and other grid parameters [3–6]. However, the performance of model-based control heavily relies on the accuracy of these models, which are inherently limited by assumptions and simplifications. Also, as power networks become more dynamic and uncertain due to the integration of renewables and other distributed resources, accurately modeling their dynamics becomes an increasingly daunting task. Moreover, even if the accurate model is computed, re-adjusting it in realtime and re-designing the feedback control law (every time the model changes) becomes highly inefficient and impractical. Thus, there is a growing motivation to transition toward completely model-free feedback control strategies.

Reinforcement learning (RL) is a key tool in the development of model-free feedback controllers due to its self-learning capabilities. RL allows the controller to autonomously learn the optimal control policy through continuous interaction with the environment while achieving predefined objectives. As a result, a variety of RL-based control algorithms have been proposed for power systems in recent years. For example, [7, 8] introduces a frequency control algorithm aimed at minimizing both the frequency nadir and the required control inputs. In [9], a voltage control strategy is developed using deep RL, where the RL algorithm minimizes voltage deviations across buses by actively adjusting generator bus voltages. Additionally, various RL-based automatic generation control (AGC) schemes have been proposed to regulate system frequency during transients. These include approaches such as Q-learning, actor-critic-based integral RL, and policy gradient (PG) techniques, as seen in [10], [11], and [12]. Similarly, reactive power control methodologies designed to improve system voltages using different RL

approaches are discussed in [13] and [14]. For a comprehensive review of RL applications in power system control, recent surveys such as [15] and [16] provide in-depth insights.

However, it's important to note that much of the current literature simplifies power system models by (1) neglecting algebraic constraints (power flow and balance equations), (2) using lower-order models for synchronous machine dynamics, and (3) overlooking the complexities of power electronics-based models for solar and wind generation. Moreover, many studies focus exclusively on minimizing frequency and voltage deviations, without addressing the critical issue of damping system-wide low-frequency and ultra-low-frequency oscillations (LFOs/ULFOs). This omission stems from not solving the complete optimal feedback control problem, such as the traditional linear-quadratic regulator (LQR)-type control design. As previously mentioned, addressing LFOs and ULFOs is essential for maintaining grid stability and optimizing power transfer capabilities [1].

Some recent efforts have been made to design various decentralized and centralized damping feedback controllers using RL. For example, in [17] using a reduced power system model a wide-area damping controller (WADC) has been proposed to improve system oscillations after disturbance. Similarly, in [18] authors have developed a WADC using a policy iteration algorithm, while the study [19] has introduced a decentralized feedback controller design via an off-policy iteration-based RL technique. However, these studies also rely on simplified power system models, considering lower-order generator dynamics and neglecting algebraic constraints, dynamics of loads, wind, and solar power plants. Using a simplified power system model during training can cause serious stability issues as the learned policy might be unstable when applied to the actual power system model with detailed dynamics.

Furthermore, it is clear that designing feedback controllers using a model-free approach via RL has a key advantage as it does not require the knowledge of system parameters, dynamics, or topology. However, some basic questions arise while solving an optimal state feedback control problem for renewables heavy power systems: How computationally efficient is model-free control as compared to model-based design? Does the model-free approach yield a *better* control law as compared to the model-based approach? To that end, this paper addresses the aforementioned literature gaps and research questions by designing model-free and model-based WADCs for renewables heavy power grids with detailed wind, solar, and composite load dynamics (given in Sec. 2). The problem formulation is given in Sec. 3. The proposed model-free WADC is based on reinforcement learning (Sec. 4) while the model-based controller is designed via Lyapunov control theory (Sec. 5), and Sec. 6 presents thorough case studies and some conclusions.

The technical contributions of this paper are as follows:

- We propose WADC for renewables heavy power systems (with majority of power being generated by wind and solar-based power plants). The considered power systems comprise of conventional power plant modeled via a detailed nonlinear 9th-order dynamical model (model-ing synchronous machine, exciter, and turbine dynamics), comprehensive power electronics based wind/solar power plants models, and composite load dynamics (constant power, constant impedance, and motor loads). The proposed WADCs act as a secondary control layer and are directly actuated via the primary controller layer (by sending additional control signals) of the power network.

- Two different approaches have been utilized to solve the same WADC problem for the considered power system. One is based on a completely model-free approach via reinforcement learning while the other methodology utilizes a model-based design. In particular, for the model-free approach, we leverage deep deterministic policy gradient (DDPG)-based algorithm to learn optimal control law by continuously interacting with the power system model. While for the model-based design, we use linear-matrix inequalities (LMIs)-based $\mathcal{H}_2$ stability notion along with Lyapunov control theory. Also, to improve computational efficiency in the model-free approach, we use control-theoretic techniques to properly design and initialize the corresponding actor and critic neural networks as explained in Sec. 4.

Thorough numerical simulations and discussions have been provided on modified IEEE 9-bus and 39-bus systems showcasing the different pros and cons of model-based and model-free WADC designs. Furthermore, to showcase the advantages of the proposed WADCs a comparative analysis has also been carried out by studying the transient response of the system with only primary controllers and with WADCs acting on top of them.

Notations: Matrices and vectors are written in bold, while sets are represented using calligraphic fonts, such as $\mathcal{G}$ or $\mathcal{W}$. The notation $\mathbb{R}^b$ refers to the set of real-valued column vectors with 'b' elements, and $\mathbb{R}^{c \times d}$ denotes the set of real-valued matrices with 'c' rows and 'd' columns. The zero matrix is denoted by $\boldsymbol{O}$, and the identity matrix of appropriate size by $\boldsymbol{I}$. The union of two sets is represented by $\cup$, and the Kronecker product is indicated by $\otimes$. Additionally, $\mathbb{S}_{++}^{c \times d}$ refers to the set of positive definite matrices of size 'c' by 'd'. The asterisk $*$ in a symmetric matrix indicates that the entries are symmetric with respect to the main diagonal. All quantities are given in per unit (p.u.) unless otherwise noted. For simplicity, the time dependence of vectors is sometimes omitted in equations, e.g., $\boldsymbol{x}_d(t)$ is written as $\boldsymbol{x}_d$.

2. Renewables heavy power system with solar farms, wind farms, and composite loads

We consider a power system model consisting of G traditional power plants (both steam and hydro-based), S solar power plants, W wind power plants, and various loads: L_k motor-based loads, L_p constant impedance loads, and L_z constant power loads. The electrical grid is represented as a graph, where $\mathcal{E}$ denotes the set of transmission lines and $\mathcal{N} = \{1, ..., N\}$ represents the set of buses. These buses are categorized into different types: $\mathcal{G} = \{1, ..., G\}$ corresponds to buses connected

to traditional power plants, $\mathcal{S} = \{1, ..., S\}$ to buses with solar farms, and $\mathcal{W} = \{1, ..., W\}$ to buses connected to wind-based power plants. The set $\mathcal{L}$ represents buses connected to motor-based, constant impedance, and constant power loads, while $\mathcal{U}$ includes non-unit buses that are not connected to any generation or load elements.

With this setup, the power system is described by a set of nonlinear differential-algebraic equations (NDAEs) as follows[20]:

$$\dot{x}(t) = f(x_d, x_a, w, u) \tag{1a}$$
$$0 = h(x_d, x_a, w, u). \tag{1b}$$

In the above equations, differential equations (1a) describe the models of traditional power plants, solar farms, wind power plants, and dynamics of composite loads while the algebraic equations (1b) model the algebraic constraints (the power/current balance equations) in the network. The notation $x_a(t) \in \mathbb{R}^{n_a}$ represents algebraic variables, $x_d(t) \in \mathbb{R}^{n_d}$ denotes dynamic variables, $u(t) \in \mathbb{R}^{n_u}$ contains the control inputs, and $w(t) \in \mathbb{R}^{n_w}$ denotes system disturbances. The detailed explanations of these vectors are given in Appendix A. For brevity, the complete dynamical equations (set of ordinary differential equations) describing the models of generators, solar, and wind power plants are not included in this paper and can be found in [20–23].

That being said, we can also express the power system model (1) in the following state-space format:

$$E\dot{x} = \bar{A}x + \bar{B}_u u + f(x, w, u) + \bar{B}_w w \tag{2}$$

where $x(t) = \begin{bmatrix} x_d^\top & x_a^\top \end{bmatrix}^\top \in \mathbb{R}^n$ represents the overall state vector and $E \in \mathbb{R}^{n \times n}$ is a singular matrix that encodes algebraic constraints, characterized by rows of zeros. The constant matrices $\bar{A} \in \mathbb{R}^{n \times n}$, $\bar{B}_u \in \mathbb{R}^{n \times n_u}$, $\bar{B}_w \in \mathbb{R}^{n \times n_w}$ map the state vector, control inputs, and the disturbance vector w in the system dynamics. These matrices are determined by capturing the linear components of the model (1) while the function $f(x, w, u)$ accounts for any linearization errors present in the power system model.

3. Preliminaries and problem formulation

In this section, we address the state feedback control problem for the renewables-heavy interconnected power system model (3). We first define the overall state feedback control problem and then propose two distinct solutions: one using a model-free approach via reinforcement learning, and the other using a model-based approach based on system matrix information and approximations of the nonlinear function $f(x, w, u)$, employing Lyapunov stability theory.

In the designed system (3), the dynamic states x_d—which represent conventional generators, renewable energy sources, and motor-based loads—are treated as dynamic components, while loads, non-unit buses, and other interconnections are considered static. For the purpose of designing a feedback controller, the algebraic state vector x_a, which includes voltage and current phasors, can be regarded as redundant and thus eliminated [24]. Given that power system models are typically index-1 DAEs, these algebraic variables can be removed, allowing the

conversion of the DAE into an equivalent ODE as follows: Let us consider

$$\bar{A} = \begin{bmatrix} A_{dd} & A_{da} \\ A_{ad} & A_{aa} \end{bmatrix}, \ \bar{B}_u = \begin{bmatrix} B_{ud}^\top & B_{ua}^\top \end{bmatrix}^\top, \ \bar{B}_w = \begin{bmatrix} B_{wd}^\top & B_{wa}^\top \end{bmatrix}^\top$$

and assume that A_{aa} is invertible (a common assumption in the literature of power systems). Then, we can extract the equation for x_a and substitute it in the dynamic states equation to get the following ODE system:

$$\dot{x}_d = Ax_d + B_u u + f_d(x_d, w, u) + B_w w \tag{3}$$

where $f_d(.)$ represents the corresponding nonlinearity and the rest of the matrices are given as:

$$A = A_{dd} - A_{da}A_{aa}^{-1}A_{ad}, \quad B_w = B_{wd} - A_{da}A_{aa}^{-1}B_{wa}$$
$$B_u = B_{ud} - A_{da}A_{aa}^{-1}B_{ua}.$$

That being said, to set up the state feedback control problem in the control literature, one first needs to design the perturbed closed-loop system dynamics. For this purpose, for a dispatch time period $kT \leq t < (k+1)T$, let us consider a control policy given as:

$$\boxed{\textbf{Control policy:} \ \ u_{cl}(t) = u_0^k - K\left(x_d(t) - x_d^k\right)} \tag{4}$$

where x_d^k is the state equilibrium value before the occurrence of disturbances, u_0^k is the set-point of the input u which is determined for every k^{th}-dispatch time-period by running power flow (PF), and $K \in \mathbb{R}^{n_u \times n_d}$ is a gain matrix (a design parameter). Then, the corresponding closed-loop system can be written as:

$$\dot{x}_d = Ax_d + B_u u_{cl} + f_d(x_d, u_{cl}, w) + B_w w \tag{5}$$

Now, let us assume there is an unknown disturbance in the power system. This disturbance will push the system to a new equilibrium state, let us denote that by x_d'. Then, the closed-loop system at this new equilibrium can be written as:

$$0 = Ax_d' + B_u u_{cl}' + f_d(x_d', u_{cl}', w') + B_w w'. \tag{6}$$

Then, the perturbed closed-loop dynamics can be expressed as follows (which is essentially computed by subtracting (6) from (5)):

$$\Delta\dot{x}_d = (A - B_u K)\Delta x_d + \Delta f_d(\Delta x_d, \Delta u_{cl}, \Delta w) + B_w \Delta w \tag{7}$$

with $\Delta x_d = x_d - x_d'$, $\Delta w = w - w'$, $\Delta f_d(\Delta x_d, \Delta u_{cl}, \Delta w) = f_d(x_d, u_{cl}, w) - f_d(x_d', u_{cl}', w')$. For simplicity, from now onward, with little abuse of notation, we drop the Δ notation from Δx_d, Δf_d, and Δw and simply use f_d, x_d, and w instead, respectively. Thus the perturbed closed-loop system (7) is rewritten as follows:

$$\dot{x}_d = (A - B_u K)x_d + f_d(x_d, u_{cl}, w) + B_w w. \tag{8}$$

That being said, in state feedback control literature the primary goal is to design a control law (or control gain matrix K)

3

with two main objectives. Firstly, the control law should converge the perturbed closed-loop system asymptotically to zero. This implies that the control policy (4) endeavors to restore the power system (3) to its steady-state equilibrium following a significant disturbance. Secondly, the designed control law should achieve this objective with the minimum effort required. This ensures that the control strategy minimizes the resources and inputs required to achieve the desired stabilization. We want to emphasize here that, in the end, the control policy u_{cl} needs to be applied to the complete power system model (3), then, it can simply be mapped back by computing $\hat{K} = \begin{bmatrix} K & O \end{bmatrix} \in \mathbb{R}^{n_u \times n}$ and redesigning the control policy as: $\hat{u}_{cl} = u_0^k - \hat{K}\left(x(t) - x^k\right)$. Now, since the dimensions are same, we can use $\hat{u}_{cl}$ as a control policy for the complete power system NDAE.

Before introducing the state feedback control problem, we outline a key assumption used throughout the paper:

Assumption 1. *The pair (A, B) is impulse controllable and finite dynamics stabilizable.*

Assumption 1 is standard in control theory literature [25, 26], and several power system models have been shown to satisfy controllability and stabilizability, as demonstrated in [24, 27].

The overall infinite-horizon optimal state feedback control problem can then be formulated as:

$$\min_{K \in \mathbb{R}^{n_u \times n_d}} \int_0^\infty (Cx_d + DKx_d)^\top (Cx_d + DKx_d)dt \qquad (9)$$

subject to   Dynamics (8)

where $C \in \mathbb{R}^{n_d \times n_d}$ and $D \in \mathbb{R}^{n_d \times n_u}$ are fixed penalty matrices, analogous to the weight matrices Q and R in traditional LQR control. The selection of C and D is driven by the grid operator's preferences, which determine the specific states or control inputs that should be penalized during the design of the controller gain matrix K.

To solve the above feedback control problem for the test system considered, we employ two distinct approaches. The first approach is a fully data-driven, RL-based technique, while the second is a model-based framework that leverages Lyapunov stability theory. Both approaches are discussed in detail in the following sections.

4. Model-free optimal state feedback controller design

Here, we propose a model-free approach to solve the optimal state feedback control problem given in Eq. (9). The presented approach is based on deep reinforcement learning in which the agent learns optimal control policy (which essentially means learning the feedback controller gain matrix K) such that it maximizes cumulative rewards over time (the negative of quadratic cost function given in Eq. (9)) by interacting with the environment (the complete NDAE power system model). That being said, since (9) is a *continuous* action space infinite horizon problem, then deep-deterministic policy gradient based RL algorithms are well suited for such optimization problem [28]. An overview of DDPG-based RL is given below

followed by the proposed methodology which properly shapes (using knowledge of state feedback control theory) the DDPG-based technique to efficiently solve the formulated state feedback control problem in (9).

4.1. DDPG for state feedback control in power systems

DDPG employs an actor-critic architecture, where the actor learns a policy for selecting actions, and the critic evaluates the value of those actions. Both the actor and critic functions are approximated using neural networks, let us denote them by $\pi_\theta(x_d) : \mathbb{R}^{n_d} \rightarrow \mathbb{R}^{n_u}$ and $Q_\phi(x_d, u_\pi) : \mathbb{R}^{n_d} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}$, respectively, where u_π is the output of the actor-network. The actor neural network is parameterized by θ and it takes states as inputs and returns action as output while for the critic network weights/parameters are denoted by ϕ and it takes the pair states-actions as input and returns the Q-value (the cumulative long term reward for the state-action pair) as output.

That being said, generally speaking, the main idea in DDPG is to run gradient descent on the actor-network parameter θ as: $\theta \leftarrow \theta - \epsilon \nabla_\theta J(\theta)$, where ϵ represents the step size and $\nabla_\theta J(\theta)$ denote the policy gradient which in DDPG is approximated as follows [28]:

$$\nabla_\theta J(\theta) \approx \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \nabla_u Q_\phi(.)\big|_{x_d[i], \pi_\theta(x_d[i])} \nabla_\theta \pi_\theta(x_d[i]) \qquad (10)$$

where $x_d[i], u[i]$ with $i \in \mathcal{B}$ are set of samples extracted from the replay buffer which stores the history of observations (x_d, u, r, x_d^n), here, r is the reward and x_d^n is the new state after taking action u. In (10) the critic network $Q_\phi(x_d, u)$ is learned via temporal difference learning given as [28]:

$$\min_\phi L(\phi) = \mathbb{E}[Q_\phi(x_d, u) - (r + \gamma Q_{\phi_t}(x_d^n, \pi_{\theta_t}(x_d^n)))^2] \qquad (11)$$

where $\mathbb{E}$ denotes the expectation notation. In (11) π_{θ_t}, Q_{ϕ_t} represent the target actor and critic networks whose weights θ_t, ϕ_t are computed using θ, ϕ via Polyak averaging as follows:

$$\phi_t \leftarrow \rho\phi_t + (1 - \rho)\phi \qquad (12a)$$
$$\theta_t \leftarrow \rho\theta_t + (1 - \rho)\theta \qquad (12b)$$

where ρ is a positive constant usually selected close to 1. The target networks are essentially copies of the original networks which trail behind the original networks and are updated slowly using the original network parameters (in DDPG this is usually done to improve the training stability). Further detailed explanations about DDPG can be seen in [28].

With that in mind, to efficiently solve the state feedback control problem (9) using the DDPG-based RL technique, we propose to do the following: Firstly, notice that, the actor neural network tries to approximate the feedback controller gain matrix K. For the perturbed closed-loop system, the control law implemented by the actor is essentially linear mapping $u_\pi = Kx_d = k_1 x_{d_1} + k_2 x_{d_2} + ...$ where $k_1, k_2 + ...$ are the weights of the actor neural network. Then, a shallow neural network for the actor has been designed with an input layer and a fully connected layer to provide a linear mapping between states to

actions. Notice that to reduce the number of learnable parameters (and thus increasing the convergence rate) no `relu` layer is added as no nonlinear mapping is needed to train the actor. Also, it is well known from the control theory that the closed-loop system is stable if these gains are negative, therefore, initializing them to take negative values can speed up the convergence. Moreover, since the learned control law approximated by the actor neural network does not have any extra biasing (constant coefficients such as $\boldsymbol{u}_\pi = \boldsymbol{K}\boldsymbol{x}_d + \boldsymbol{b}$), then while training, the actor neural network weights are only updated and the biasing learn rate is set to be zero.

Now, as discussed earlier, the critic learns the Q-value function. The critic accepts an observation-action pair as inputs and returns a scalar (the discounted long-term reward) as output. To that end, from feedback control theory, we know that the long-term reward function (or the value function) for (8) is known to be quadratic (the Lyapunov function $\boldsymbol{x}_d^\top \boldsymbol{P} \boldsymbol{x}_d$ with $\boldsymbol{P} > 0$, which tells us the optimal cost-to-go). Thus, the critic network is designed to have a quadratic layer (which returns a vector of quadratic monomials) followed by a fully connected layer (providing a linear mapping of its inputs). Furthermore, in the context of our setting, we can essentially express the structure of the Q-value function as follows:

$$\boldsymbol{Q}(\boldsymbol{x}_d, \boldsymbol{u}_\pi) = l_1 x_{d_1}^2 + l_2 x_{d_1} x_{d_2} + l_3 x_{d_2}^2 \cdots l_{n_d} x_{d_1} u_1 + \cdots \quad (13)$$

where $l_1, l_2, \cdots$ are the weights of the critic neural network, or alternatively Eq. (13) in matrix form can be written as:

$$\boldsymbol{Q}(\boldsymbol{x}_d, \boldsymbol{u}_\pi) = \begin{bmatrix} x_{d_1} & x_{d_2} \cdots u_1 \cdots \end{bmatrix} \begin{bmatrix} l_1 & \frac{l_2}{2} & \frac{l_4}{2} \cdots \\ \frac{l_2}{2} & l_3 & \frac{l_5}{2} \cdots \\ \frac{l_4}{2} & \frac{l_5}{2} & l_6 \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} x_{d_1} \\ x_{d_2} \\ \vdots \\ u_1 \\ \vdots \end{bmatrix} \quad (14)$$

$$\iff \begin{bmatrix} \boldsymbol{x}_d^\top & \boldsymbol{u}_\pi^\top \end{bmatrix} \boldsymbol{L} \begin{bmatrix} \boldsymbol{x}_d^\top & \boldsymbol{u}_\pi^\top \end{bmatrix}^\top$$

Since $\boldsymbol{u}_\pi = \boldsymbol{K}\boldsymbol{x}_d$, then, one can rewrite (14) as:

$$\boldsymbol{Q}(\boldsymbol{x}_d, \boldsymbol{K}\boldsymbol{x}_d) = \boldsymbol{x}_d^\top \begin{bmatrix} \boldsymbol{I} & \boldsymbol{K}^\top \end{bmatrix} \boldsymbol{L} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{K}^\top \end{bmatrix}^\top \boldsymbol{x}_d = \boldsymbol{x}_d^\top \boldsymbol{P} \boldsymbol{x}_d \quad (15)$$

Now, in our case, since we are maximizing the negative of the quadratic cost function, both $\boldsymbol{P}$ and $\boldsymbol{L}$ should be negative definite. Thus, initializing the critic network to be a negative definite matrix can overall stabilize the learning and speed up the convergence.

Moving on, in DDPG for every episode, the actor generates a random action, applies it to the environment, and tries to maximize the cumulative long-term reward. However, for the power system model, any random unbounded action cannot be chosen as it will destabilize the system and the episode might not even start. Thus, for each episode, the random actions are selected in a bounded region with $\boldsymbol{u}_{max}$ and $\boldsymbol{u}_{min}$ as upper and lower bounds, respectively. For the considered test system the control actions are voltage and power/valve position set points as discussed in Sec. 2. Then, voltage setpoints are constrained between ±5% while the power/valve position setpoints are selected to be between ±10%. This is reasonable as in normal operation the voltage should be between $0.95 - 1.05$pu and also the

power output from each power generator usually changes proportionally in response to large disturbances [24]. Furthermore, since the output of the actor-network is bounded between $\boldsymbol{u}_{max}$ and $\boldsymbol{u}_{min}$, then during training to avoid the actor output being saturated frequently a `tanh` (which scales the output between -1 and 1) and scaling layer (which scales back the output to their desired range) have been added. That being said, the overall proposed model-free approach to solving the state feedback control problem is given in Algorithm 1.

Algorithm 1: Model-free state feedback controller

1 Design actor $\boldsymbol{\pi}_\theta$ and critic $\boldsymbol{Q}_\phi$ networks and initialize them as discussed in Sec 4.
2 Initialize ρ and target actor $\boldsymbol{\pi}_{\theta_t}$ and critic $\boldsymbol{Q}_{\phi_t}$ networks.
3 Initialize $\boldsymbol{C}$, $\boldsymbol{D}$ matrices and control input bounds $\boldsymbol{u}_{max}$, $\boldsymbol{u}_{min}$.
4 Initialize replay buffer $\mathcal{R}$.
5 **for** *episode = 1 to M* **do**
6   Initialize power system model with a random initial condition $\boldsymbol{x}_0$ having 5% maximum deviation from steady-state values.
7   **for** *iteration = 1 to J* **do**
8   Select input $\boldsymbol{u}_\pi$ from the actor network.
9   Execute $\boldsymbol{u}_\pi$ on system (8) using $\boldsymbol{x}_0$ and observe quadratic cost $\boldsymbol{r}$ and next state $\boldsymbol{x}_d^n$.
10   Store observation in $(\boldsymbol{x}_d, \boldsymbol{u}_\pi, \boldsymbol{r}, \boldsymbol{x}_d^n)$ in $\mathcal{R}$.
11   Sample mini-batch $\mathcal{B}$ of observation from $\mathcal{R}$.
12   **for** *observation $b \in \mathcal{B}$* **do**
13   Compute: $y_b = \boldsymbol{r}(\boldsymbol{x}_d, \boldsymbol{u}) + \gamma \boldsymbol{Q}_{\phi_t}(\boldsymbol{x}_d^n, \boldsymbol{\pi}_{\theta_t}(\boldsymbol{x}_d^n))$.
14   Update the critic network by minimizing the loss as given in (11).
15   Update the actor network via gradient descent and with policy gradient $\nabla \boldsymbol{J}(\boldsymbol{\theta})$ computed using Eq. (10).
16   Update target actor and crtic network through polyak averaging using Eq. (12).
17 Return trained actor $\boldsymbol{\pi}_\theta$ and critic $\boldsymbol{Q}_\phi$ networks.
18 Extract controller gain $\boldsymbol{K}$ from $\boldsymbol{\pi}_\theta$.
19 Design $\hat{\boldsymbol{K}}$, plug it in the control policy $\hat{\boldsymbol{u}}_{cl}$, and apply it to the complete NDAE power system (3) with $\boldsymbol{x}^k$ and $\boldsymbol{u}_0^k$ being updated every OPF/PF dispatch time period $kT \leq t < (k + 1)T$.

---

5. Model-based optimal state feedback controller design

To solve the formulated optimal state feedback control problem (9), here, we develop a completely model-based approach (without relying on any data or interaction with the power system environment). The proposed technique in this section is based on the Lyapunov stability theory and it consider information of the constant system matrices $(\boldsymbol{A}, \boldsymbol{B}_u, \boldsymbol{B}_w)$ along with approximation on nonlinear function $\boldsymbol{f}_d(\boldsymbol{x}_d, \boldsymbol{u}, \boldsymbol{w})$ to compute the controller gain matrix $\boldsymbol{K}$.

That being said, to solve the optimal feedback control problem (9), we present the following results:

Proposition 1. *Suppose Assumption 1 hold. Then, there exists a solution to problem* (9) *(meaning the perturbed closed-loop dynamics asymptotically converges to zero with minimum control effort required), if there exist matrices* $Z \in \mathbb{S}_{++}^{n_d \times n_d}$, $W \in \mathbb{R}^{n_d \times n_d}$, *and* $F \in \mathbb{R}^{n_u \times n_d}$ *that are solution to the following semi-definite program (SDP)*

$$(\text{SDP}) \quad \min_{F, Z, W} \quad Tr(W)$$

$$\text{subject to} \quad \text{LMI (16),} \quad Z > O \begin{bmatrix} Z & * \\ \hat{B}_w^\top & W \end{bmatrix} \geq 0$$

where LMI (16) *is as follows:*

$$\begin{bmatrix} Z^\top A^\top + AZ - F^\top B_u^\top - B_u F & * & * \\ \hat{B}_w^\top & O & * \\ CZ + DF & O & -I \end{bmatrix} \prec O \quad (16)$$

Upon solving the above SDP the controller policy can be computed as $K = FZ^{-1}$.

Proof. Before presenting the proof, to formulate a tractable convex LMI-based formulation for the model-based controller design, we assume that the perturbation in the nonlinearity in Eq. (8) is $\mathcal{L}_2$-norm bounded and can be expressed as $\Delta f_d(x_d, u_{cl}, w) = B_{fd} w_{fd}$ with $B_{fd} = B_w$. We want to emphasize here that, the above assumption on $\Delta f_d(x_d, u_{cl}, w)$ is only carried out to design tractable convex SDP formulation for the controller design. At the end, the final designed feedback controller is applied to the complete NDAE power network without any simplifications. Now, we define:

$$\hat{w} = \begin{bmatrix} w^\top & w_{fd}^\top \end{bmatrix}^\top, \ \hat{B}_w = \begin{bmatrix} B_w & B_{fd} \end{bmatrix}. \quad (17)$$

With these definitions in place, to begin the proof of Proposition 1, we consider a candidate Lyapunov function for the perturbed system (8) as $V(x_d) = x_d^\top P x_d$ where $V : \mathbb{R}^{n_d} \to \mathbb{R}_+$ and $P \in \mathbb{R}^{n_d \times n_d}$. Then, the derivative of $V(x_d)$ with respect to x_d along the trajectories of (8) can be expressed as

$$\dot{V}(x_d) = (P x_d)^\top \dot{x}_d + \dot{x}_d^\top (P x_d).$$

Now, for the overall system stability including the objective function (as given in (9)), we need $\dot{V}(x_d) + x_d^\top \Gamma^\top \Gamma x_d < 0$ where $\Gamma = C - DK$, which can expanded as:

$$(P x_d)^\top (A_C x_d + \hat{B}_w \hat{w}) + x_d^\top \Gamma^\top \Gamma x_d + (A_C x_d + \hat{B}_w \hat{w})^\top P x_d < 0$$

where $A_C = A - B_u K$. We can rewrite the above equation also as $\Psi^\top \Xi \Psi < 0$ with:

$$\Psi = \begin{bmatrix} x \\ \hat{w} \end{bmatrix}^\top, \ \Xi = \begin{bmatrix} A_C^\top P + P^\top A_C + \Gamma^\top \Gamma & P^\top \hat{B}_w \\ \hat{B}_w^\top P & O \end{bmatrix} \quad (18)$$

Notice, $\Psi^\top \Xi \Psi < 0$ holds if and only if $\Xi \preceq O$.

Now, let us define $Z := P^{-1}$ and pre-multiply and post-multiply (18) with $\text{diag}\left(\begin{bmatrix} Z^\top & I \end{bmatrix}\right)$ and $\text{diag}\left(\begin{bmatrix} Z & I \end{bmatrix}\right)$, respectively, to get the following equivalent representation of (18):

$$\begin{bmatrix} Z^\top A_C^\top + A_C Z + Z^\top \Gamma^\top \Gamma Z & \hat{B}_w \\ \hat{B}_w^\top & O \end{bmatrix} \prec O$$

Then, applying the Schur complement lemma [29], the above LMI can equivalently be represented as:

$$\begin{bmatrix} Z^\top A_C^\top + A_C Z & \hat{B}_w & Z^\top \Gamma^\top \\ \hat{B}_w^\top & O & O \\ \Gamma Z & O & -I \end{bmatrix} \prec O \quad (19)$$

Now, defining $F := KZ$, then we get the convex LMI (16) which represents the necessary and sufficient stability conditions for the existence of the controller policy. Finally, to minimize the impact of w on the system dynamics according to $\mathcal{H}_2$ notion, we need to minimize $Tr(\hat{B}_w^\top Z \hat{B}_w)$ where Tr denotes the Trace notation. Now, let us upperbound $\hat{B}_w^\top Z \hat{B}_w$ it by matrix variable W as $\hat{B}_w^\top Z \hat{B}_w \leq W$. Then, by minimizing $Tr(W)$ and taking Schur compliment of $\hat{B}_w^\top Z \hat{B}_w < W$ we get the second LMI along with the objective function in the formulated SDP. This completes the proof. ∎

Proposition 1 presents a completely model-based approach to solve the formulated optimal state feedback control problem (9). The computed gain matrix K guarantees the asymptotic stability of the perturbed system (8). In other words, it ensures that after a large disturbance, the NDAE system model (5) converges back to its steady values (while adding damping to the system oscillation) with minimum control effort required.

We want to emphasize here that, in the end, both proposed approaches provide us the feedback gain matrix K via solving exactly the same state feedback control problem given in (9). In the model-free approach, the controller tries to learn K by continuously interacting with the power system model and observing the quadratic cost/reward given in (9) while in the model-base design, it requires an accurate system model to compute K. As renewable energy integration and the widespread adoption of distributed energy resources continue to evolve, the accurate modeling of power systems becomes increasingly challenging. Thus, the model-free approach has a major advantage over model-based design as it does not require any system information and modeling.

In the following section, we implement both presented approaches and do a thorough comparison on a renewables heavy power system model with detailed synchronous generator, wind, solar, and composite loads (constant power, constant impedance, and motor-based loads) dynamics as discussed in Sec. 2.

6. Case studies

Both the proposed model-based and model-free controllers have been tested on IEEE 9-bus and 39-bus systems. Both these power systems have been modified to include composite loads, and high penetration of wind, solar-based renewable resources. The 9-bus system includes one steam-based traditional power plant at Bus 1, one wind-based power generator at Bus 3, a solar farm at Bus 2, and a motor-based load at Bus 8. Similarly, the 39-bus system is comprised of four traditional power plants, four solar power plants, two wind-based power plants, and a motor-based load at Bus 14. The one-line diagrams of both power systems are presented in Appendix A, Figs. A.7,

A.8. Further details about modeling and parameters of traditional power plants, wind/solar farms, and loads of the systems used in this work are given in [21, 23, 30, 31].

All the case studies are performed on a computer with Intel $i7$ processor and 64GB RAM. The power systems have been modeled in MATLAB 2021a and are simulated using `ode15s` MATLAB DAE solver. The initial conditions for the power systems are computed using power flow studies via MATPOWER [32] `runpf` function. The baseline for the frequency and volt-ampere of the power network is set to be $w_b = 120\pi$rad/s, $S_b = 100$MVA, respectively. Also, it's worth mentioning that in both the 9-bus and 39-bus systems the total shares of power production from renewables are around 63% and 67%, respectively.

To design the actor and critic networks and to implement the proposed DDPG-based model-free feedback controller, MATLAB RL-toolbox is utilized. During learning for the 9-bus system the agent interacts with the power system model for a sample time of $T_s = 0.1s$ and a total of 200 samples are taken in each episode while the total number of episodes is set to be 4000. The rest of the hyper-parameters for both 9-bus and 39-bus systems used during the learning process are given in Tab. 1. These hyper-parameters are essential and need to be properly tuned to have stable training and to get optimal control policy. Although there is no systematic way to tune these parameters, they can be determined based on human operator's knowledge or trials and errors. One can play with these parameters to get better performance. In Tab. 1, we set the critic learning rate to be larger as compared to the actor learning rate. This seems to work well in our case, the intuitive reasoning behind this can be that since the critic takes the action and observation and tells how good the taken action is, thus pushing the critic to learn quickly as compared to the actor seems to stabilize the overall learning.

In every episode, the initial conditions for the system are randomly chosen with 5% maximum deviation from steady-state values. Notice that, while learning in every episode, if for any particular state-action pair the `ode15s` solver does not converge to a solution (meaning the power system is not simulated) then no learning is performed and the episode is terminated. Regarding training, the overall learning times for both test systems are given in Tab. 2 while the cumulative long-term rewards (the negative of the quadratic cost function) for both test networks are presented in Fig. 1. We can see that for the 9-bus test system the algorithm converges to optimal policy in around 4000 episodes while for the 39-bus test system, it took almost 5000 episodes to converge and learn the optimal control policy.

On the other hand, the model-based controller is implemented in YALMIP [33] with MOSEK [34] as an optimization solver to solve the proposed SDP given in Proposition 1. Notice that, in both types of designs, the proposed controller acts in realtime as a secondary control loop and is directly actuated via the primary controllers (using their voltage and power/valve-position set points).

Figure 1: Average and episodic rewards for 9-bus (left) and 39-bus (right) test systems.
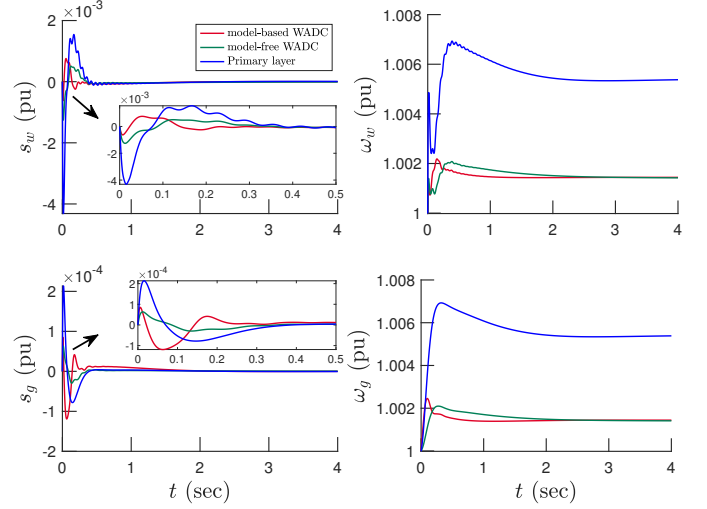

Figure 2: Comparative analysis under $\Delta_L = -0.5$ for 9-bus system: relative slip and angular speed of wind power plant (above), slip and rotor frequency of synchronous machine (below).
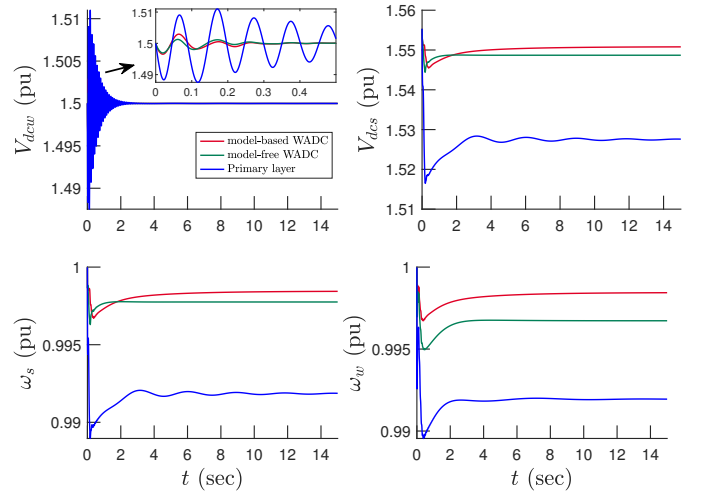

Figure 3: Comparative analysis under $\Delta_L = 0.7$ for 9-bus system: DC link voltage of the wind and solar plant (above), relative speed of wind and solar plants (below).
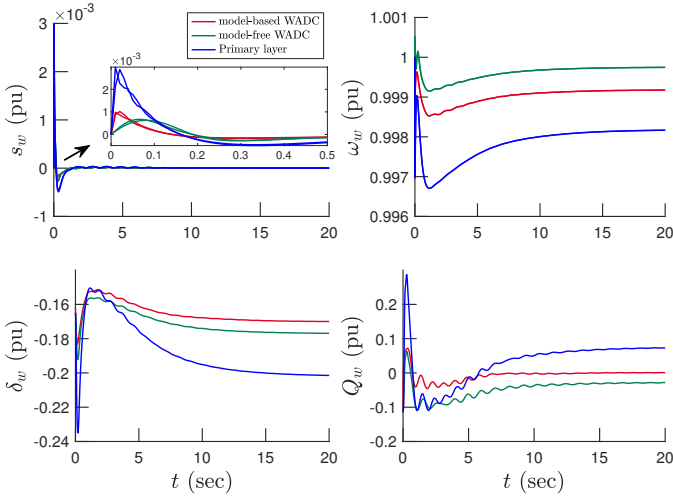
Figure 4: Comparative analysis under $\Delta_L = 0.003$ for 39-bus test system: relative slip and angular speed of all wind power plants (above), relative angle and reactive power output of wind plant connected at Bus 32 (below).
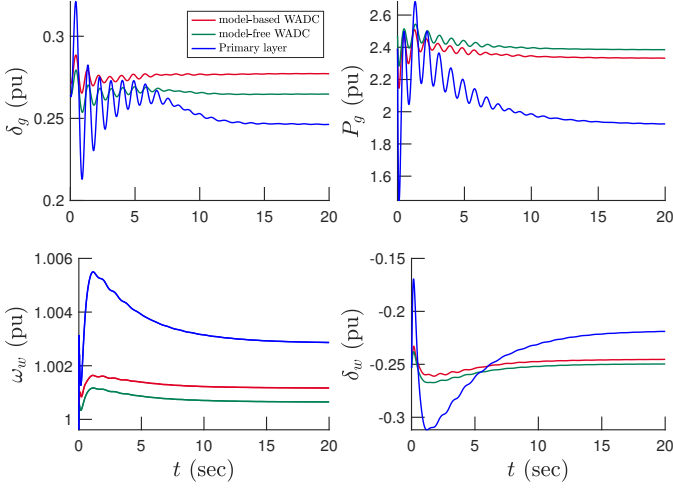

Figure 5: Comparative analysis under $\Delta_L = -0.001$ and $\Delta_{I_s} = 0.1$ for 39-bus test system: rotor angle and real power output of Generator at Bus 38 (above), relative speed of all wind power plants and relative angle of wind plant connected at Bus 34 (below).

Table 1: Parameters used in model-free state feedback controller design.

Parameter	Value, 9-bus	Value, 39-bus
Sample time, T_s	$0.1s$	$0.3s$
Replay buffer length, $\mathcal{R}$	10^6	10^8
Mini batch size, $\mathcal{B}$	10^1	10^2
actor learning rate	10^{-4}	10^{-4}
critic learning rate	10^{-1}	10^{-1}
Max episodes, M	4000	5000
Max steps per episode	200	300

6.1. Comparative analysis under uncertainty in load demand

In this section, we do a thorough comparative analysis of the proposed model-based and model-free state feedback controllers under abrupt disturbances in load demand. To that end, we carried out the numerical simulations as follows: Initially, the power network operates under steady-state conditions, meaning the power generation is equal to the overall load demand and thus the system rests at an equilibrium. Then, right at the start of simulation, at $t > 0$, an abrupt change in over-

all load demand occurs as follows: $P_L^n = (1 + \Delta_L)P_L^0$, where Δ_L denotes the severity of the load disturbance and P_L^0 represents the initial load demand before the disturbance, which, for the considered test systems are; 0.77pu for the 9-bus test system and 19.8pu for 39-bus test network. With that in mind, two different simulation studies are conducted for both test systems. In the first study, we assume Δ_L to be positive—meaning the overall load demand has suddenly been increased, which in other words can roughly be interpreted as a generator-trip event. That being said, for the 9-bus test system, we select $\Delta_L = 0.7$ while for 39-bus network we select it to be $\Delta_L = 0.003$. For the second simulation, we assume Δ_L to be negative, meaning the overall load demand of the power network has suddenly been decreased, thus it can be seen as a load-trip event. With that in mind, for this simulation, we set Δ_L for the 9-bus and 39-bus test systems to be -0.5 and -0.001, respectively.

The above disturbance is going to push the power system to a new equilibrium from its initial steady-state or it might potentially destabilize the whole system. The objective of the proposed WADCs is to hedge against this disturbance and enhance system transient stability by damping LFOs/ULFOs, improve frequency nadir, and bring the system back to its nominal operating conditions. The results are presented in Figs. 2, 3, and 4. To demonstrate the effectiveness of the proposed WADCs, a comparison of the system transient response with only primary control layer and with developed WADCs acting on top of them has also been presented. Note that, by primary control layer, we refer to the conventional control mechanism of the power system, which in the considered test system for synchronous machines are; power system stabilizers (PSSs), governor, automatic voltage regulators (AVRs), and machine inertial response. Similarly, wind and solar power plants are acting in grid-forming mode with droop controllers and PI-type current and voltage regulators acting as their control mechanisms. Providing detailed descriptions and mathematical equations of synchronous machine, solar/wind models and their primary controllers is out of the scope of this study and readers are referred to [23, 35] for synchronous machines and [21, 31, 36] for wind and solar power plants.

To that end, for the 9-bus test system, we can see that with both the proposed damping controllers there are significant improvements in LFOs/ULFOs, this can be verified by looking at the plots of the synchronous machine and solar/wind slips given in Figs. 2 and 3. We can see that with only the primary control layer right after the load disturbance during transient period (the initial 2-3 seconds of the simulations) there are significant oscillations, while with the additional proposed secondary control layer, the oscillations have been damped out. Similar results have been achieved for the 39-bus test system as given in Fig. 4. Thus, with the proposed WADCs the overall power system transient stability has been improved.

6.2. Comparative analysis under renewable uncertainty

Here we thoroughly study the transient behavior of the considered test system with and without the proposed damping controllers under uncertainty in the renewables. To that end, the simulations here are carried out as follows: At the begin-

ning, the solar irradiance on all solar power plants is set to be 1000 W/m^2 (which is the standard solar irradiance), then, at $t > 0$ we decrease the irradiance of PV plant connected at Bus 1 for 9-bus system and PV plants at Buses 30, 37 for 39-bus test system as follows: $I_s^n = (1 - \Delta_{I_s})I_s^0$, where I_s^0 is the solar irradiance before the disturbance and I_s^n represent its new value, while Δ_{I_s} represent the severity of disturbance and is set to be 0.1 (meaning the irradiance is decreased by 10%).

To further add transients, the load disturbance from the previous section has also been kept intact and to further mimic realistic load uncertainty, we add some Gaussian noise to it also as: $P_L^n = (1 + \Delta_L)P_L^0 + \omega_L(t)$, where $\omega_L(t)$ denotes Gaussian noise with zero mean and standard deviation of $0.2\Delta_L$. Under these transient conditions, the system is initially stabilized with only the primary control layer and then the proposed damping controllers are also added on top of them.

The results are presented in Figs. 5 and 6. We can see that with both the proposed WADCs there is an improvement in system oscillations. This can be verified from the plots of the slips of all the power plants. Similarly, from Fig. 5 the oscillations in the power output of synchronous machines have also been damped out during transient periods. Note that, in Figs. 5 and 6, we compute the slip s, inverters relative angular speed ω, and DC-link voltages V_{dc} from the state vector using the following equations [21]:

$$\omega(t) = 1 - k_d(P(t) - P^*), \quad V_{dc}(t) = \sqrt{E_{dc}(t)}, \quad s(t) = (w_c - \omega(t))/w_c$$

where k_d is the droop constant of the power plant, P is real power output of the power plant, while w_c is the overall weighted-average system frequency and is defined as [21, 22]:

$$\omega_c = \frac{\sum_{i=1}^{G} \omega_g(i)H_g(i) + \sum_{j=1}^{S} \omega_s(j)H_s(j) + \sum_{l=1}^{W} \omega_w(l)H_w(l)}{\sum_{i=1}^{G} H_g(i) + \sum_{j=1}^{S} H_s(j) + \sum_{l=1}^{W} H_w(l)}$$

with H_g, H_s, and H_w representing the inertial constants of synchronous machines, solar, and wind power plant, respectively. That being said, we can also notice that for both test systems, there is a noticeable improvement in frequency nadir (the overall frequency dip or increase during the transient period). For 39-bus system, from Fig. 6 with only primary control layer after disturbance the synchronous machines frequency rises upto almost 1.006pu while with proposed WADCs it is limited only to around 1.0019pu thus improving the overall system frequency nadir. Similar results are achieved for the inverter-based resources, we can see that there is a significant improvement in the relative frequency nadir for both solar and wind power plants. These results are also corroborated from Tab. 3 where the overall weighted-average frequency deviation of the power system is presented, we can see that for both test systems there are significant improvements under all types of disturbances.

6.3. Discussion on model-based vs model-free feedback control for the future power grid

Notice that in both proposed approaches the main difference lies in the fact that in the case of model-free design, the controller only requires realtime information of state vector x, which these days can be obtained accurately as there exist
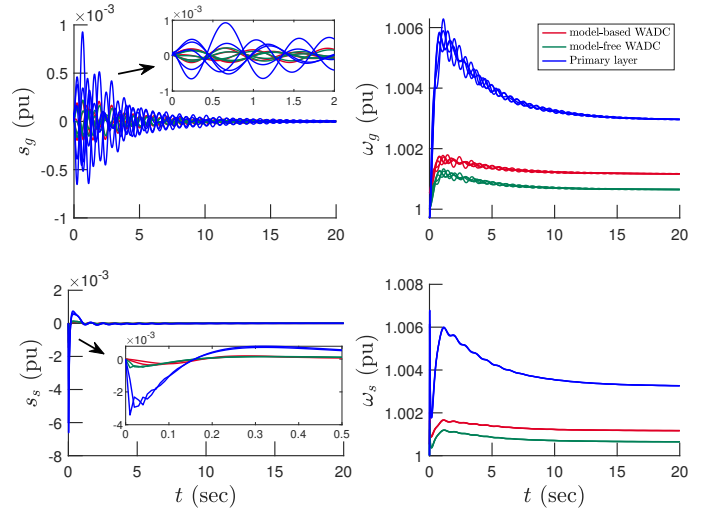

Figure 6: Comparative analysis under $\Delta_L = -0.003$ and $\Delta_{I_s} = 0.1$ for 39-bus test system: all synchronous generators slip and rotor frequency (above), slip and relative speed of all solar power plants (below).

highly robust state estimation algorithms. These estimation algorithms based on measurements received from a few PMUs can accurately estimate all the states including the states of solar/wind, composite loads, and synchronous machines [37]. While for the model-based approach, we not only need the information of state vector x but also accurate information about the system (matrices A, B_u, and so on) in realtime, which can be highly problematic for a larger power system and with high penetration of uncertain renewables. On the other hand, although the model-free approach seems reasonable for the future power grid as knowledge of the system is not required, however, the long training time, tuning of hyper-parameters, and no solid theoretical stability certificate/guarantee of the learned policy (since it merely solves an optimization problem and maximizes the long term reward) can be problematic. In the case of model-based approach, if there is a solution to Proposition 1, then, there is a solid theoretical guarantee (based on the Lyapunov stability notion) that obtained control policy will asymptotically stabilize the perturbed system (or in other, words push the original power system model back to its original steady state equilibrium).

Furthermore, notice that, according to the control theory literature, the model-based feedback control approaches can easily be extended to other more robust feedback control designs without adding significant complexities to the controller architecture, such as $\mathcal{H}_\infty$ or $\mathcal{L}_\infty$-type feedback controller design which has shown to have superior performance [38, 39]. While in the case of RL-based model-free designs, one can only maximize the long-term reward (since no information about the system is available) then it might be even much more harder to solve robust feedback controller ($\mathcal{H}_\infty$, $\mathcal{L}_\infty$ and so on) design as compared to their model-based counterparts.

Moreover, in the case of model-based designs, if the LMI-based formulations (which are used in this paper) become numerically unstable (or ill-conditioned, which they sometimes become depending on the input system matrices)—see [40], then, they can also be formulated as continuous-time alge-

Table 2: Comparison of Computational time between the proposed two approaches.

Network	Time required to compute the optimal control policy	
	model-based WADC	model-free WADC
9-bus test system	2.2 seconds	6.28 hrs
39-bus test system	53.2 seconds	8.32 hrs

Table 3: Comparative analysis of the deviation of overall weighted-average system frequency for the whole power system under various disturbances in load and renewable generations.

Network	Disturbance	$(\omega_0 - \omega_c)$		
		Primary	model-based	model-free
9-bus	$\Delta_L = 0.7$	0.015	0.004	0.0036
	$\Delta_L = -0.5$	-0.013	-0.0037	-0.0038
	$\Delta_L=0.6, \Delta_{I_s}=0.1$	0.019	0.0027	0.0029
39-bus	$\Delta_L = 0.003$	2.6×10^{-3}	1.3×10^{-5}	1.9×10^{-5}
	$\Delta_L=-0.001$	-1×10^{-3}	-2×10^{-5}	-2×10^{-5}
	$\Delta_L=0.04, \Delta_{I_s}=0.1$	7.9×10^{-3}	1.9×10^{-4}	2.1×10^{-4}

braic Riccati equations (CAREs) formats which can be solved highly efficiently and there exist well-built off-the-shelf software's/tools available to solve these types of feedback control problems easily [24, 41, 42]. In short, these are interesting future research avenues about how to efficiently and tractably solve (similar to the model-based approach) the optimal feedback (and robust optimal feedback) control problems via RL-based model-free approach for larger detailed renewables heavy power system models.

7. Concluding Remarks

In this paper, we solve the optimal feedback control problem for renewables heavy power systems with detailed solar, wind, composite load dynamics. Two distinct approaches are used: one is based on completely model-free design using DDPG-based reinforcement learning while the second methodology is based on model-based approach via Lyapunov and control theory. The designed WADCs are tested on 9-bus and 39-bus IEEE system, the results show that by adding additional feedback control loop the overall system transient stability after disturbance can significantly be improved. Moreover, thorough comparative analysis and discussion have also been carried out showcasing various pros and cons of using model-based vs model-free WADCs for future renewables heavy power grid.

The limitations of the presented work are as follows: Firstly, for the both proposed WADCs, the designed control policy is dense (requires all the power plant to take part in the control action) and not sparse. Secondly, both approaches does not take into account delays and/ or cyber-attacks in the communication network. Lastly, for the proposed model-free WADCs, the long training time and no solid theoretical stability guarantee (similar to its model-based counterpart which guarantees stability using Lyapunov criterion) are the drawbacks. Future work will

be about addressing the aforementioned limitations and designing stability-aware tractable model-free WADC.

References

[1] Pooja Gupta, Anamitra Pal, and Vijay Vittal. Coordinated wide-area damping control using deep neural networks and reinforcement learning. *IEEE Transactions on Power Systems*, 37(1):365–376, 2021.

[2] Yu Liu, Abhinav Kumar Singh, Junbo Zhao, A. P. Sakis Meliopoulos, Bikash Pal, Mohd Aifaa bin Mohd Ariff, Thierry Van Cutsem, Mevludin Glavic, Zhenyu Huang, Innocent Kamwa, Lamine Mili, Abdul Saleem Mir, Ahmad Taha, Vladimir Terzija, and Shenglong Yu. Dynamic state estimation for power system control and protection. *IEEE Transactions on Power Systems*, 36(6):5909–5921, 2021.

[3] Gianmario Rinaldi, Michele Cucuzzella, and Antonella Ferrara. Third order sliding mode observer-based approach for distributed optimal load frequency control. *IEEE Control Systems Letters*, 1(2):215–220, 2017.

[4] D.D. Siljak, D.M. Stipanovic, and A.I. Zecevic. Robust decentralized turbine/governor control using linear matrix inequalities. *IEEE Transactions on Power Systems*, 17(3):715–722, 2002.

[5] Song Zhang and Vijay Vittal. Design of wide-area power system damping controllers resilient to communication failures. *IEEE Transactions on Power Systems*, 28(4):4292–4300, 2013.

[6] Ramtin Hadidi and Benjamin Jeyasurya. Reinforcement learning based real-time wide-area stabilizing control agents to enhance power system stability. *IEEE Transactions on Smart Grid*, 4(1):489–497, 2013.

[7] Wenqi Cui, Yan Jiang, and Baosen Zhang. Reinforcement learning for optimal primary frequency control: A lyapunov approach. *IEEE Transactions on Power Systems*, 38(2):1676–1688, 2022.

[8] Jie Feng, Wenqi Cui, Jorge Cortés, and Yuanyuan Shi. Online event-triggered switching for frequency control in power grids with variable inertia. *arXiv preprint arXiv:2408.15436*, 2024.

[9] Shengyi Wang, Jiajun Duan, Di Shi, Chunlei Xu, Haifeng Li, Ruisheng Diao, and Zhiwei Wang. A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning. *IEEE Transactions on Power Systems*, 35(6):4644–4654, 2020.

[10] Ziming Yan and Yan Xu. A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system. *IEEE Transactions on Power Systems*, 35(6):4599–4608, 2020.

[11] Huaizhi Wang, Zhenxing Lei, Xian Zhang, Jianchun Peng, and Hui Jiang. Multiobjective reinforcement learning-based intelligent approach for optimization of activation rules in automatic generation control. *IEEE access*, 7:17480–17492, 2019.

[12] Vijay Pratap Singh, Nand Kishor, and Paulson Samuel. Distributed multi-agent system-based load frequency control for multi-area power system in smart grid. *IEEE Transactions on Industrial Electronics*, 64(6):5151–5160, 2017.

[13] Yuanyuan Shi, Guannan Qu, Steven Low, Anima Anandkumar, and Adam Wierman. Stability constrained reinforcement learning for real-time voltage control. In *2022 American Control Conference (ACC)*, pages 2715–2721. IEEE, 2022.

[14] Linfei Yin, Chenwei Zhang, Yaoxiong Wang, Fang Gao, Jun Yu, and Lefeng Cheng. Emotional deep learning programming controller for automatic voltage control of power systems. *IEEE Access*, 9:31880–31891, 2021.

[15] Xin Chen, Guannan Qu, Yujie Tang, Steven Low, and Na Li. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. *IEEE Transactions on Smart Grid*, 13(4):2935–2958, 2022.

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

[17] Sayak Mukherjee, Aranya Chakrabortty, He Bai, Atena Darvishi, and Bruce Fardanesh. Scalable designs for reinforcement learning-based wide-area damping control. *IEEE Transactions on Smart Grid*, 12(3):2389–2401, 2021.

[18] Draguna Vrabie, O Pastravanu, Murad Abu-Khalaf, and Frank L Lewis. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477–484, 2009.

[19] Yu Jiang and Zhong-Ping Jiang. Robust adaptive dynamic programming for large-scale systems with an application to multimachine power sys-

tems. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 59(10):693–697, 2012.

[20] Muhammad Nadeem, MirSaleh Bahavarnia, and Ahmad F Taha. Wide-area damping controller via reinforcement learning for power networks with wind and solar farms. In *2024 American Control Conference (ACC) (underview.* IEEE, 2024.

[21] Soummya Roy and Hugo Nestor Villegas Pico. Transient stability and active protection of power systems with grid-forming pv power plants. *IEEE Transactions on Power Systems*, pages 1–1, 2022.

[22] Hugo N Villegas Pico and Vahan Gevorgian. Blackstart capability and survivability of wind turbines with fully rated converters. *IEEE Transactions on Energy Conversion*, 37(4):2482–2497, 2022.

[23] P.W. Sauer, M.A. Pai, and J.H. Chow. *Power System Dynamics and Stability: With Synchrophasor Measurement and Power System Toolbox*. Wiley - IEEE. Wiley, 2017.

[24] Tomonori Sadamoto, Aranya Chakrabortty, Takayuki Ishizaki, and Jun-ichi 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*, 39(2):34–65, 2019.

[25] K. Takaba, N. Morihira, and T. Katayama. H_∞ control for descriptor systems-a J-spectral factorization approach. In *Proceedings of 33rd IEEE Conference on Decision and Control*, volume 3, pages 2251–2256 vol.3, 1994.

[26] Daniel Cobb. Controllability, observability, and duality in singular systems. *IEEE Transactions on Automatic Control*, 29(12):1076–1082, 1984.

[27] Sebastian A. Nugroho and Ahmad F. Taha. Load-and renewable-following control of linearization-free differential algebraic equation power system models. *IEEE Transactions on Control Systems Technology*, 31(4):1774–1786, 2023.

[28] Lillicrap Timothy, P, Alexander Jonathan J. Hunt, Heess Pritzel, Nicolas, Erez Tom, Tassa Yuval, Silver David, and Wierstra Daan. Continuous control with deep reinforcement learning. *arXiv preprint:1509.02971*, 2015.

[29] Fuzhen Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.

[30] Philip J. Hart, R.H. Lasseter, and T.M. Jahns. Symmetric droop control for improved hybrid ac/dc microgrid transient performance. In *2016 IEEE Energy Conversion Congress and Exposition (ECCE)*, pages 1–8, 2016.

[31] O. Wasynczuk, S.D. Sudhoff, T.D. Tran, D.H. Clayton, and H.J. Hegner. A voltage control strategy for current-regulated pwm inverters. *IEEE Transactions on Power Electronics*, 11(1):7–15, 1996.

[32] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, 2011.

[33] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[34] Erling D. Andersen and Knud D. Andersen. *The Mosek Interior Point Optimizer for Linear Programming: An Implementation of the Homogeneous Algorithm*, pages 197–232. Springer US, Boston, MA, 2000.

[35] Graham J. W. Dudgeon, William E. Leithead, Adam Dysko, John O'Reilly, and James R. McDonald. The effective role of avr and pss in power systems: Frequency response analysis. *IEEE Transactions on Power Systems*, 22(4):1986–1994, 2007.

[36] Hugo N Villegas Pico and Dionysios C Aliprantis. Voltage ride-through capability verification of wind turbines with fully-rated converters using reachability analysis. *IEEE Transactions on Energy Conversion*, 29(2):392–405, 2014.

[37] Muhammad Nadeem and Ahmad F Taha. Robust dynamic state estimation of multi-machine power networks with solar farms and dynamics loads. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 7174–7179. IEEE, 2022.

[38] Feng Yu and Yagoubi Mohamed. *Robust control of linear descriptor systems*. Studeis in system decison and control 102. Springer, 2017.

[39] Ahmad F Taha, Mohammadhafez Bazrafshan, Sebastian Adi Nugroho, Nikolaos Gatsis, and Junjian Qi. Robust control for renewable-integrated power networks considering input bound constraints and worst case uncertainty measure. *IEEE Transactions on Control of Network Systems*, 6(3):1210–1222, 2019.

[40] Johan Lofberg. Pre-and post-processing sum-of-squares programs in practice. *IEEE transactions on automatic control*, 54(5):1007–1011, 2009.

[41] X.P. Li and B.C. Chang. On the convexity of h/sup infinity / riccati solutions and its applications. *IEEE Transactions on Automatic Control*, 38(6):963–966, 1993.

[42] J.R. Cloutier and D.T. Stansbery. The capabilities and art of state-dependent riccati equation-based design. In *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, volume 1, pages 86–91 vol.1, 2002.

[43] Paul C Krause, Oleg Wasynczuk, Scott D Sudhoff, and Steven D Pekarek. *Analysis of Electric Machinery and Drive Systems*. John Wiley & Sons, 3 edition, 2013.

[44] NERC. *Dynamic Load Modelling*. North American Electric Reliability Corporation, Atlanta, GA, USA, Tech. Rep., 2016.

Appendix A. Details of considered power system model

In the NDAE (1), $\boldsymbol{x}_a(t) \in \mathbb{R}^{n_a}$ is given as:

$$\boldsymbol{x}_a(t) = \begin{bmatrix} \boldsymbol{V}_{\mathrm{Im}}^\top & \boldsymbol{V}_{\mathrm{Re}}^\top & \boldsymbol{I}_{\mathrm{Im}}^\top & \boldsymbol{I}_{\mathrm{Re}}^\top \end{bmatrix}^\top \quad\quad\quad (A.1)$$

where $\boldsymbol{V}_{\mathrm{Im}} = \{V_{\mathrm{Im}_i}\}_{i\in\mathcal{N}}$ represent the imaginary part of voltages and $\boldsymbol{V}_{\mathrm{Re}} = \{V_{\mathrm{Re}_i}\}_{i\in\mathcal{N}}$ are the real parts of voltage phasors. Similarly, $\boldsymbol{I}_{\mathrm{Im}} = \{I_{\mathrm{Im}_i}\}_{i\in\mathcal{N}}, \boldsymbol{I}_{\mathrm{Re}} = \{I_{\mathrm{Re}_i}\}_{i\in\mathcal{N}}$ are the imaginary and real parts of current phasors. The input vector $\boldsymbol{u}(t) \in \mathbb{R}^{n_u}$ lumps the control inputs for all the power plants and is given as:

$$\boldsymbol{u}(t) = \begin{bmatrix} \boldsymbol{u}_G^\top & \boldsymbol{u}_S^\top & \boldsymbol{u}_W^\top \end{bmatrix}^\top \quad\quad\quad (A.2)$$

where $\boldsymbol{u}_G(t) = \begin{bmatrix} \boldsymbol{P}_v^{*\top} & \boldsymbol{V}_g^{*\top} \end{bmatrix}^\top \in \mathbb{R}^{2G}$ represents the control inputs of synchronous generators with $\boldsymbol{V}_g^*$ denoting voltage set points of AVRs of the generators and $\boldsymbol{P}_v^*$ representing turbine valve position set-points. Similarly, $\boldsymbol{u}_S(t) = \begin{bmatrix} \boldsymbol{P}_s^{*\top} & \boldsymbol{V}_s^{*\top} \end{bmatrix}^\top \in \mathbb{R}^{2S}$ and $\boldsymbol{u}_W(t) = \begin{bmatrix} \boldsymbol{P}_w^{*\top} & \boldsymbol{V}_w^{*\top} \end{bmatrix}^\top \in \mathbb{R}^{2W}$ are the control inputs for solar and wind power plants with $\boldsymbol{P}_s^*, \boldsymbol{P}_w^*$ and $\boldsymbol{V}_s^*, \boldsymbol{V}_w^*$ denoting active power and voltage set-points for solar and wind-based power plants, respectively.

Also, in the system dynamics (1) the disturbance vector $\boldsymbol{w}(t) \in \mathbb{R}^{n_w}$ is modeled as $\boldsymbol{w}(t) = \begin{bmatrix} \boldsymbol{I}_s^\top & \boldsymbol{P}_L^\top \end{bmatrix}^\top$ in which $\boldsymbol{I}_s$ is the solar irradiance (W/m^2) and $\boldsymbol{P}_L$ represent the system active power load demand.

Moreover, the vector $\boldsymbol{x}_d(t) \in \mathbb{R}^{n_d}$ in (1) lumps the dynamic states of traditional power plant, solar, wind-based plant, and loads which is expressed as:

$$\boldsymbol{x}_d(t) = \begin{bmatrix} \boldsymbol{x}_G^\top & \boldsymbol{x}_S^\top & \boldsymbol{x}_W^\top & \boldsymbol{x}_M^\top \end{bmatrix}^\top \quad\quad\quad (A.3)$$

where $\boldsymbol{x}_G(t)$ represents the dynamic states of the conventional power plant, $\boldsymbol{x}_S(t)$ lumps dynamic states of solar power plant, $\boldsymbol{x}_W(t)$ contains wind power plant states, and $\boldsymbol{x}_M(t)$ denotes the dynamic states of the motor-based loads.

That being said, we model the conventional/traditional power plant via a comprehensive 9th-order dynamical model representing generators swing equations, excitation system, governor, and turbine dynamical models. Then, $\boldsymbol{x}_G(t) \in \mathbb{R}^{9G}$ is represented as follows [21, 23]:

$$\boldsymbol{x}_G(t) = \begin{bmatrix} \boldsymbol{\delta}_g^\top & \boldsymbol{\omega}_g^\top & \boldsymbol{E}_{\mathrm{fd}}^\top & \boldsymbol{E}_q^\top & \boldsymbol{E}_d^\top & \boldsymbol{T}_M^\top & \boldsymbol{P}_v^\top & \boldsymbol{r}_f^\top & \boldsymbol{v}_a^\top \end{bmatrix}^\top$$

where $\boldsymbol{\delta}_g$ denotes the rotor angle of the generator, $\boldsymbol{\omega}_g$ is the generator rotor speed, $\boldsymbol{E}_{\mathrm{fd}}$ is the field voltage of the generator, $\boldsymbol{E}_q$,

E_d are the transient voltages of the generator along q-axis and d-axis, respectively, T_M represents torque of the prime mover of the turbine, P_v denotes turbine valve position, while r_f and v_a are the stabilizer output and amplifier voltages, respectively. For further detailed explanations about the generator dynamics used in this study readers are referred to [23].

In (A.3), we model the solar plant dynamics via 12th-order dynamical system representing solar plants working in grid-forming mode as detailed in [20, 21, 31]. The complete dynamical equations modeling solar farms represents, DC side dynamics modeling PV arrays, DC link, current/voltage regulators dynamic equation, and AC side dynamics representing LCL filter equations, and AC/DC converter dynamics. That being said, the state vector for the solar plants $x_S(t) \in \mathbb{R}^{12S}$ can be written as follows:

$$x_S(t) = \begin{bmatrix} E_{dc}^\top & P_s^\top & Q_s^\top & \delta_s^\top & i_{df}^\top & i_{qf}^\top & v_{dc}^\top & v_{qc}^\top & z_{df}^\top & z_{qf}^\top & z_{do}^\top & z_{qo}^\top \end{bmatrix}^\top$$

where E_{dc} denotes the energy stored in the DC side capacitor, P_s is the real power while Q_s represents the reactive power injected by solar plants to the power grid, δ_s is the solar plant relative angle, $i_{df}, i_{qf}, v_{df}, v_{qf}$ are the dq-axis current and voltages from the solar power plants at their terminal bus, respectively, while $z_{df}, z_{qf}, z_{do}, z_{qo}$ are the dynamic states of voltage and current regulators along dq-axis, respectively. Interested readers are referred to [21, 31] for further in-depth explanations of the solar power plant model considered in this study.

Similarly, the dynamics of wind power plants have been modeled via 13th-order dynamical system as detailed in [22, 36]. The overall model describes double-fed induction generator (DFIG)-based wind turbine acting in a grid forming-mode and thus the state vector for the wind plants $x_W(t) \in \mathbb{R}^{13W}$ can be expressed as follows:

$$x_W(t) = \begin{bmatrix} \delta_w^\top & E_{dcw}^\top & i_{wf}^\top & v_{wc}^\top & P_w^\top & Q_w^\top & z_{dc}^\top & z_{wf}^\top & z_{wo}^\top \end{bmatrix}^\top \quad \text{(A.4)}$$

where δ_w represents the inverter relative angle of the wind plant, E_{dcw} is the energy stored in the DC link capacitor of the wind plant, $i_{wf} = [i_{dw_f}^\top \ i_{qw_f}^\top]^\top$ denotes the inverter output current at the terminal bus along dq-axis, similarly $v_{wc} = [v_{dw_c}^\top \ v_{qw_c}^\top]^\top$ represents the dq-axis AC capacitor voltages, P_w, Q_w respectively denote the real and reactive power output from the wind power plant to the power grid, while $z_{dc}, z_{wf} = [z_{dw_f}^\top \ z_{qw_f}^\top]^\top$, $z_{wo} = [z_{dw_0}^\top \ z_{qw_0}^\top]^\top$, are the states of the current and voltage regulators of the wind power plant along dq-axis, respectively. Further details explanations about the wind power plants used in this work can be found in [22, 36].

Furthermore, in (A.3) $x_M(t) \in \mathbb{R}^{L_k}$ denotes the speed $\omega_{m_i}(t)$ of the motor-based load and is given as [43]:

$$\dot{\omega}_{m_i}(t) = \frac{1}{2H_{M_i}}(T_{e_i} - T_{M_i}) \quad \text{(A.5)}$$

where $i \in \mathcal{L}$, T_{e_i} represent electromagnetic torque of motor, T_{M_i} is mechanical torque while H_{M_i} is the inertia constant of the motor-based load [43]. The constant impedance/power-based loads satisfy the following equations [44]:

$$I_{z_i} Z_i + V_{z_i} = 0 \quad \text{(A.6a)}$$

$$(P_{p_i} + Q_{p_i}) + \text{conj}(I_{p_i})V_{p_i} = 0 \quad \text{(A.6b)}$$

where $i \in \mathcal{L}$, Z_i represent the impedance of the load, P_{p_i}, Q_{p_i} are the real and reactive power of constant power loads, the term conj denotes the complex conjugate operator while $I_{z_i}, V_{z_i}, I_{p_i}, V_{p_i}$ are the current and voltage phasors of buses connected to constant impedance and constant power loads, respectively. This completes the modeling of differential equations (1a) of the power system NDAE model (1).

The algebraic constraints (1b) model the current balance equation between all the power plants and loads and thus can be written as [23]:

$$\underbrace{\begin{bmatrix} I_G(t) \\ I_S(t) \\ I_L(t) \\ I_W(t) \end{bmatrix}}_{I(t)} - \underbrace{\begin{bmatrix} Y_{GG} & Y_{GS} & Y_{GL} & Y_{GW} \\ Y_{SG} & Y_{SS} & Y_{SL} & Y_{SW} \\ Y_{LG} & Y_{LS} & Y_{LL} & Y_{LW} \\ Y_{WG} & Y_{WS} & Y_{WL} & Y_{WW} \end{bmatrix}}_{Y} \underbrace{\begin{bmatrix} V_G(t) \\ V_S(t) \\ V_L(t) \\ V_W(t) \end{bmatrix}}_{V(t)} = 0 \quad \text{(A.7)}$$

where $V(t)$, $I(t)$ are the net voltages and currents and Y represents the system admittance matrix. In (A.7) the terms $I_G = \{I_{Re_i}\}_{i \in \mathcal{G}} + j\{I_{Im_i}\}_{i \in \mathcal{G}}$, $V_G = \{V_{Re_i}\}_{i \in \mathcal{G}} + j\{V_{Im_i}\}_{i \in \mathcal{G}}$, denote real and imaginary parts of synchronous generators terminal buses current and voltage phasors, respectively. Similarly, I_S, I_W, I_L and V_S, V_W, V_L denote current and voltage phasors of all solar, wind-based power plants, and loads, respectively.
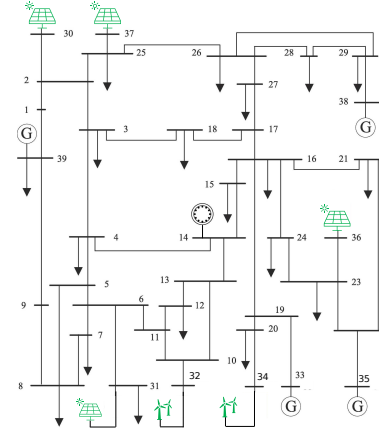

Figure A.7: Diagram of the 39-bus test system with PV plants at Buses 30, 31, 36, 37, wind-based plants at Buses 32, 34, conventional power plants at Buses 33, 35, 38, 39, and a motor-based load at Bus 14.
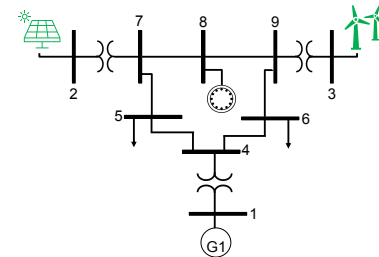

Figure A.8: Diagram of the 9-bus system with a PV-based plant at Bus 2, a wind plant at Bus 3, steam-based plant at Bus 1, and a motor-based load at Bus 8.