6-10-2022

Modified Q-Learning Method for Automatic Voltage Regulation in Wide-Area Multigeneration Systems

Brook Abegaz

Sina Zarrabian

WILEY | Hindawi

Research Article

Modified Q-Learning Method for Automatic Voltage Regulation in Wide-Area Multigeneration Systems

Brook W. Abegaz [1] **and Sina Zarrabian** [2]

[1]*Department of Engineering, Loyola University Chicago, Chicago, IL, USA*
[2]*Department of Electrical Engineering, State University of New York, Throggs Neck, New York, NY, USA*

Correspondence should be addressed to Brook W. Abegaz; babegaz@luc.edu

The state-estimation and optimal control of multigeneration systems are challenging for wide-area systems having numerous distributed automatic voltage regulators (AVR). This paper proposes a modified Q-learning method and algorithm that aim to improve the convergence of the approach and enhance the dynamic response and stability of the terminal voltage of multiple generators in the experimental Western System Coordinating Council (WSCC) and large-scale IEEE 39-bus test systems. The large-scale experimental testbed consists of a six-area, 39-bus system having ten generators that are connected to ten AVRs. The implementation shows promising results in providing stable terminal voltage profiles and other system parameters across a wide range of AVR systems under different test scenarios including N-1 contingency and fault conditions. The approach could provide significant stability improvement for wide-area systems as compared to the implementation of conventional methods such as using standalone AVR and/or power system stabilizers (PSS) for the wide-area control of power systems.

1. Introduction and Literature Review

Automatic Voltage Regulators (AVRs) play a significant role in supporting the generator's terminal voltage, enhancing transient stability, and improving the damping of oscillations in power systems. AVRs are typically associated with power system stabilizers (PSS) since an AVR may not be adequate for damping oscillations in a standalone operation. Therefore, a combination of AVR and PSS has been much focused on the topic of voltage regulation and stability in the past decade. Additionally, the use of AVRs has been seen as a better alternative as compared to bulk equipment such as FACTS which will be utilized mainly for damping oscillations in interconnected grids.

The state prediction, estimation, and control are challenging areas in wide-area systems due to uncertainties in the optimal operating voltages of multiple distributed generators. Research in this area has focused on improving the transient stability of power systems and damping interarea oscillations that could negatively affect the power transfer capability. In the past, researchers have focused on the study of power systems oscillations damping, angle stability improvement, and avoiding cyber uncertainties such as communication delay and package dropouts [1]. Among the various approaches used are a radial basis function network [2], a sparse control architecture [3, 4], a hierarchical controller-based stability enhancement [5], coordination of AVR and PSS to damp large-scale power systems [6], optimization algorithms for the wide-area damping controller to improve the damping rates of low-frequency oscillation modes considering power system operating uncertainties, variations in the time delay and robustness to a permanent failure of one WADC communication channel [7], and reinforcement learning approaches [8]. However, the wide-area coordination and control of voltage regulator-based systems is still a challenge considering the increasing numbers of generators being connected with various types of AVR systems. Literature has addressed the rotor angle stability [8] and the adaptivity of the linear model-based control algorithm. However, it has been

indicated in [1, 9] and [10] that there are still many open problems to investigate with control of cyber-physical WAPS of such systems but not much on the terminal voltage stability. This initiates the main motivation of this work to find reinforcement learning-based methods and algorithms to coordinate the operation of automatic voltage regulators in a smart power grid scenario.

Driven by the need for maintaining the terminal voltage stability in power systems, different adaptive control approaches [1–19], such as neural network methods [11] and optimization approaches [12] of AVRs, are being explored for terminal voltage stability purposes. An efficient, robust, and adaptive AVR control is necessary due to uncertainty, nonlinearity, and high-level penetration of distributed energy resources in modern power grids. For example, a nonlinear algorithm was proposed in [13] by designing a combined AVR and load frequency control (LFC) method. The results are convincing; however, the tests are operated under small disturbances but not fault conditions and the two-area testbed characteristics are unknown. In [14], an excitation control method was proposed based on wide-area measurement to improve transient stability. The drawback of the proposed method is that the setting of the parameters was done offline after each fault and not based on real-time changes in the system. In another work [15], an AVR control scheme was presented based on a PID controller that utilizes a tree-seed algorithm to optimize PID controller gains. In [16], a kidney-inspired algorithm was proposed for optimized tuning of the PID gains for a PID-based AVR system. The drawback of the work is that the proposed method is only tested under changes in some AVR system parameters and did not reflect the robustness of the AVR control system under fault or disturbance conditions or from the instability point of view. Papers [17, 18] make use of the fractional-order PID in addition to the second-order derivative controller (FOPIDD) to achieve a better transient response at a terminal voltage of AVR. In [17], the parameters of the controller are tuned optimally by the Equilibrium Optimizer (EO) algorithm, whereas [18] utilizes a multiverse optimizer (MVO) algorithm to tune the parameters of the controller. These algorithms have been tested under different disturbances such as changes in time constant and gain constant. However, they did not assess the AVR performance in terms of wide-area voltage stability and dynamic response to contingencies, e.g., N-1 or severe fault analysis. In [19], three different intelligent methods, i.e., fuzzy logic, artificial neural network, and brain emotional learning coordinators are used for coordinating AVR and PSS in a multimachine power system. The proposed algorithm is only evaluated on a small, four-machine test system, not in a wide-area and complex power system.

The learning-based excitation control of the terminal voltages of synchronous generators (SGs) is an interesting research problem to solve due to two main reasons:

(1) AVRs of SGs that are electrically far from a fault location could have a low contribution to excitation voltage change as compared to AVRs of SGs near the fault location. In this paper, a method is presented

that improves the excitation systems of SGs that are farthest away from the fault locations to enhance the voltage and transient stability.

(2) There are trade-offs in the effectiveness of AVRs and PSSs in transient stability analysis as PSSs specifically contribute to damping but do not necessarily contribute to first swing stability improvement, whereas AVRs do contribute to the first swing stability enhancement.

To explore the research problem further, three research questions are formulated in this paper.

(i) How wide is the impact of AVR in a power system?

(ii) Do the settings of AVRs affect an interconnected power system effectively in wide-area connections? Or are their impacts only locally?

(iii) Why and what information do AVRs need to communicate through wide-area connections?

In this paper, a modified Q-learning-based adaptive AVR system is applied to a multiarea large-scale power system to enhance the voltage stability and improve the excitation system's performance. The main contributions of the paper that answer the original research questions posed in this paper are as follows:

(a) The Q-learning algorithm can compute the impact of changes on the gains of the AVR systems.

(b) The settings of AVR gains directly affect the voltage stability of the interconnected power system and they are not limited only locally.

(c) The communication of AVR parameters helps aid the coordinated operation of the power system across wide areas.

This paper is organized as follows. The description of the problem and literature review are presented in Section 1. Next, the reinforcement learning approach is presented in Section 2. The modified Q-learning approach and its implementation are presented in Section 3. The results are presented in Section 4, and the discussion and conclusion are presented in Section 5.

2. The Reinforcement Learning Approach

Reinforcement Learning (RL) is a machine learning-based algorithm in which an agent interacts with the environment through the states, actions, and rewards to learn an optimal policy to reach a predefined target. At each step of the learning process, a reward is received by the agent through the transition from a state to action. The main objective of the RL is to explore an optimal solution in which the expected cumulative rewards are maximized using (1):

$$R_t(s, a) = \sum_{n=0}^{\infty} \gamma^n r_{t+n+1} \equiv r_{t+1} + \gamma \max Q_t(s_{t+1}, a_{t+1}), \quad (1)$$

where $\gamma \in [0, 1]$ is the discount factor.

Q-learning (QL) is a well-established model-free algorithm based on the Temporal Difference (TD) learning method of solving the RL problem. In the QL algorithm, the action-value function (*Q*-value) is estimated. The derived expression of QL is based on the Bellman optimality function as given in (2), where $\alpha \in [0, 1]$ is the learning rate, $Q_t(s_t, a_t)$ is first initialized (estimated), and action a_t is selected in the state s_t:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha[r_{t+1} + \gamma \max Q_t(s_{t+1}, a_{t+1}) - Q_t(s.a)]. \tag{2}$$

The state s_{t+1} is obtained with an immediate reward r_{t+1} and the $\max Q_t(s_{t+1\Delta}, a)$ associated with the new state s_{t+1}. To store and update the expected future *Q*-values, a lookup *Q*-table is created that is indexed by state and action. Then, the optimal *Q*-values are approximated iteratively by initializing arbitrary values. To obtain an optimal *Q*-value, there is a trade-off between exploration and exploitation. This means that all possible actions are considered in every state with a nonzero probability in addition to actions with the highest *Q*-value and probabilities. In this paper, SoftMax (Boltzmann exploration) action selection policy is used to get the right trade-off where the probability of selected actions is weighted based on their *Q*-values as given in (3), where $T \geq 0$ is the temperature parameter:

$$P\left(\frac{a}{s}\right) = \frac{e^{Q(s,a)/T}}{\sum_a e^{Q(s,a)/T}}. \tag{3}$$

Smaller values of T lead to an action selection policy with a more greedy strategy; however, higher values of T result in a random strategy. The value of T can be adjusted during the learning process to achieve a better trade-off in exploration-exploitation.

3. Modified Q-Learning Method for AVR Control

To obtain an accurate solution for terminal voltage control and transient stability enhancement, the AVR gains could be adjusted using a modified Q-learning-based method. Conventional Q-learning algorithms suffer from convergence issues because of state-action transition and reward for the same transition which results from stochastic characteristics of the process. To overcome this issue and guarantee a fast convergence, we deploy the Monte Carlo method to obtain the expected *Q*-value by averaging the perceived rewards [20]. Thus, (2) can be rewritten as

$$Q_{t+1}(s,a) = Q_t(s,a) + \frac{1}{n}[R_t(s,a) - Q_t(s,a)], \tag{4}$$

where n specifies the number of selected (s, a) and the modified learning rate is defined as

$$\alpha(n) = \frac{1}{n}. \tag{5}$$

Based on the stochastic approximation theory, learning rate constraints are defined as follows to ensure convergence with a probability of unity [21]:

$$\sum_{n=1}^{\infty} \alpha(n) = \infty \& \sum_{n=1}^{\infty} \alpha^2(n) < \infty. \tag{6}$$

To resolve the issue that future reward in every state-action transition $r(s_{t+1}|s,a)$ varies with iterations, the first item in (1) is replaced by

$$\overline{r_t'}(s_{t+1}|s,a) = \frac{1}{n'}[(n'-1)\overline{r_{t-1}}(s_{t+1}|s,a) + r_t]. \tag{7}$$

$n\prime$ specifies the number of times that (s, a, s_{t+1}) is observed.

After modification of the Q-learning algorithm, the three main functions of the state, action, and reward should be defined properly. In this article, the objective is to adjust the terminal voltage of generators in an interarea system. Hence, the state space is defined as the terminal voltage of generators, where $V_{i,t}$ is the ith generator voltage at the time step t as given in (8).

$$S = \{s|s_{i,t} = V_{i,t}\}. \tag{8}$$

The action of the terminal voltage adjustment is taken by the generator's agent and is defined as given in (9), where $\Delta K_{A,i}$ is a change in AVR gain of the generator i. This change is applied through the exciter control unit of each generator.

$$A = \{a| - \Delta K_{A,i}, 0, +\Delta K_{A,i}\}. \tag{9}$$

The action space is discretized into three values: "decrease or decrement" $(-\Delta K_{A,i})$, "no change" (0), and "increase or increment" $(+\Delta K_{A,i})$. This means that, at each stage, the agent can select among three actions for each excitation system. In each time step, the agent selects an action from the action space based on the current state and the action selection policy to update the power adjustment signal.

The reward function should be defined properly based on the predefined objective of the agent. Based on the goal of the agent, the reward function is defined in (10) where $V_{i,t}$ and $V_{i,t+1}$ are the voltages of the ith generator at two sequential iterations and $\sigma > 0$ is a small number taken, for example, as 0.05.

$$r(s_{i,t}, a_{i,t}) = \begin{cases} +1, & \text{if } V_{i,t+1} - V_{i,t} < -\sigma, \\ 0, & \text{if } |V_{i,t+1} - V_{i,t}| \leq \sigma, \\ -1, & \text{if } V_{i,t+1} - V_{i,t} > \sigma. \end{cases} \tag{10}$$

After defining the three main functions which are essential for the Q-learning process, the parameters expressed in (1) and (2) are presented in Table 1.

3.1. Q-Leaning Algorithm Implementation Steps. Figure 1 shows the block diagram of the excitation system with the proposed Q-learning algorithm for further understanding. Table 2 demonstrates a summarized implementation of the proposed algorithm. Since the original Q-learning algorithm is not efficient in terms of run-time, in this paper, a modification is implemented which considers a threshold value for the ordinary differential equation of d*S*/d*t* equal to $10e - 4$ to improve the algorithm's convergence by

TABLE 1: Q-learning parameters.

Q-learning parameters	Value
Learning rate (α)	0.01
Discount factor (γ)	0.9
Temperature $(\mathbf{T})$	$T_{\min} = 1.0, T_{\max} = 10$

FIGURE 1: Block diagram of the excitation system with Q-learning control.

TABLE 2: Steps of the modified Q-learning implementation.

Modified Q-learning algorithm with Boltzmann exploration

Input: learning rates (α), discount factor (γ),
temperature $(\mathbf{T})$, $\Delta \mathbf{K}_{A,i} = [-5, 5]$
(1) Initialize Q-function using Q-table, e.g., $\mathbf{Q}_0 \longleftarrow 0$
(2) Measure initial states, $\mathbf{s}_{i,0} \in \mathbf{S}$
(3) For each time interval (iteration) $\mathbf{t} = 0, 1, 2, \cdots$, **Do:**
   (3.1) Action selection using,
   $\mathbf{a}_{i,t} \in \mathbf{A} \longleftarrow \mathbf{P}(\mathbf{a}|\mathbf{s}) = \mathbf{e}^{\mathbf{Q}(s,a)/T} / \sum_a \mathbf{e}^{Q(s,a)/T}$
   (3.2) Apply $\mathbf{a}_{i,t}$, measure $\mathbf{s}_{i,t+1}$ and reward $\mathbf{R}_t(\mathbf{s}, \mathbf{a})$
     (3.3) Update Q-function:
   $\mathbf{Q}_{t+1}(\mathbf{s}, \mathbf{a}) \longleftarrow \mathbf{Q}_t(\mathbf{s}, \mathbf{a}) + \alpha(\mathbf{n})[(\mathbf{r}_{t+1}+)/\gamma \max_a \mathbf{Q}_t(\mathbf{s}_{t+1}, \mathbf{a}_{t+1}) - \mathbf{Q}(\mathbf{s}_t \cdot \mathbf{a}_t)]$
(4) Check for convergence, $|\mathbf{Q}_{t+1}(\mathbf{s}_t, \mathbf{a}_{t,i}) - \mathbf{Q}_t(\mathbf{s}_t, \mathbf{a}_{t,i})| \leq \varepsilon$
(5) Update AVR variables: if $(\mathbf{d}S/\mathbf{d}t) \geq 1\mathbf{e} - 4 \times [0 \cdot 050 \cdot 05]$
$\mathbf{E}_{x+1}(i, j) = \mathbf{E}_x(i, j) \pm \Delta \mathbf{K}_{A,i} * \mathbf{AVR}(i)$
(6) Output: ODE Solution $[t, x]$; AVR Exciter variables $\mathbf{E}_x(i, j)$
(7) End For

FIGURE 2: WSCC 9-bus test system.

minimizing the search space, where S is the state space of the terminal voltage of generators in the system.

4. Results and Discussion

To test the effectiveness and robustness of the proposed method, three different scenarios are considered which include (1) a disturbance in the AVR system, (2) an N-1 contingency, and (3) a fault scenario. These scenarios are tested on the Western System Coordinating Council (WSCC) 9-bus test system that contains three synchronous generators and six transmission lines as shown in Figure 2. Additionally, the N-1 contingency and fault scenarios are implemented in the large-scale, New England IEEE 39-bus test system that contains 10 machines and 46 transmission lines.

4.1. Test Results for WSCC 9-Bus Test System. In the first scenario, a sudden 0.1 p.u. step increase is applied to the reference voltage of the G1 excitation system (V_{ref1}) at $t = 1.5$ sec. The response of the system is shown in Figures 3 and 4. In Figure 3(a), the terminal voltage of generator buses is shown. Figure 3(b) shows the terminal voltage of non-generation buses. It is seen that after the disturbance is applied, the AVR systems of generators with the Q-learning control scheme act very fast, damp oscillations quickly, keep voltage terminals within acceptable ranges, and provide voltage stability for the entire system. Similarly, Figures 4(a) and 4(b) show the bus voltage angle of SGs and non-generation buses, respectively. They also show the effectiveness of the proposed control in balancing the bus voltage's angle after the disturbance.

In the second scenario, an N-1 contingency is applied to the test system. To achieve this scenario, the breaker connected to the transmission line 4-5 is disconnected at
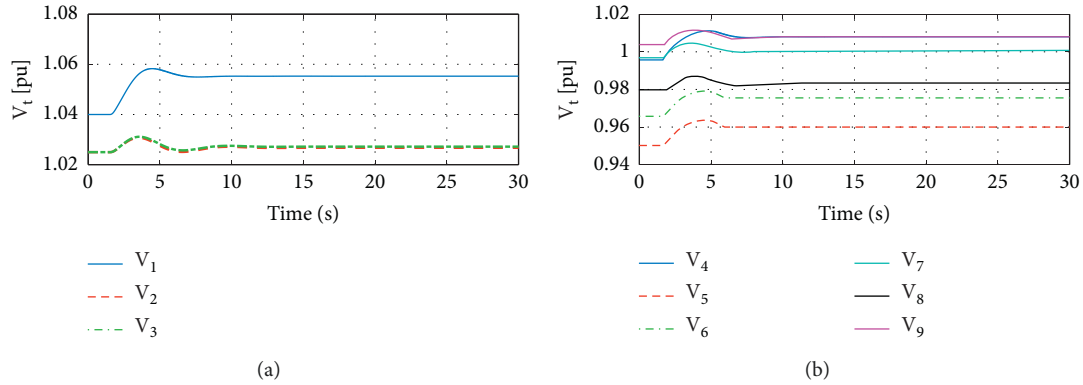
FIGURE 3: (a) Terminal voltages after disturbance of generator buses. (b) Terminal voltages after disturbance of nongeneration buses.
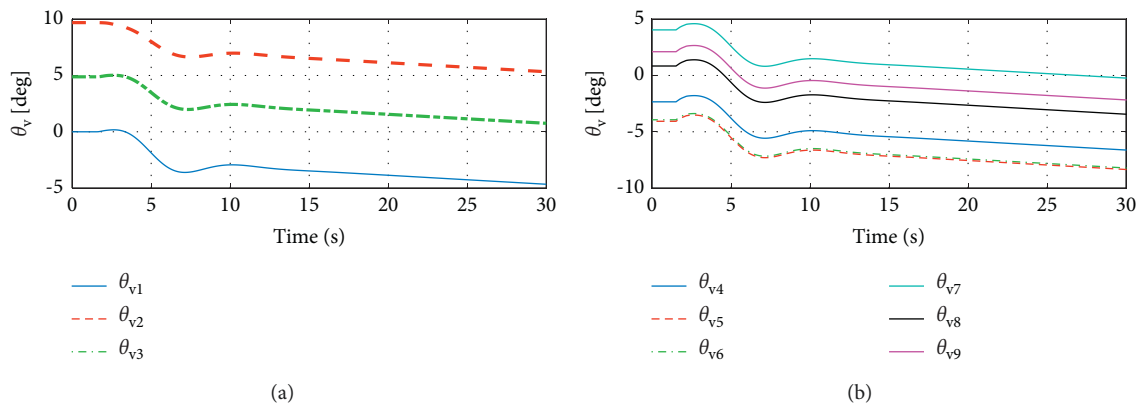

FIGURE 4: (a) Angle of terminal voltage after disturbance for SGs. (b) The angles of terminal voltages after disturbance for nongenerating buses.

$t = 1.5$ sec. The response of the system is shown in Figures 5 and 6. In Figures 5(a)–5(c), the terminal voltage of G1, G2, and G3 is shown, respectively. To check the robustness and efficacy of the proposed Q-learning control algorithm, the results are compared with the conventional combination of AVR and PSS. It is observed that after an N-1 contingency occurs, the AVR system equipped with the Q-learning algorithm has significantly less voltage deviation right after the contingency happens, damps terminal voltage oscillations much faster than the traditional excitation system, and enhances voltage stability for the entire system. On the other hand, the terminal voltage of generators equipped with AVR + PSS encounters considerable voltage dips after contingency, and voltage oscillations and instability remain in the system for a longer time.

Figures 6(a) and 6(b) depict the voltage angle of the generators referenced to the slack bus (bus 1). In Figures 6(a) and 6(b), the voltage angle of G2 and G3 are shown, respectively. For the sake of effectiveness review and analysis, the results are compared with the conventional combined AVR and PSS. It is observed that the AVR systems with the help of the proposed Q-learning control can damp oscillations much faster, keep angle deviations within an acceptable margin, and stabilize the voltage angle profile on time. On the other hand, the AVR system with conventional PSS has a slower response

in damping angle oscillations and it is not effective to limit magnitude deviations shortly after the contingency occurs.

In the third scenario, a three-phase to ground fault is applied to the 9-bus system. The fault occurs at bus 6, at $t = 1.5$ sec with a clearing time of 0.01 sec, and fault resistance and reactance of 0.01 and 0.001 p.u., respectively. The behavior of the system under the fault condition is shown in Figure 7. In Figures 7(a)–7(c), the terminal voltage of G1, G2, and G3 is shown, respectively. Additionally, to better observe the effectiveness and robustness of the proposed control algorithm, the results are compared with the conventional combination of AVR and PSS. Again, it is demonstrated that after the three-phase fault is implemented, the AVR systems equipped with the Q-learning control smooth voltage deviations significantly in a very short period, damp terminal voltage oscillations, and improve voltage stability of the system. In contrast, in the system with only AVR + PSS, the terminal voltage of the generators drops significantly before the fault is cleared and the voltage oscillations and the instability remain in the system after the fault is cleared as well. Unlike the N-1 scenario, the voltage magnitude oscillations reduce at a very slow rate as time elapses and are not eliminated even after a long time following the fault clearance. In the case of a conventional AVR system without adaptive control, the system needs significant time to reach stability.
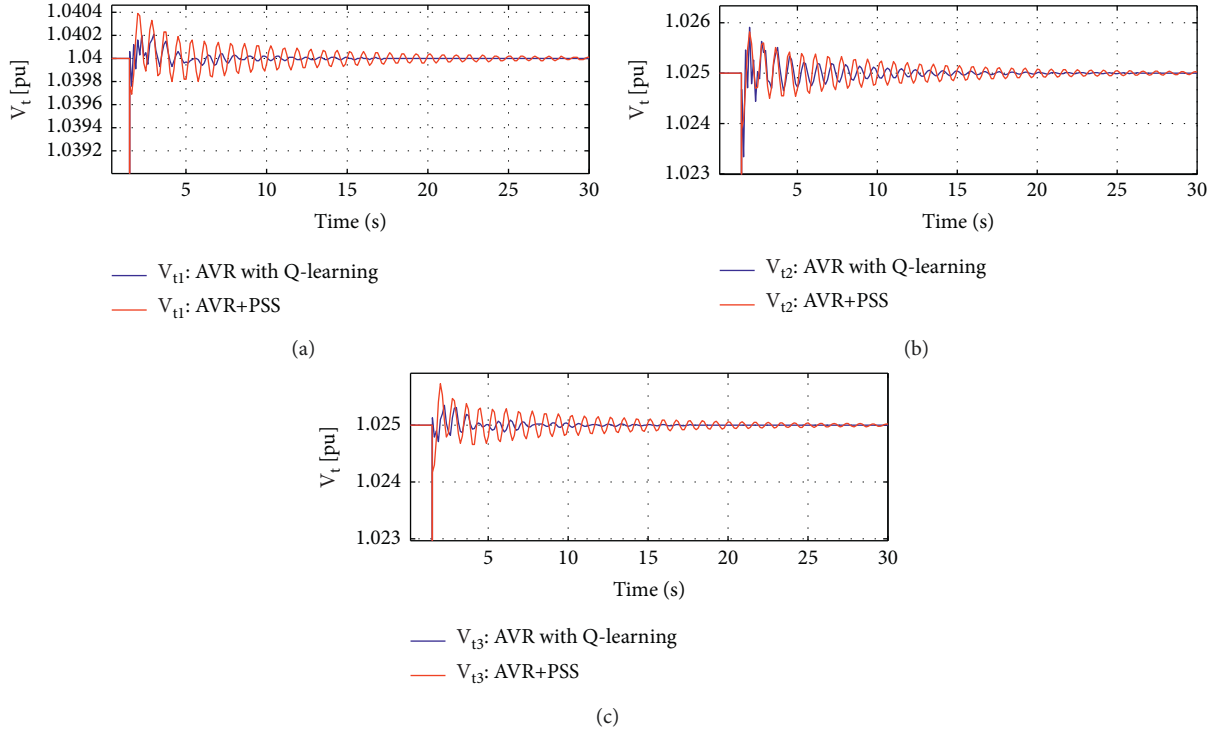
(a)

(b)

(c)

FIGURE 5: The terminal voltages of the generator buses for an N-1 contingency scenario: (a) bus voltage of G1; (b) bus voltage of G2; (c) bus voltage of G3.
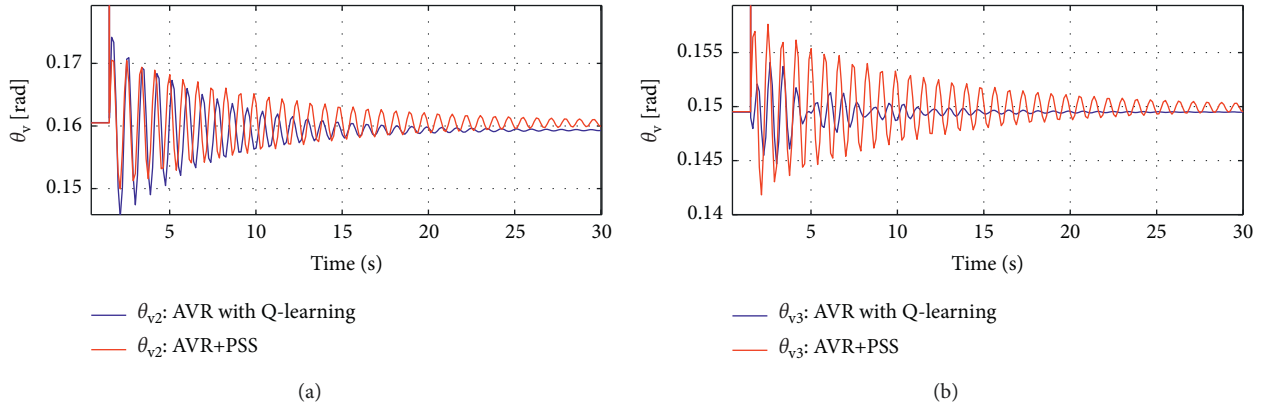

(a)

(b)

FIGURE 6: Voltage angle of generators referenced to the slack bus for an N-1 contingency scenario: (a) bus voltage of G1; (b) bus voltage of G2.

In another comparison during the three-phase to ground fault scenario and due to the importance and severity of this contingency, the angle of the generators' voltage referenced to the slack bus (bus 1) is shown in Figure 8. In Figures 8(a) and 8(b), the voltage angles of G2 and G3 are shown, respectively. For the sake of effectiveness review and analysis, the results are compared with the conventional combined AVR and PSS. It is observed that the AVR systems with the support of the proposed Q-learning control can damp oscillations much faster, keep angle deviations within an acceptable margin even right after the fault occurs, and stabilize the angle profile on time. However, with the combined AVR and PSS control scheme, the oscillations in

the angle profile result in instability with very high oscillating magnitudes. With the presence of the conventional AVR and PSS control, the voltage angles remain undamped after the fault is cleared and the system does not settle down. Unlike the N-1 contingency case, the magnitude of voltage angles does not reduce even after the fault is cleared and the system remains unstable.

4.2. Test Results for IEEE 39-Bus Test System.
To further demonstrate the efficacy and robustness of the proposed method in a large and complex system, the N-1 contingency and three-phase fault scenarios are implemented in the New
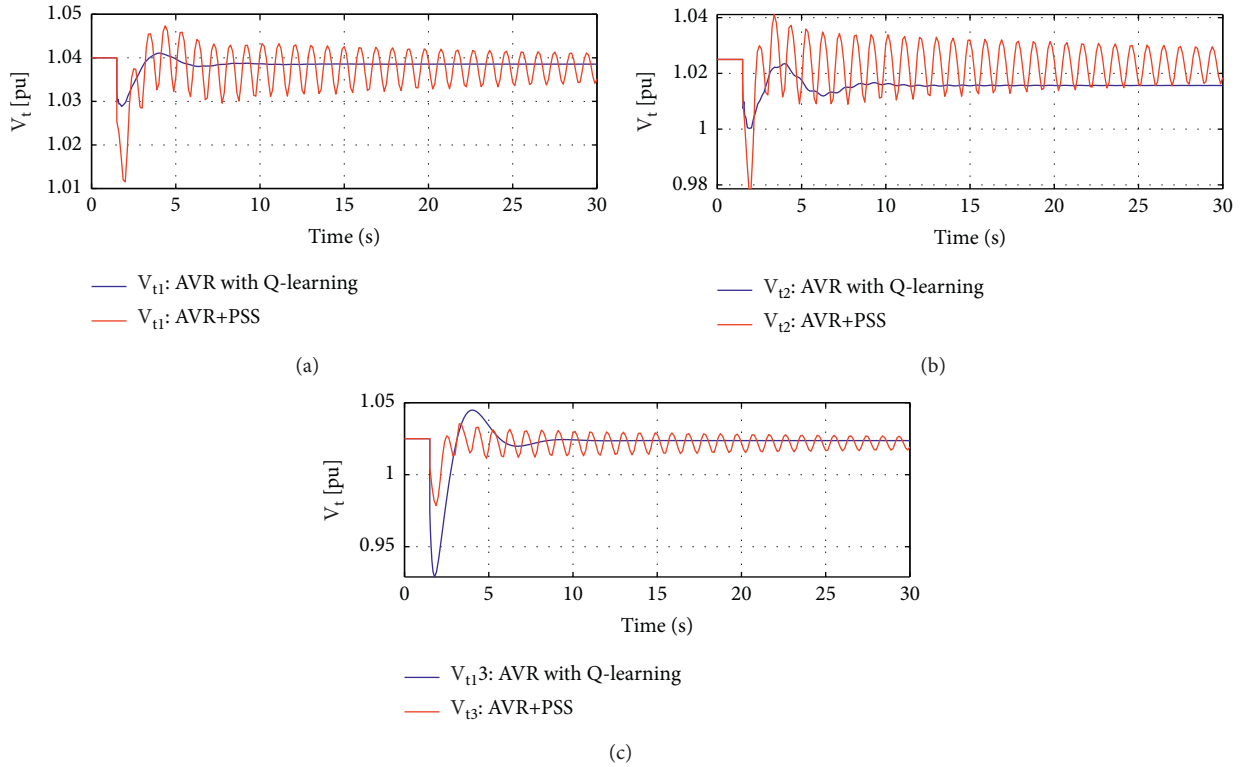
(a)

(b)

(c)

Figure 7: Terminal voltage of generator buses for three-phase fault scenario: (a) bus voltage of G1; (b) bus voltage of G2; (c) bus voltage of G3.
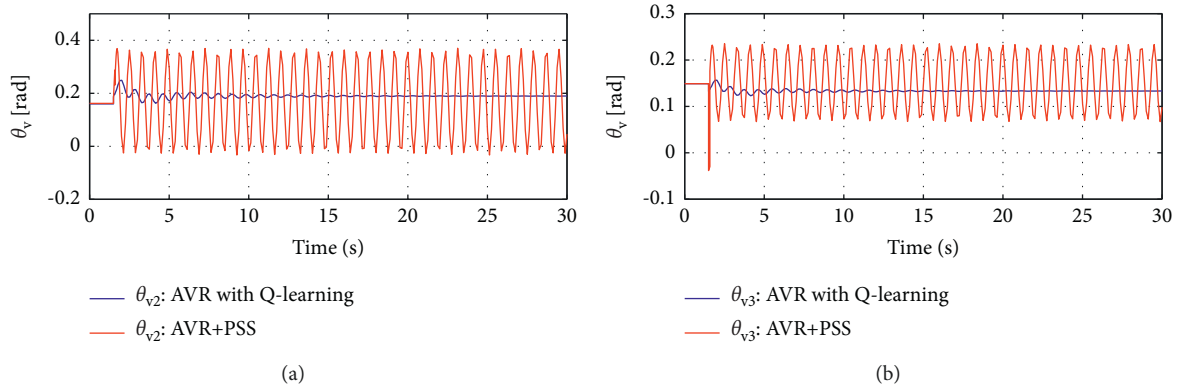

(a)

(b)

Figure 8: Voltage angle of generators referenced to the slack bus for three-phase fault scenario: (a) bus voltage of G1; (b) bus voltage of G2.

England IEEE 39-bus test system. Figure 9 shows the response of the AVR system in regulating the generators' terminal voltage when an N-1 contingency is applied to the test system. To achieve this scenario, the breaker connected to the transmission line 8-9 is disconnected at $t = 1.0$ sec. Figure 9 depicts the terminal voltage of all the generators. To check the robustness and efficacy of the proposed Q-learning control algorithm, the results are compared with the conventional AVR and PSS combination. It is observed that after the N-1 contingency occurs, the AVR system equipped with the Q-learning algorithm has significantly less voltage deviation right after the contingency happens, damps terminal voltage oscillations much faster than the traditional excitation system,

and enhances voltage stability for the entire system. It is observed that the proposed Q-learning control is very capable of adjusting AVR gains accurately across all the generators to provide voltage stability for the entire system when a critical contingency occurs. Additionally, the proposed control system can maintain all terminal voltages at their prefault level after the contingency happens. On the other hand, the terminal voltage of generators equipped with AVR + PSS encounters considerable voltage dips right after the contingency, and voltage oscillations remain in the system for a longer time. It is also clear that G7 and G8 are not able to maintain their voltages at their prefault values when conventional AVR control is deployed.
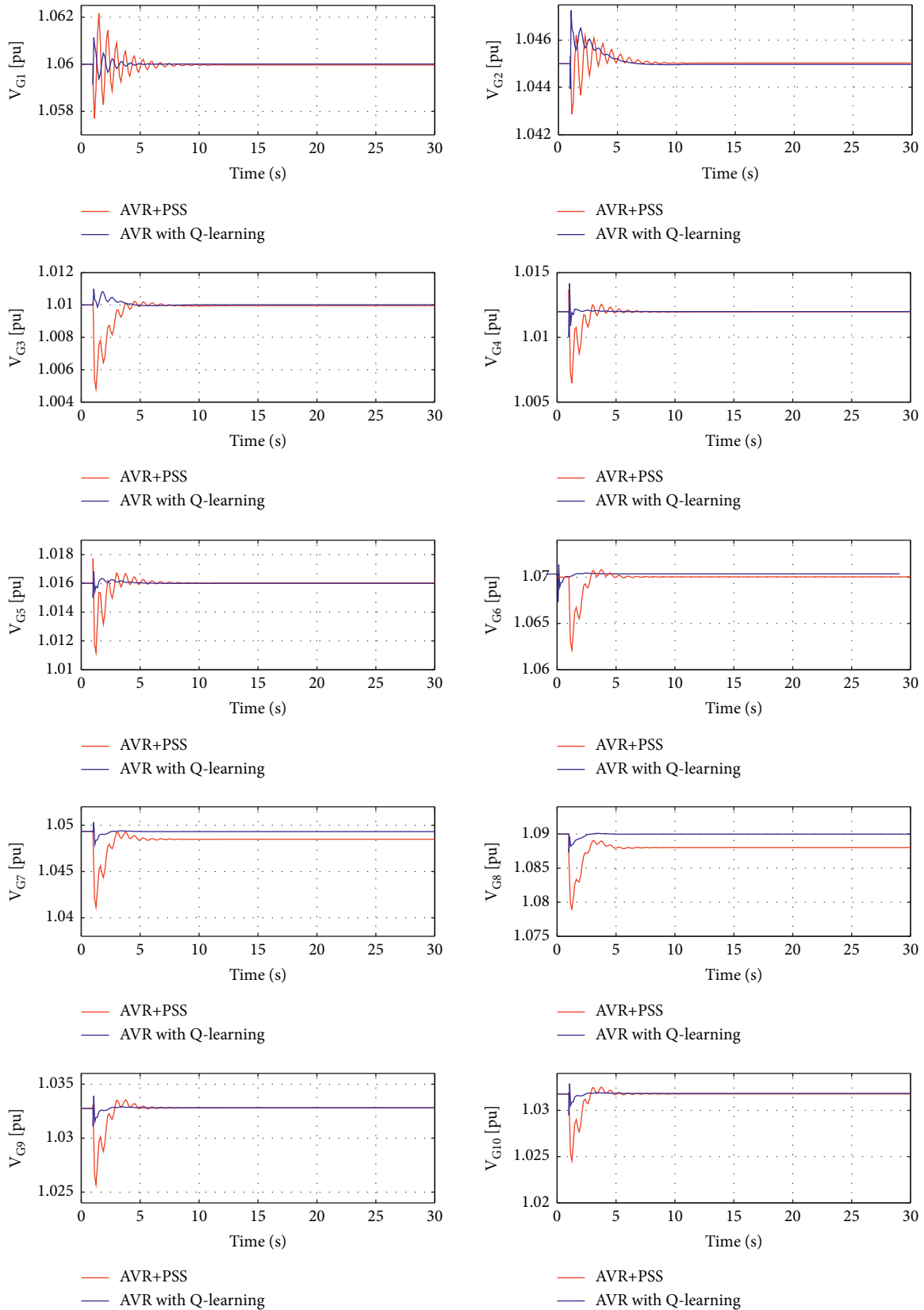
FIGURE 9: The terminal voltage of generator buses for an N-1 contingency scenario.
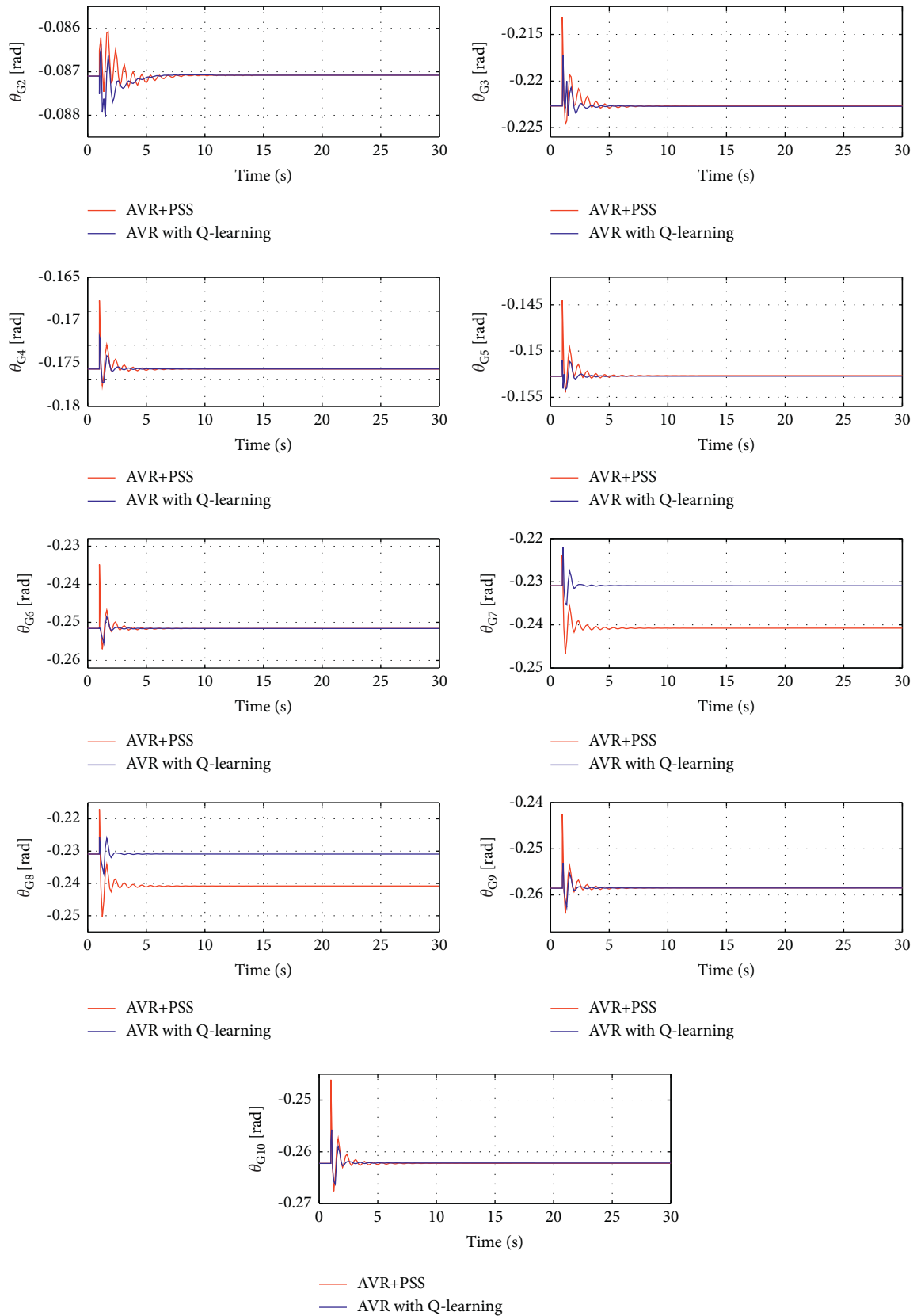
FIGURE 10: Voltage angle of generators referenced to the slack bus for an N-1 contingency scenario.
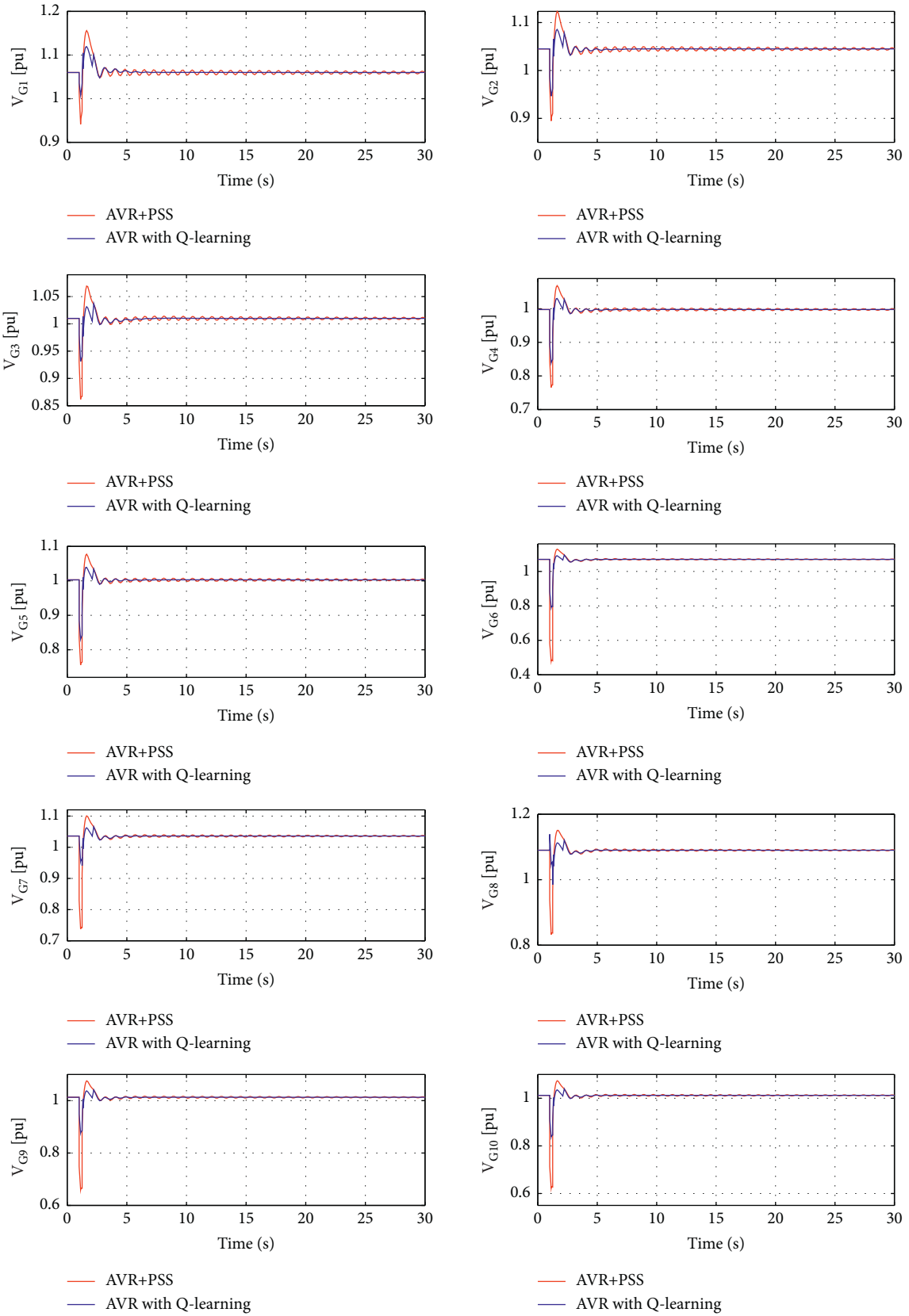
Figure 11: The terminal voltage of generator buses for a three-phase fault scenario.
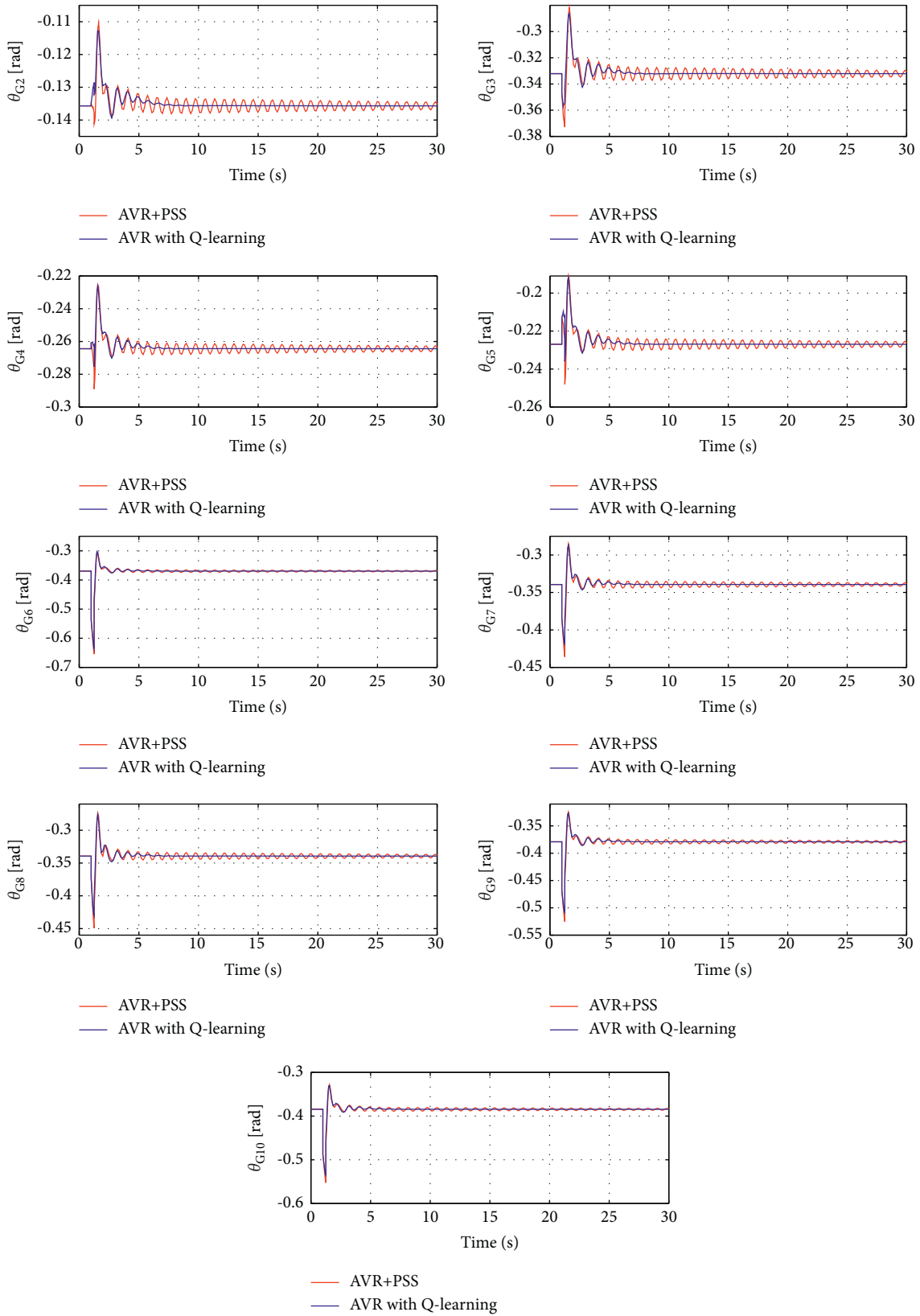
FIGURE 12: Voltage angle of generators referenced to the slack bus for three-phase fault scenario.

Figure 10 shows the voltage angle of all the generators referenced to the slack bus (bus 31). For the sake of effectiveness review and analysis, the results are compared with the conventional AVR and PSS combination. It is observed that the AVR systems with the help of the proposed Q-learning algorithm can damp oscillations faster, keep angle deviations within an acceptable margin, and stabilize the angle profile quickly. On the other hand, the AVR system with conventional PSS has a slower response in damping angle oscillations and it is not effective to limit magnitude deviations shortly after the contingency occurs. It is also obvious that G7 and G8 are not able to maintain their voltage angles at their prefault values when conventional AVR control is implemented.

In another scenario, a three-phase to ground fault is applied to the 39-bus test system. The fault occurs at bus 9, at $t = 1.0$ sec with a clearing time of 0.025 sec and fault resistance and reactance of 0.01 and 0.001 p.u., respectively. The behavior of the system under the fault condition is shown in Figures 11 and 12. In Figure 11, the terminal voltage of the generators is shown. Additionally, to better observe the effectiveness and robustness of the proposed control algorithm, the results are compared with the conventional combination of AVR and PSS. Again, it is demonstrated that after the three-phase fault is implemented, the AVR systems equipped with the Q-learning control algorithm smooth voltage deviations significantly in a very short period, damp terminal voltage oscillations, and improve voltage stability of the system. In contrast, in the system with only AVR + PSS, the terminal voltage of the generators drops significantly before the fault is cleared and the voltage oscillations and the instability remain in the system after the fault is cleared as well. In this situation, unlike in the N-1 scenario, the voltage magnitude oscillations reduce at a very slow rate as time elapses and are not eliminated even after a long time following the fault clearance. Therefore, in the case of a conventional AVR system without adaptive control, the system needs significant time to reach stability.

Figure 12 demonstrates the voltage angle of the generators referenced to the slack bus (bus 31). For the sake of effectiveness review and analysis, the results are compared with the conventional combined AVR and PSS. It is observed that the AVR systems with the support of the proposed Q-learning control can damp oscillations much faster and stabilize the angle profile quickly. However, with the combined AVR and PSS control approach, the oscillations in the angle profile result in instability with very high oscillating magnitudes. With the presence of the conventional AVR and PSS control, the voltage angles remain undamped after the fault is cleared and the system does not settle down. Unlike the N-1 contingency case, the magnitude of voltage angles does not reduce even after the fault is cleared and the system remains unstable.

5. Conclusion and Future Work

The application of a modified Q-learning algorithm for wide-area control of AVR systems is presented in this paper. The use of reinforcement learning with AVR systems has provided more stability than conventional power systems that use AVR and PSS. In this paper, a modified Q-learning control has been applied to the WSCC 9-bus and IEEE 39-bus test systems. The proposed control algorithm has been tested under three different critical contingency scenarios and the obtained results are very promising and satisfactory in providing voltage stability compared to the traditional combination of AVR and PSS. The proposed intelligent adaptive control of AVRs provides a fast-damping response, stable operation, and smooth transient voltage profile through all critical test conditions. It was demonstrated that the modified Q-learning algorithm can optimally compute the impact of system changes on the gains of the AVR systems to provide robust voltage stability in a wide-area interconnected performance rather than solely local impacts. Future work would focus on integrating intelligent PSS tuning and a self-adjustable AVR system and comparing them with the modified Q-learning approach.

Abbreviations

AVR:	Automatic voltage regulator
PSS:	Power system stabilizer
α:	Learning rate
γ:	Discount factor
$\mathbf{Q}_t(\mathbf{s}_t, \mathbf{a}_t)$:	First initialized (estimated) action a_t is selected in a state s_t
$\mathbf{R}_t$:	Expected cumulative rewards
$\mathbf{P}(\mathbf{a}/\mathbf{s})$:	SoftMax (Boltzmann exploration) action selection policy
$\mathbf{T}$:	Temperature parameter
$\mathbf{S}$:	State space of the terminal voltage of generators
$\mathbf{V}_{i,t}$:	i^{th} generator voltage at the time step t
$\Delta\mathbf{K}_{A,i}$:	Change in AVR gain of the generator i
$\mathbf{r}(\mathbf{s}_{i,t}, \mathbf{a}_{i,t})$:	Reward function
$\mathbf{K}_{A,i}$:	The gain of the amplifier at AVR i
$\mathbf{K}_{E,i}$:	The gain of the exciter at AVR i
$\mathbf{K}_{G,i}$:	The gain of the generator at AVR i
$\mathbf{K}_{P,i}$:	$\mathbf{K}_{I,i}$, $\mathbf{K}_{D,i}$ the proportional, integral, and derivative gains of the AVR i
$\mathbf{G}_{s,i}$:	The transfer function of the sensor system at AVR i
$\mathbf{E}_x(i, j)$:	AVR exciter variables.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Duan, H. Xu, and W. Liu, "Q-learning-based damping control of wide-area power systems under cyber uncertainties," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6408–6418, 2018.

[2] S. Mohagheghi, G. K. Venayagamoorthy, and R. G. Harley, "Optimal wide area controller and state predictor for a power system," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 693–705, 2007.

[3] F. Dorfler, M. Jovanovic, M. Chertkov, and F. Bullo, "Sparse and optimal wide-area damping control in power networks," in *Proceedings of the American Control Conference*, Washington, DC, USA, June 2013.

[4] F. Dorfler, M. R. Jovanovic, M. Chertkov, and F. Bullo, "Sparsity-promoting optimal wide-area control of power networks," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2281–2291, 2014.

[5] F. Okou, L. A. Dessaint, and O. Akhrif, "Power systems stability enhancement using a wide-area signals based hierarchical controller," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1465–1477, Aug 2005.

[6] H. Quinot, H. Bourles, and T. Margotin, "Robust coordinated AVR + PSS for damping large scale power systems," *IEEE Transactions on Power Systems*, vol. 14, no. 4, pp. 1446–1451, Nov 1999.

[7] M. E. C. Bento, "A hybrid particle swarm optimization algorithm for the wide-area damping control design," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 592–599, 2022.

[8] R. Hadidi and B. Jeyasurya, "Reinforcement learning based real-time wide-area stabilizing control agents to enhance power system stability," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 489–497, 2013.

[9] B. Abegaz and J. Kueber, "Smart control of automatic voltage regulators using *K*-means clustering," in *Proceedings of the IEEE Annual Conference System of Systems Engineering*, Anchorage, AK, USA, 2019.

[10] B. Abegaz, "Stability improvement of power networks using cluster-based dynamic excitation of AVR," in *Proceedings of the IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference*, New York, NY, USA, October 2019.

[11] A. K. Bhullar, R. Kaur, and S. Sondhi, "Modified neural network algorithm based robust design of AVR system using the kharitonov theorem," *International Journal of Intelligent Systems*, vol. 37, no. 2, pp. 1339–1370, 2021.

[12] R. Devarapalli, B. Bhattacharyya, and N. K. Sinha, "An intelligent EGWO-SCA-CS algorithm for PSS parameter tuning under system uncertainties," *International Journal of Intelligent Systems*, vol. 35, no. 10, pp. 1520–1569, 2020.

[13] N. Nahas, M. Abouheaf, A. Sharaf, and W. Gueaieb, "A self-adjusting adaptive AVR-LFC scheme for synchronous generators," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5073–5075, 2019.

[14] K. Kawabe, M. Masuda, and T. Nanahara, "Excitation control method based on wide-area measurement system for improvement of transient stability in power systems," *Electric Power Systems Research*, vol. 188, Article ID 106568, 2020.

[15] E. Köse, "Optimal control of AVR system with tree seed algorithm-based PID controller," *IEEE Access*, vol. 8, pp. 89457–89467, 2020.

[16] S. Ekinci and B. Hekimoglu, "Improved kidney-inspired algorithm approach for tuning of PID controller in AVR system," *IEEE Access*, vol. 7, pp. 39935–39947, 2019.

[17] A. Tabak, "Maiden application of fractional order PID plus second order derivative controller in automatic voltage regulator," *International Transactions on Electrical Energy Systems*, vol. 31, no. 12, 2021.

[18] N. Paliwal, L. Srivastava, and M. Pandit, "Equilibrium optimizer tuned novel FOPID-DN controller for automatic voltage regulator system," *International Transactions on Electrical Energy Systems*, vol. 31, no. 8, 2021.

[19] R. Khezri, A. Oshnoei, A. Yazdani, and A. Mahmoudi, "Intelligent coordinators for automatic voltage regulator and power system stabilizer in a multi-machine power system," *IET Generation*, Transmission & Distribution, vol. 14, no. 23, pp. 5480–5490, 2020.

[20] X. Luo, Y. Gao, S. Huang, Y. Zhao, and S. Zhang, "Modification of Q-learning to adapt to the randomness of environment," in *Proceedings of the International Conference on Control, Automation and Information Sciences (ICCAIS)*, Chengdu, China, 2019.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, London, UK, 2017.