Journal of Power Technologies 103 (4) (2023) 230 -- 242

Stabilizer Design of PSS3B based on the KH algorithm and Q-Learning for Damping of Low Frequency Oscillations in a Single-Machine Power System

Farshid Mohamadi, Alireza Sedaghati

Electrical Engineering Faculty of Shahabdanesh University, Qom, Iran
✉ eerfaculty@yahoo.com

Abstract

The aim of this study is to use the reinforcement learning method in order to generate a complementary signal for enhancing the performance of the system stabilizer. The reinforcement learning is one of the important branches of machine learning on the area of artificial intelligence and a general approach for solving the Marcov Decision Process (MDP) problems. In this paper, a reinforcement learning-based control method, named Q-learning, is presented and used to improve the performance of a 3-Band Power System Stabilizer (PSS3B) in a single-machine power system. For this end, we first set the parameters of the 3-band power system stabilizer by optimizing the eigenvalue-based objective function using the new optimization KH algorithm, and then its efficiency is improved using the proposed reinforcement learning algorithm based on the Q-learning method in real time. One of the fundamental features of the proposed reinforcement learning-based stabilizer is its simplicity and independence on the system model and changes in the working points of operation. To evaluate the efficiency of the proposed reinforcement learning-based 3-band power system stabilizer, its results are compared with the conventional power system stabilizer and the 3-band power system stabilizer designed by the use of the KH algorithm under different working points. The simulation results based on the performance indicators show that the power system stabilizer proposed in this study underperform the two other methods in terms of decrease in settling time and damping of low frequency oscillations.

Keywords: 3-band power system stabilizer, reinforcement learning, Q-learning.

Introduction

Stability of power systems is one of the most important aspects of the performance of the electricity network because the frequency and voltage of the power system must always be in their nominal values, even under extreme turbulences, such as a sudden cloudburst rise, a sudden outage of a generator, or getting out of a transmission line during an error. Power system can be envisioned as large and interconnected systems with very complicated dynamics. The connection between different components of power systems impose different oscillations on the whole system. Meanwhile, the low frequency oscillations (0.2-0.3 Hz) are very important because they continue for a long time after starting. Sometimes, in the absence of appropriate damping, oscillations ranges become larger and lead to instability in the power systems. Also, these

oscillations impose a lot of restrictions on the capability of transferring the power of the system [1]. To improve damping of the power system oscillations, the generator is equipped with PSS, which is capable of damping of these oscillations. In [2], the authors have designed PSS in a resistant form for the single-machine power system. In [3] and [4], the authors have shown the

superiority of the multiple-band power system stabilizer to the conventional system. In [5], a smart approach based on neural networks and artificial intelligence has been used which replaces in the place of AVR and PSS completely. Although this control method is resistant and adjustable, its implementation in practice needs a very complicated hardware, which makes its use impossible in practice. In [6], incorporating the characteristics of the neural networks and fuzzy logic, the authors have designed the neural networks and fuzzy logic-based PSS for damping of the power system stabilizer oscillations, which is a very complicated method and has a difficult design.

Many researchers have emphasized on an intelligent and systematic education to control power system. Intelligent agents can update their decision-making power at every moment [7, 8, 9]. This need can be addressed by the use of a computational method for learning, named reinforcement learning. The purpose of reinforcement learning for designing a controller is to make the automated and intelligent agents able to make decisions at each state of the system and act in order to increase their long-term compensations. In the recent decade, the reinforcement learning has found a special place in controlling the system power and has been successfully employed on topics such as small-signal stability, voltage stability, and electricity market. In [10], the principles of employing the reinforcement learning in controlling the stability of the power system have been investigated and the efficiency of this method in controlling TSCS for improving the oscillations of the power transmission between two regions of a four-machine system have been shown. It can be understood from the results of this study that the reinforcement learning is applicable to any system with any large size and any dynamical complexity. Also, this control method is resistant and adjusts itself with any changes in the conditions of the system. In [11], the authors have shown two application of the reinforcement learning. In the first, the reinforcement learning have been used to adjust the performance of the conventional power system stabilizer, and in the second, the power system stabilizer is completely replaced by the reinforcement learning. The both applications show that the reinforcement learning can complement the power system stabilizer or be an appropriate substitute for it. In [12], the reinforcement learning has been used to control the reactive power and compared with the CLF probabilistic methods, and the results establish the superiority of the reinforcement learning. [13], the authors show the applicability of the reinforcement learning in the topic of electricity market.

In this study, three types of control methods for the stability of the single-machine power system are investigated: conventional power system stabilizer (CPSS), 3-band power system stabilizer (PSS3B), and 3-band power system stabilizer with reinforcement learning (PSS3B-RL). PSS3B is designed in a resistant form under different points of operation by optimization of the objective function based on the damping coefficient and the damping ratio of the unstable electromechanical modes with weak damping using the new optimization KH algorithm in such a way that the unstable electromechanical modes with weak damping are transmitted to a specified area of the complex plane, and then, its controlling efficiency in an non-linear system based on the proposed reinforcement learning method “Q-learning” is improved in the real time in order to better damp the low frequency oscillations. The reinforcement learning has been used to generate a complementary signal for improving the performance of the 3-band power system

stabilizer. The control strategy combines the characteristics of the 3-band power system stabilizer and the Q-learning-based reinforcement learning, which leads to a simple and flexible control structure and is considered as a powerful method for damping of the low frequency oscillations and improvement of the dynamical stability of the power system. The

conventional power system stabilizer has been designed by the phase compensation method. Then, the results are compared and the superiority of the proposed control method are shown in terms of overshoot, undershoot, settling time, and ITAE, ISTSE, and ISE performance measures.

Types of Power System Stabilizers

Power system stabilizer is an electronic feedback control of the excitation system of the generation unit that its duty is to damp the oscillations and increase the rotor angle stability (PDF) limit of the power system by modulating the excitation voltage of the generator [14]. IEEE has defined different models for PSS. In this paper, the conventional PSS and PSS3B are investigated.

Conventional Power System Stabilizer (CPSS)

Figure 1 shows IEEE model of a PSS. Input of this stabilizer is signal of angular velocity changes, which is so-called CPSS.

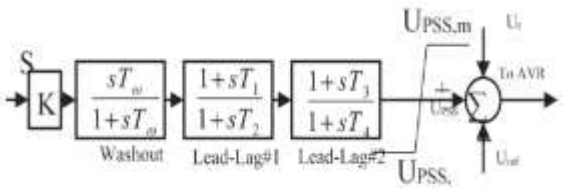


Figure 1: Structure of stabilizer CPSS.

This block of the diagram contains the block "washout" which leads to decrease in the response over than the tolerance of the system when a turbulence occurs. Since PSS must create electric torque in phase as the velocity changes, the pre-phase-post-phase block is employed. The number of the pre-phase-post-phase block depends on the nature of the system. In this study, two blocks are assumed. The extent of PSS damping is created by the gain K_ω. This stabilizer is very sensitive to noise and always contains torsional oscillations. In this research, for simplicity, T_1 is considered as equal to T_3, T_2 is considered as equal to T_4, and the stabilizer has been designed by the phase compensation method [15]. The data are presented in Table 1.

Table 1: Data related to the CPSS used in this paper.

Name	Value
K_ω	12.5
$T_\omega(s)$	5
$T_1(s)$	0.0738
$T_2(s)$	0.028
$U_{PSS,}^{min}[pu]$	-0.15
$U_{PSS,}^{max}[pu]$	0.15

Power System Stabilizer PSS3B

The IEEE model of the stabilizer PSS3B is shown in Figure 2. The stabilizer PSS3B uses two inputs: electrical power changes (ΔP) and rotor angle velocity changes $\Delta\omega$. In this stabilizer, T_1 and T_3 are time constants of the convertor and T_2 and T_4 are time constants of the torsional filters. The optimal stabilizer gain is obtained by adjusting K_1, K_2, and K_3. Also, T_{1n}, T_{1d}, T_{2n}, and T_{2d} are the coefficients of the phase compensator of the stabilizer. In the stabilizer output, also, an excitation voltage limiter is used [16]. In the structure of Figure 2, $T_2 = T_4 = 10$, and for simplicity, we assumed that $T_{1n} = 0.02$, $T_{1d} = 0.01$, $T_{2n} = 0.03$, and $T_{2d} = 0.01$. Therefore, for the stabilizer PSS3B, the parameters K_1, K_2, K_3, T_1, and T_3 are adjustable.

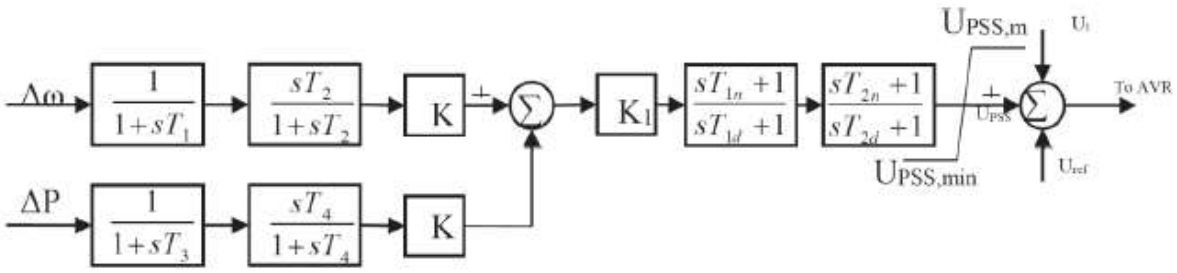


Figure 2: Structure of the stabilizer PSS3B.

The problem of optimizing the parameter setting of this stabilizer are defined and resolved using the KH algorithm.

Overview on Krill Herd Algorithm

The KH algorithm has a structure similar to PSO, but according to [17], it performs better than PSO does and is also easily applicable. Thus, in this study, this algorithm is used to resolve the optimization problem. Krill is the name of a crustacean that is found in the waters of the whole world. These crustaceans are always on motion in large swarms, and the KH algorithm has been inspired by the rules governing their mass movements. In the KH algorithm, it is assumed that the motion of each krill particle is influenced by three factors: motion caused by other particles, exploratory motion (searching for food), and random dissemination. The Lagrange equation shown in the formula 1 models this assumption mathematically:

$$\frac{dX_i}{dt} = N_i + F_i + D_i \tag{1}$$

Where N_i is motion caused by other particles, F_i is exploratory motion, and D_i is physical dissemination of i-th particle. The motion caused by other particles is expressed according to the formulas (1) and (2):

$$N_i^{new} = N^{max}\alpha_i + \omega_n N_i^{old} \tag{2}$$

Where

$$\alpha_i = \alpha_i^{local} + \alpha_i^{target} \tag{3}$$

In the above relations, N^{max} is the maximum induced velocity, ω_n is the inertia coefficient of the induced motion, which is a number in $[0,1]$, N_i^{old} is the previous induced motion, α_i^{local} is the local effect caused by neighbors, and α_i^{target} is the effect of the target direction that is caused by the best particle. The exploratory motion is formulated based on two main variables; the first is food position and the second is the previous experience of the food position. This disposition is defined by the formulas (4) and (5):

$$F_i = V_f\beta_i + \omega_f F_i^{old} \tag{4}$$

Where

$$\beta_i = \beta_i^{food} + \beta_i^{best} \tag{5}$$

In the above relations, V_f is the exploration velocity, ω_n is the inertia coefficient of exploration that is a number in $[0,1]$, β_i^{food} is the food absorption coefficient, and β_i^{best} is the best experience position of i-th particle so far. The physical dissemination of the krill particles is considered as a random process and defined by the formula (6):

$$D_i = D^{max}\delta \tag{6}$$

Where D_{max} is the maximum dissemination velocity and δ is a directed random vector between $[0,1]$. Finally, the new position of i-th krill particle at the moment $t + \Delta t$ is calculated by the formula (7):

$$X_i(t + \Delta t) = X_i(t) + \Delta t\frac{dx_i}{dt} \tag{7}$$

It should be noted that the fixed parameter Δt is very important and must be determined carefully with regard to the optimization problem. Since the value of this quantity depends on the search space, it can be determined by the formula (8):

$$\Delta t = C_t \sum_{j=1}^{NV} \left(UB_j - LB_j\right) \tag{8}$$

Where NV is the total number of variables, UB_j and $\mathrm{LB_j}$ are the upper and lower limit of j-th variable, respectively, and C_t is a constant between $[0,2]$, which allows the krill particles to search the search space carefully. The flowchart of the KH algorithm is shown in Figure 4 by blue boxes. For more details on the KH algorithm, the reader is referred to [17].

Cost Function

In order that the stabilizer to be resistant against changes in the working points of the system, the optimization has been performed with respect to changes in P_t, Q_t, and X_e within defined limits. The working points used for optimization are as following:

- Active power (P_t): from 0.4 to 1 by steps of 0.1;
- Reactive power (Q_t): from -0.2 to 0.5 by steps of 0.1;
- Line reactance (X_e): from 0.2 to 0.7 by steps of 0.1.

The cost function is calculated as follows: for each working point, the system is linearized, the eigenvalues of the closed loop system are obtained, and the objective function is calculated using the unstable eigenvalues or less damped eigenvalues of the system that need to be displaced toward the complex plane. Formula (9) shows the objective function used in this study:

$$J = \sum_{j=1}^{np} \sum_{\sigma_{ijz}\sigma_0} \left[\sigma_0 - \sigma_{ij}\right]^2 + a \sum_{j=1}^{np} \sum_{\zeta_{ij}\geq\zeta_0} \left[\zeta_0 - \zeta_{ij}\right]^2 \tag{9}$$

Where np is the number of working points, σ is the real part of eigenvalues, ζ is the damping coefficient, and a is the weight coefficient. In relation (1), we assume that $a = 10, \sigma_0 = -1$, and $\zeta_0 = 10\%$. Figure 3 describes the objective function of formula (9). For more details, refer to [18].

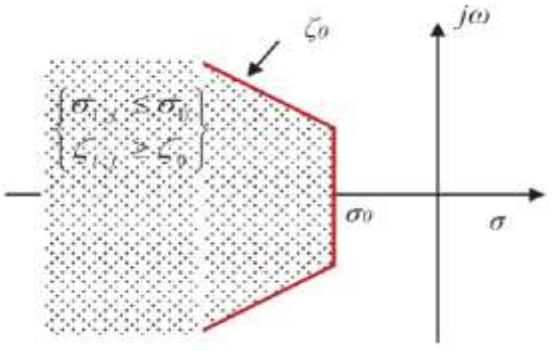


Figure 3: Area specified for objective function.

The process of designing the destabilizer PSS3B can be considered as an optimization problem with following restrictions:

Minimize J

subject to:

$$\begin{aligned} &K_1^{\min} < K_1 < K_1^{\max}, K_2^{\min} < K_2 < K_2^{\max}, \\ &K_3^{\min} < K_3 < K_3^{\max}, T_1^{\min} < T_1 < T_1^{\max} \\ &T_3^{\min} < T_3 < T_3^{\max} \end{aligned} \tag{10}$$

The results from resolving the optimization problem of formula (10) is shown in Table 2.

Table 2: Results from optimization of PSS3B.

Name	Value
K_1	2.0304
K_2	13.7193
K_3	0.5
T_1	0.1002
T_3	0.01

Reinforcement Learning

The reinforcement learning is a method in which one or more agents in interaction with the environment learn an optimal control policy in order to realize a predefined objective. Optimal policy refers to selecting the best action among the available ones for each position of the agent in the environment. In general case, the agent does not have an initial knowledge about the environment and learns the control policy using a trial and error method. The reinforcement learning methods can control any non-linear system without simplifying. Q-learning is one of the well-known reinforcement learning methods used in this research. The reason for using it is its simplicity and independence on the system model. Some of the salient features of the Q-learning-based controllers can be their independence on the system model, their resistance in changing the operation conditions and uncertainty of the system parameters, their adaptive control, and their implementation simplicity. This control method can be used as an appropriate complement for the traditional control methods; in this paper, this feature has been used and the efficiency of the power system has increased by utilizing Q-learning and creating a complementary signal. The reinforcement learning assumes that the environment (control system) has been divided into limited states and is shown by {S}. At each step t, the agent sees itself in the state st and select the action a among a set of available actions {A}. The agent receives a reward as soon as it does an action. The given reward is defined in such a way that shows the satisfaction with the performed action. Then, the agent sees itself in the state st+1, select the appropriate action again and this trend continues until the specified objective is realized. The purpose of the reinforcement learning is to find a policy, a map between the states and actions of the system, and as a result, the decreased long-term reward reaches its maximum. The decreased long-term reward of the system is given as follows:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \qquad (11)$$

Where γ is a number between $[0,1]$ and is called "reduction factor". This factor shows the importance of the future rewards in decision-making. If its value is assumed 0 , then the next rewards will be ineffective in the decision-making process, and if its value is assumed 1 , then the next rewards will be effective in the decision-making process. The reinforcement learning has a value function, which is called "Q function" in Q-learning and defined as:

$$Q^{\pi}(s.a) = E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid S_t = S.a_t = a\right\} \qquad (12)$$

Where π is the control policy, s is the current state, a is the selected action, and r is the reward received from the environment. The reinforcement learning method finds the optimal policy π^* in such a way that the value of the Q function of formula (12) reaches its maximum. In the Qlearning method of the reinforcement learning at each time step, formula (12) is updated in the interaction with the environment. The updated relation is called "Bellman optimal equation" and is given by formula (13):

$$\Delta Q = \alpha[r_{r+1} + \gamma_a^{\max} Q(S_{t+1} \cdot a_{t+1}) - Q(s_t \cdot \alpha_t)] \qquad (13)$$

Where α is a number between $[0,10]$ and is called "attenuation coefficient", which represents the real error value. In Table 3, the implementation steps of the Q-learning algorithm is shown briefly.

The semi-greedy algorithm is used to select the action in each state as follows: with the probability $1-\varepsilon$, the action with higher Q is selected, and with the probability ε, an action is randomly selected among all actions.

Table 3: Step-by-step implementation of the Q-learning algorithm.

a. Find the optimal policy.
a.1. Define the set of states, actions, and rewards.
a.2. Determine the values of γ, α, and ε.
a.3. Initializing $Q(s,a) = 0$ for all states and actions.
a.4. For each run (episode):
a.4.1. Calculate the current state (s) of the system.
a.4.2. Repeat until reaching the goal.

a.4.2.1. Select the action a among the available actions of the system for the state (s) using the semi-greedy algorithm (ε-greedy).
a.4.2.2. Do the action a and receive the reward (r) and the next state.
a.4.2.3. Update the function Q using the below relation:

$$Q(s.a) = Q(s.a) + \Delta Q \quad (14)$$

a.4.2.4. Set the next state of the system as the current state.
End of repetition (go to a.4.2)
End of a run (go to a.4)

b. Implement the optimal policy.
b.1. For the current state of the system, select the action that maximizes the value of the function Q.
b.2. Select the next state of the system and set it as the current state.
b.3. Go to $b.1$ and continue this process.

Reinforcement Learning Parameters Used in This Study

States

The main purpose of using stabilizer is to damp the low frequency oscillations in the power system. In other words, the stabilizer must damp the power oscillations or frequency oscillations. Thus, ΔP and/or $\Delta\omega$ and/or the combination of these two can be used as system states. In this study, $\Delta\omega$ has been considered as a state. The interval $[-0.01, +0.01]$ (in the pre-unit system) has been divided into 12 parts, each of which represents a state of the system. Therefore, the set of states can be defined as following:

$$S = \{(-\infty, -0.01], (-0.01, -0.0082], (-0.0082, -0.0064], (-0.0064, -0.0027], (-0.0027, -0.0009], (-0.0009, 0.0009], (0.0009, 0.0027], (0.0027, 0.0045], (0.0045, 0.0064], (0.0064, 0.0082], (0.0082, 0.01], (0.01, +\infty]\} \quad (15)$$

In the set of states, the state (-0.0009, 0.0009] is considered as the normal state.

Actions

Defining the set of actions is very complicated and important. Given the restrictions placed on the output of the stabilizer, the range (limit) of these actions can be estimated. According to [19], it can be estimated that actions are in the interval [-0.2, 0.2]. For simplicity, the set of actions is defined as following:

$$A = \{-0.2, 0.2\} \quad (16)$$

Reward

In general, the aim of designing an stabilizer is to damp the power and frequency oscillations, the choice of the intended reward is set to the distance $\Delta\omega$ from 0 in two time steps t and t-1. In this study, it is assumed that each action is applied to the system for 50 milliseconds, and then, the next state and reward are calculated. Thus, the reward is given by:

$$Reward_t = \sum_{k=t-1}^{t} \Delta\omega(t) \quad (17)$$

Also, the values of $\alpha, \varepsilon\ and\ \gamma$ are assumed as 0.02, 0.05, and 0.98, respectively. In Figure 4, the power system model of this study and the way by which the reinforcement learning is applied to the power system stabilizer in order to optimize its performance is shown.

Simulation Results

To show the performance of the proposed control method, the computer simulations are performed by the use of the software MATLAB. The studied system is shown in Figure 4 and the details of its components is available in [10]. The simulations for a three-phase error of 100 ms in the generator bus have been done in different operation conditions. Figures 5 and 6 show the results from occurring a three-phase error of 100 ms in the generator bus at two different working situations.

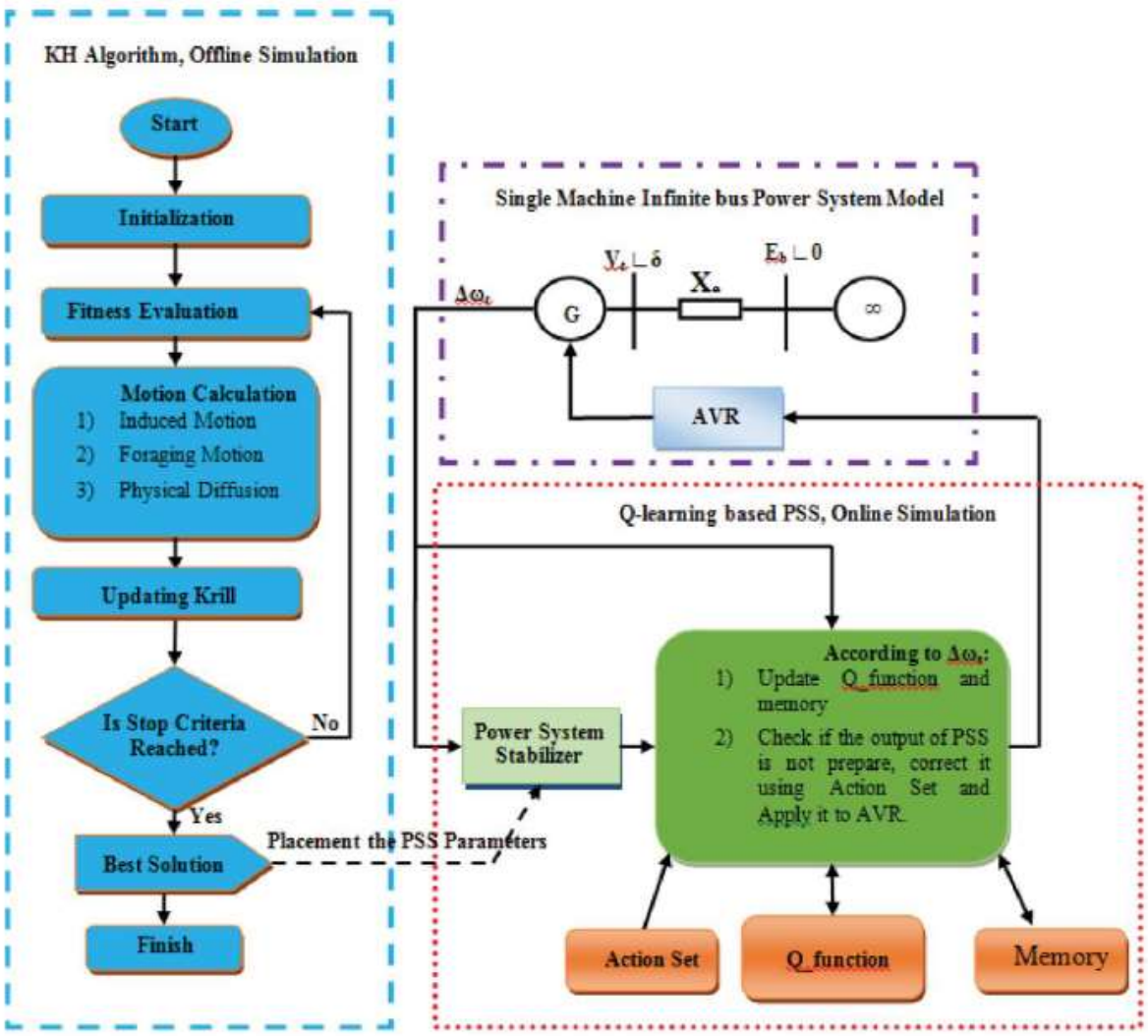
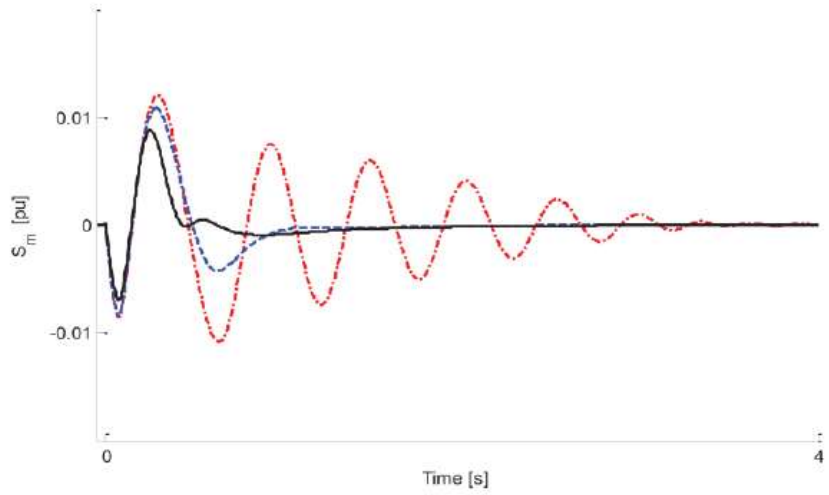
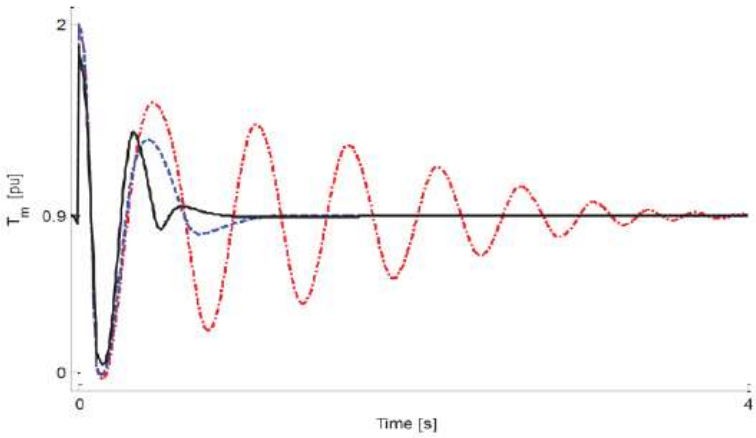


Figure 4: The studied power system model and the proposed method for optimizing the performance of the power system stabilizer using Q-learning and the Krill Heard algorithm. Blue dashed line: flowchart of the KH algorithm; dotted line: Q-learning-based stabilizer; violet dashed line: the studies power system model.



A



B

Figure 5: Results from a three-phase error of 100 ms in the working situation: P_t = 1.2, Q_t = 0.2, X_e = 0.7, and T_m = 0.9. A) S_m. B) T_m. Integrated: PSS3B + RL. Dashed line: PSS3B. Dotted line: CPSS.

These figure obviously show that the proposed control method, in dependent of the operation conditions, is very more effective than the conventional stabilizer and the stabilizer PSS3B. In the following, the simulation results of four different working situations are investigated statistically and the measures of overshoot, undershoot, ITAE, ISTSE, and ISE for the changes $\Delta\omega$ are calculated in Table 4 in order to compare the performance of stabilizers at different working situations. As the data in Table 4 show, it can be concluded that the Q-learning-based complementary control method has improved the performance of PSS3B independent of different working situations. The value of overshoot at each of four working situation has been improved by about 20% compared to PSS3B without complementary control. The value of undershoot at the working situations 1 and 2 and at the working situations 3 and

4 has been improved by about 16% and 13%, respectively. The settling time at the working situations 1 and 3, the working situation 2, and the working situation 4 has been increased by about 0.1, 0.7, and 0.33, respectively.

The measure ITAE has been increased by about 40% at the working situations 1 and 2 and 60% at the working situations 3 and 4. The measure ISTSE has been also improved by about 95%, 56%, 69%, and 71% at the four working states, respectively. Therefore, Figures 5 and 6 and also the data analysis of Table 4 prove that the proposed reinforcement learning-based control method is more effective than other methods.

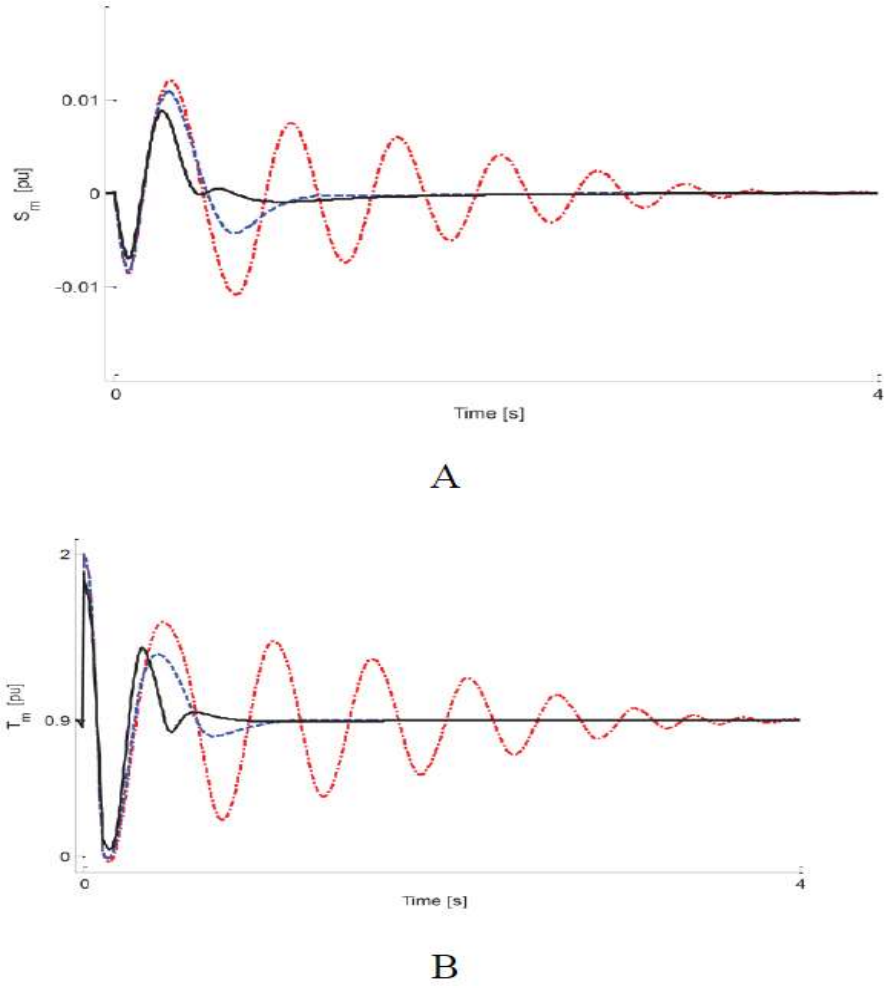


Figure 6: Results from a three-phase error of 100 ms in the working situation: P_t = 0.7, Q_t = 0.5, X_e = 0.3, and T_m = 1.2. A) S_m. B) T_m. Integrated: PSS3B + RL. Dashed line: PSS3B. Dotted line: CPSS.

Table 4: Comparison of different types of PSS at different working situations.

PSS3B+RL	PSS3B	CPSS	Measure	Working situation	
0.88	1.09	1.2	%OS		
0.7	0.842	1.08	%US		
1.009	1.0101	3.2784	Ts [S]	T_m = 0.9, P_{t0} = 1.2	1
0.0248	0.0401	0.1262	ITAE	Q_{t0} = 0.2, X_e = 0.7	
0.001	0.0024	0.0082	ISTSE		
1e-6	2.2e-5	7.08e-5	ISE		
0.87	1.08	1.19	%OS		
0.6956	0.8376	1.06	%US		
1.3556	1.4599	3.2483	Ts [S]	T_m = 0.9, P_{t0} = 1.2	2
0.0247	0.0399	1.1224	ITAE	Q_{t0} = 0.2, X_e = 0.7	
0.0008	0.0022	0.0078	ISTSE		
9.97e-6	2.26e-5	6.77e-5	ISE		
1.37	1.69	1.83	%OS		
1.22	1.39	1.73	%US	T_m = 0.9, P_{t0} = 1.2	
1.3502	1.5069	10.9621	Ts [S]	Q_{t0} = 0.5, X_e = 0.3	3
0.0410	0.1026	0.4227	ITAE		
0.003	0.0106	0.0557	ISTSE		
2.85e-5	9.72e-5	3.98e-4	ISE		
1.37	1.7	1.84	%OS		
1.23	1.42	1.76	%US	T_m = 1.2, P_{t0} = 1	
1.3388	1.7864	10.5223	Ts [S]	Q_{t0} = 0, X_e = 0.5	4
0.0405	0.1027	0.4214	ITAE		
0.0029	0.0107	0.00554	ISTSE		
2.81e-5	9.74e-5	9.96e-4	ISE		

Conclusion

In this study, the power system stabilizer PSS3B was designed by the use of the KH smart algorithm. To design a resistant controller, the used objective function shifted the unstable or less damped eigenvalues of the system towards stability and dampness. After optimal setting of the PSS3B

parameters, its performance was optimized in real time using the proposed Q-learning-based reinforcement learning algorithm. Some of the fundamental features of the proposed reinforcement learning-based combined stabilizer is its simplicity, its independence of the system model, and its resistance against disturbances imposed on the system and changes in the working points of operation. After applying the proposed control method, the performance of three kinds of control method, i.e. the conventional power system stabilizer, PSS3B, and PSS3B+RL was evaluated by imposing the system to disturbances at different operation conditions and applying the three-phase error. The simulation results showed that by combining the features of the 3-band power system stabilizer and the Q-learning-based reinforcement learning, the proposed control method leads to a simple and flexible control structure and has a high ability to damp the low frequency oscillations and improve the dynamical stability of the power system. To show the superiority of the proposed stabilizer, the simulation results were compared at different working situations in terms of the values of overshoot, undershoot, settling time, ITAE, ISTSE, and ISE and the results apparently confirm the superiority of the proposed method. Therefore, it can be concluded that the reinforcement learning can a complement of and even an appropriate substitute for power system stabilizers.

Appendix

Formula (18) shows the dynamical equations of the power system used in this study:

$$\frac{d\delta}{dt} = \omega_n S_m$$

$$\begin{aligned}
\frac{dS_m}{dt} &= \frac{1}{2H}\left[-DS_m + T_m - T_e\right] \\
\frac{dE'_d}{dt} &= \frac{1}{T'_{q0}}\left[-E'_d - \left(X_d - X'_q\right)i_q\right] \\
\frac{dE'_q}{dt} &= \frac{1}{T'_{d0}}\left[-E'_q - \left(X_d - X'_q\right)i_q + E_{fd}\right] \\
\frac{dE_{fd}}{dt} &= \frac{1}{T_a}\left[K_a\left(V_{ref} + V_s - V_t\right) - E_{fd}\right] \\
\begin{bmatrix}U_d\\U_q\end{bmatrix} &= \begin{bmatrix}E'_d\\E'_q\end{bmatrix} - \begin{bmatrix}0 & X'_q\\-X'_d & 0\end{bmatrix}\begin{bmatrix}i_d\\i_q\end{bmatrix} \\
\begin{bmatrix}i_d\\i_q\end{bmatrix} &= \begin{bmatrix}0 & X_e\\-X_e & 0\end{bmatrix}^{-1}\left[\begin{bmatrix}U_d\\U_q\end{bmatrix} + E'_b\begin{bmatrix}\sin\delta\\-\cos\delta\end{bmatrix}\right]
\end{aligned} \tag{18}$$

The equations of the measures used in this study is given by formula (19):

$$\begin{aligned}
\text{ITAE} &= \int_0^{t_{sim}} t|\Delta\omega|dt \\
\text{ISTSE} &= \int_0^{t_{sim}} t^2\Delta\omega^2 dt \\
\text{ISE} &= \int_0^{t_{sim}} \Delta\omega^2 dt
\end{aligned} \tag{19}$$

References

[1] Anderson, M. and Fouad, A. A., Power system control and stability, Ames: IA: Iowa State Univ. Press, 1977.

[2] Dehghani M, Nikravesh S, Karrari M. “Decentralized Robust Power System Stabilizer Design”, Journal of Iranian Association of Electrical and Electronics Engineers, Vol. 4, No. 1, pp. 36-43, 2007.

[3] Khodabakhshian, A., Hemmati R. and Moazzami M., "Multi-band power system stabilizer design by using CPCE algorithm for multi-machine power system," Electric Power Systems Research, Vol. 101, pp. 36-48, 2013.

[4] He, P., Wen, F., Ledwich, G., Xue, Y. and Wang, K., "Effects of various power system stabilizers on improving power system dynamic performance," Electrical Power and Energy Systems, Vol. 46, pp. 175-183, 2013.

[5] Farahani, M., "A multi-objective power system stabilizer," IEEE Transactions on Power Systems, Vol. 28, No. 3, pp. 2700-2707, 2013.

[6] Malik, O. P. and Hariri, A., "Power system stabilizer based on a self-learning adaptive network fuzzy inference system," Transactions of the Institute of Measurement and Control, vol. 24, no. 2, pp. 153-173, 2002.

[7] Taylor, C. W., "Response-based, feedforward wide-area control," in NSF/DOE/EPRI Sponsored Workshope on Future Research Directions for Complex Interactive Networks, Washington DC, USA, 2000.

[8] Liu, C. C., Jung, J., Heydt, G. T. and Vittal, V., "The strategic power infrastructure defense (SPID) system," IEEE Control System Magazine, pp. 40-52, 2000.

[9] Diu, A. and Wehenkel, L., "EXaMINE-Experimentat on of a monitoring and control system for managing

vulnerabilitis of the european infrastructure for electrical power exchange," in IEEE PES Summer Meeting, Chicago, USA, 2002.

[10] Ernst, D., Glavic, M. and Wehenkel, L., "Power system stability control: Reinforcement learning framwork," IEEE Transaction on Power Systems, Vol. 19, No. 1, pp. 427- 435, 2004.

[11] Yu, T. and Zhen, W. G., "A reinforcement learning approach to power system stabilizer," in IEEE Power & Energy Society General Meeting, Calgary, AB, 2009.

[12] Vlachogiannis, J. G. and Hatziargyriou, N. D., "Reinforcement learning for reactive power control," IEEE Transaction on Power Systems, Vol. 19, No. 3, pp. 1317-1325, 2004.

[13] Naduri, V. and Das, T. K., "A reinforcement learning model to assess market power under auction-based energy pricing," IEEE Transaction on Power Systems, Vol. 22, No. 1, pp. 85-95, 2007.

[14] Safari, A., Shayeghi, H., Jalilzadeh, S. , “Robust Coordinated Design of UPFC Damping Controller and PSS Using Chaotic Optimization Algorithm”, Journal of Iranian Association of Electrical and Electronics Engineers, Vol. 12, No. 3, pp. 55-62, 2015.

[15] Singh, R., "A novel approach for tuning of power system stabilizer using genetic algorithm," Master of Sience dissertaition, Department of Electrical Engineering, Indian Institute of Science, Bangalor, India, 2004.

[16] IEEE recommended practice for excitation system models for power system stability studies. [Performance]. IEEE Standard 421.5-2005, 2006.

[17] Gandomi, A. H. and Alavi, A. H., "Krill herd: A new bioinspired optimization algorithm," Commun Nonlinear Sci Number Simulat, Vol. 17, pp. 4831-4845, 2012.

[18] Abdel-Magid, Y. L. and Abido, M. A., "Optimal multiobjective design of robust power system stabilizers ssing genetic algorithms," IEEE Transaction on Power Systems, Vol. 18, No. 3, pp. 1125-1132, 2003.

[19] Padiyar, K. R., Power System Dynamics, Giniraj Lane, Sultan Bazar, Hyderabad: BS Publications, 2008.