Electrical Power and Energy Systems 63 (2014) 940–949

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Application of policy iteration technique based adaptive optimal control design for automatic voltage regulator of power system

Lal Bahadur Prasad *, Hari Om Gupta, Barjeev Tyagi

Department of Electrical Engineering, Indian Institute of Technology Roorkee, Uttarakhand 247667, India

ARTICLE INFO

Article history:
Received 12 July 2013
Received in revised form 24 June 2014
Accepted 26 June 2014
Available online 23 July 2014

Keywords:
Actor–critic
Adaptive optimal control
AVR
LQR
Policy iteration technique
Single-area power system

ABSTRACT

This paper presents an adaptive optimal control design approach for automatic voltage regulator (AVR) system using policy iteration technique based adaptive critic scheme. The infinite horizon optimal control design using linear quadratic regulator (LQR) by solution of algebraic Riccati equation (ARE) and Hamilton–Jacobi–Bellman (HJB) equation require the complete knowledge of the system dynamics, and giving solution offline. Policy iteration technique which is based on actor–critic structure consists of two-step iteration: policy evaluation and policy improvement. The online policy iteration technique based control scheme gives online the continuous-time adaptive optimal control solution without using the complete knowledge of the system's internal dynamics. The knowledge of systems internal dynamics (i.e. matrix A) is not needed for evaluation of cost or the update of control policy; only the knowledge of input matrix B is required for updating the control policy. Thus this control scheme becomes partially model-free. The proposed control scheme has been implemented for AVR system for its both models neglecting and including sensor dynamics. The simulation results and performance analysis are presented to justify the robustness & effectiveness of control scheme. The comparative performance analysis of adaptive critic control scheme and LQR is also presented.

Introduction

The industrial and domestic electrical appliances are designed to operate at a certain voltage and frequency rating and thus their performance is dependent on the quality of power supply. The performance of equipments is adversely affected and possibly may cause them damage if there is prolonged operation of the equipments outside the allowable range of voltages. Thus the terminal voltage of all the equipments in the system should be maintained within acceptable limits. Automatic voltage regulator (AVR) is an essential system in an electric power system which maintains the magnitude of terminal voltage of a synchronous generator at a specified level by its field excitation control. The incremental model of an AVR system is a continuous-time linear time-invariant (LTI) dynamical system.

The basic concept of AVR system for electric power system is discussed in [1–3]. Recently the control design using various approaches for automatic voltage regulator system and power system stabilizers has attracted researchers [4–14]. PSO-PID controller design and comparison of performance with GA-PID for AVR system using a new performance criterion for obtaining optimal controller parameters is presented in [4]. The performance of intelligent fuzzy based coordinated control of the automatic generation control (AGC) loop and the excitation loop equipped with proportional integral derivative (PID) controlled automatic voltage regulator (AVR) system and power system stabilizer (PSS) controlled AVR system is investigated in [5] using particle swarm optimization (PSO) termed as craziness based particle swarm optimization (CRPSO) as optimizing tool to get optimal tuning of PSS parameters as well as the gains of PID controllers and for on-line, off-nominal operating conditions to obtain the off-nominal optimal gains of PID controllers and parameters of PSS, Takagi Sugeno fuzzy logic (TSFL) has been applied. Chatterjee et al. [6] presents PID controlled AVR system, and PSS controlled AVR system using velocity update relaxation PSO (VURPSO) and position, velocity updating strategy and craziness PSO (CRPSO) also compares the performance with GA based approach. A fractional order (FO) $PI^{\lambda}D^{\mu}$ controller for an AVR system using an improved evolutionary non-dominated sorting genetic algorithm II (NSGA II) that is augmented with a chaotic map for greater effectiveness for the multi-objective optimization problem, is presented in [7]; and a frequency domain design approach for a fractional order PID (FOPID) controller for an AVR system using NSGA-II augmented with a chaotic henon map that is used for the multi-objective optimization based design

* Corresponding author.
E-mail addresses: erlbprasad@gmail.com, ibpeedee@iitr.ac.in (L.B. Prasad).

http://dx.doi.org/10.1016/j.ijepes.2014.06.057
0142-0615/

procedure, is presented in [8]. The steady state voltage stability assessment of power systems with automatic voltage regulator voltage limits is discussed in [9]. The effect of time delays on the stability of generator excitation control system is investigated in [10]. A parameter tuning method for AVR system using on-line measured data of the excitation control system with parameter optimization technique is presented in [11]. The design of power system stabilizer (PSS) for a small-signal stability study using sliding mode control (SMC) techniques is presented in [12]. The design of power system stabilizer using single network adaptive critic (SNAC) is discussed in [13]. In [14] the design of a sub-optimal nonlinear feedback controller for power systems based on the approximate solution of the Hamilton–Jacobi–Bellman (HJB) equation is presented. The conventional and recent control schemes for AVR system present in the literature are generally off-line and not giving adaptive and optimal control solution at the same time in real situation. Thus the investigation of adaptive optimal control solution for AVR system is desired for automation of voltage control in electric power system.

The performance of the controlled systems is desired to be optimal which should be valid also when applied in the real situation. There are many optimization & optimal control techniques and also many adaptation & adaptive control techniques which are present in the literatures for linear & nonlinear dynamical systems [15–18]. Adaptive control is an on-line design approach, and which is able to deal with uncertainties is generally not optimal in the sense of minimizing a formal performance function as specified for optimal control. Optimal control is off-line, and needs the knowledge of system dynamics for its design. Thus, to have both features of control design, it is desired to design online adaptive optimal control. Adaptive optimal control is designed either by adding optimality features to adaptive control (e.g. the adaptation of control parameters is done by seeing the desired performance improvement reflected by an optimality criterion functional) or by adding adaptive features to optimal control (e.g. the optimal control policy is improved relative to the adaptation of the parameters of system model).

Recently several researchers have tried to explore the intelligent computational techniques with adaptive and optimal control design by applying certain methodologies for certain applications [4–6,17–19]. The reinforcement learning (RL) became a third approach to adaptive optimal control which is strongly related with direct and indirect adaptive optimal control methods from a theoretical point of view. Adaptive/Approximate Dynamic Programming (ADP) [20–22] combined with RL, and backprogation [23] lead to the concept of adaptive critic designs (ACDs) which utilize two parametric structures known as the actor and the critic. There are several versions of adaptive critic designs present in the literature [23,24]. Policy Iteration (PI), a computational intelligence technique belongs to a class of RL algorithms based on actor–critic structure solves HJB equation by direct approach starting by evaluating the cost of a given initial admissible control policy. PI algorithms consists two-step iteration: policy evaluation and policy improvement [25–30]. The policy iteration technique based adaptive critic scheme gives adaptive optimal control solution of an infinite horizon optimal control problem without knowledge of the system internal dynamics.

Adaptive optimal control using various approaches has been presented in [13,19,22,26–35]. Adaptive optimal control by adaptive critic design using policy iteration technique for solving online the optimal control problem without using explicit knowledge of internal dynamics have been presented for nonlinear systems in [26–28], and for linear systems in [28–30]. In general, neural networks are used for parametric function approximations in actor–critic design [19,34], for which other choices are 'single network adaptive critic (SNAC)' [13,31,32], Takagi–Sugeno (T–S) fuzzy systems [33], PSO [35] etc. As from applications point of view PI algorithm is implemented for optimal load frequency control of a power system in [29], for F-16 aircraft in [28,30]. Thus the state-of-the-art motivates the investigations on adaptive critic design methods for adaptive optimal control of dynamical systems. Even certain recent papers have appeared in literature on adaptive critic designs and policy iteration technique with certain applications, the comprehensive performance investigation with practical applications is much desired.

In this paper the novel application of policy iteration technique based adaptive critic scheme is presented for adaptive optimal control design for AVR system for a single-area power system. This paper contributes presenting the comprehensive performance investigation of adaptive optimal control using policy iteration technique based adaptive critic scheme with application to control of a practical system. The comparative performance investigation of adaptive critic control scheme and linear quadratic regulator is also presented. The modeling, simulation results and analysis are presented for both models of AVR system neglecting and including sensor dynamics and also for cases with change in system parameter during simulation. The proposed controller adapts the change in system parameters in real situation at any moment of time. It is demonstrated that the proposed control scheme provides a promising adaptive optimal control solution for dynamical systems without complete knowledge of the system dynamics and thus it is a partially model free approach.

This paper is organized in 5 sections. Section 'Introduction' presents the relevance & the general introduction of the paper. Section 'Mathematical modeling of power system automatic voltage regulator' presents the mathematical modeling of automatic voltage regular system for both cases of neglecting and including sensor dynamics. Section 'Adaptive optimal control using policy iteration technique' describes the adaptive optimal control using policy iteration technique based adaptive critic scheme for continuous-time LTI systems. The simulation results and performance analysis of policy iteration technique based adaptive optimal control of AVR system for both cases of neglecting and including sensor dynamics and also with change in system parameters have been presented in section 4. Section 5 presents the conclusion. At the end a brief list of references is given.

Mathematical modeling of power system automatic voltage regulator

Automatic voltage regulator (AVR) maintains the generator terminal voltage and controls the reactive power flow by controlling generator field excitation. A simple AVR system comprises four main components, namely amplifier, exciter, generator, and comparator & sensor. Since the time constant of sensor is normally very small thus it may be neglected for a simplified mathematical model of AVR system of a single-area power system. Fig. 1 shows the functional block diagram of a simple AVR system neglecting the sensor dynamics [1–3]. Fig. 2 shows the functional block diagram of an AVR system including sensor dynamics [4,6–8,10]. Considering the major time constant and ignoring the saturation or other nonlinearities, and deviated from a normal state, the linearized incremental mathematical modeling of AVR system is presented as following [1–8,10].

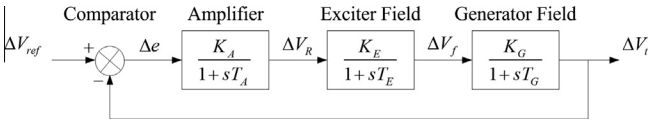

Fig. 1. Block diagram of automatic voltage regulator neglecting sensor dynamics.

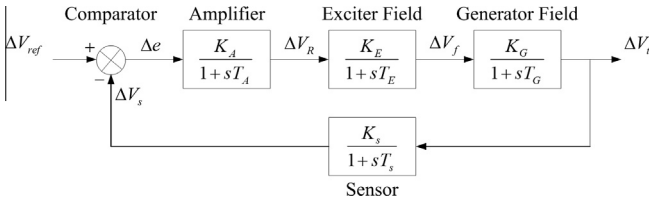

Fig. 2. Block diagram of automatic voltage regulator including sensor dynamics.

(a) Amplifier model

$$G_A(s) = \frac{\Delta V_R(s)}{\Delta e(s)} = \frac{K_A}{1+sT_A} \tag{1}$$

where nominally $10 < K_A < 400$ and a small time constant is in the range $0.02 < T_A < 0.1$.

(b) Exciter model

$$G_E(s) = \frac{\Delta V_f(s)}{\Delta V_R(s)} = \frac{K_E}{1+sT_E} \tag{2}$$

where nominally $10 < K_E < 400$ and a time constant is in the range $0.5 < T_E < 1$.

(c) Generator model

$$G_G(s) = \frac{\Delta E(s)}{\Delta V_f(s)} \cong \frac{\Delta V_t(s)}{\Delta V_f(s)} = \frac{K_G}{1+sT_G} \tag{3}$$

where nominally $0.7 < K_G < 1$ and a time constant is in the range $1 < T_G < 2$. These constants vary depending on the load.

(d) Sensor model

$$H_s(s) = \frac{\Delta V_s(s)}{\Delta V_t(s)} = \frac{K_s}{1+sT_s} \tag{4}$$

where nominally $K_s = 1$, and the time constant is in the range $0.001 < T_s < 0.06$.

Case 1: Mathematical modeling of AVR system neglecting sensor dynamics

The open loop transfer function is given by

$$G(s) = \frac{\Delta V_t(s)}{\Delta e(s)} = \frac{K_A K_E K_G}{(1+sT_A)(1+sT_E)(1+sT_G)} \tag{5}$$

The closed loop transfer function of AVR system neglecting the sensor dynamics is written as

$$G_{CL}(s) = \frac{K_A K_E K_G}{(1+sT_A)(1+sT_E)(1+sT_G)+K_A K_E K_G} \tag{6}$$

Analysis of (6) gives that the static error decreases with increased loop gain as for p% static error the loop gain $K_A K_E K_G > \frac{100}{p} - 1$. For stability compensation a series phase lead compensator $G_s(s) = 1 + sT_c$ can be included in the system.

The state space model of AVR system neglecting sensor dynamics in phase variable companion form may be written from transfer function model (6) as following:

$$\dot{X} = AX + Bu \tag{7}$$

$$y = CX + Du \tag{8}$$

where, $X = [x_1 \quad x_2 \quad x_3]^T$, $y = x_1 = \Delta V_t$, $u = \Delta V_{ref}$, and

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{1+K_A K_E K_G}{T_A T_E T_G} & -\left(\frac{1}{T_A T_E}+\frac{1}{T_E T_G}+\frac{1}{T_G T_A}\right) & -\left(\frac{1}{T_A}+\frac{1}{T_E}+\frac{1}{T_G}\right) \end{bmatrix},$$

$$B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} \frac{K_A K_E K_G}{T_A T_E T_G} & 0 & 0 \end{bmatrix}, \quad D = 0.$$

Case 2: Mathematical modeling of AVR system including sensor dynamics

The closed loop transfer function of AVR system including sensor dynamics is written as

$$G_{CL}(s) = \frac{K_A K_E K_G (1+sT_s)}{(1+sT_A)(1+sT_E)(1+sT_G)(1+sT_s)+K_A K_E K_G K_s} \tag{9}$$

Including the sensor dynamics in the AVR system modeling, a pole and a zero are added.

The state space model of AVR system including sensor dynamics in phase variable companion form may be written from transfer function model (9) similar as (7) and (8) with following modification:

$X = [x_1 \quad x_2 \quad x_3 \quad x_4]^T$, and

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -a_4 & -a_3 & -a_2 & -a_1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

$$C = \begin{bmatrix} \frac{K_A K_E K_G}{T_A T_E T_G T_s} & \frac{K_A K_E K_G}{T_A T_E T_G} & 0 & 0 \end{bmatrix}, \quad D = 0$$

where, $a_1 = \left(\frac{1}{T_A}+\frac{1}{T_E}+\frac{1}{T_G}+\frac{1}{T_s}\right)$, $a_2 = \left(\frac{1}{T_A T_E}+\frac{1}{T_A T_G}+\frac{1}{T_A T_s}+\frac{1}{T_E T_G}+\frac{1}{T_E T_s}+\frac{1}{T_G T_s}\right)$

$$a_3 = \left(\frac{1}{T_A T_E T_G}+\frac{1}{T_A T_E T_s}+\frac{1}{T_A T_G T_s}+\frac{1}{T_E T_G T_s}\right), \text{ and } a_4 = \frac{1+K_A K_E K_G K_s}{T_A T_E T_G T_s}$$

In the both cases of neglecting and including sensor dynamics AVR system is a continuous-time single-input single-output (SISO) linear time-invariant (LTI) system. The adaptive optimal control using policy iteration technique based adaptive critic scheme for continuous-time LTI systems is discussed in the following 'Implementation of online policy iteration technique for adaptive optimal control design'.

Adaptive optimal control using policy iteration technique

There are basically two ways of solving the associated optimal control problem; one is Pontryagin's minimum principle and the other is Bellman's dynamic programming (DP) [15,17,18]. However, the computational complexity of Bellman's dynamic programming and associated Hamilton–Jacobi–Bellman (HJB) equation lead to formulation of adaptive/approximate dynamic programming (ADP) [20–22]. Combining the concepts of ADP, RL, and backprogation [23], Werbos introduced an approach for ADP called adaptive critic designs (ACDs) as a way for solving dynamic programming problems forward-in-time. Adaptive critic design (ACD) utilizes two parametric structures known as the actor and the critic. The actor parameterizes the control policy. The critic approximates a value-related cost function and captures the effect that the control law will have on the future cost which describes the performance of control system. At any given time, the critic provides guidance to improve the control policy, and the actor to update the critic. Several versions of adaptive critic designs are present in the literature [23,24]. Werbos defined actor–critic online learning algorithms to solve the optimal control problem based on Value Iteration (VI), and defined a family of VI algorithms as Adaptive Dynamic Programming (ADP) algorithms. He used a critic neural network for value function approximation and an actor neural network for approximation of the control policy. Generalized Policy Iteration is a family of optimal learning techniques which has policy iteration (PI) at one extreme. Policy Iteration (PI) algorithms consist two-step iteration: policy evaluation and policy

improvement. Instead of solving HJB equation by direct approach, the PI algorithm starts by evaluating the cost of a given initial admissible control policy, which is often accomplished by solving a nonlinear Lyapunov equation. This updated cost is then used to obtain an updated improved control policy which will have a lower associated cost [25–30]. This is often accomplished by minimizing a Hamiltonian function with respect to the updated cost. This is the so-called 'greedy policy' with respect to the updated cost [28]. These two steps of policy evaluation and policy improvement are repeated until the policy improvement step no longer changes the actual policy and thus converging to the optimal control. It is noted that the infinite horizon cost can be evaluated only in the case of admissible and stabilizing control policies. Admissibility is in fact a condition for the control policy which is used to initialize the algorithm [28]. PI algorithm requires an initial stabilizing control policy, but VI does not require an initial stabilizing control policy. The policy iteration technique based adaptive critic scheme performs adaptive optimal control without using complete knowledge of the system dynamics. The online policy iteration algorithm solves the optimal control problem, along a single state trajectory, does not require knowledge of the system internal dynamics, and thus can be viewed as a direct adaptive optimal control technique. Unlike the regular adaptive controllers which rely on online identification of the system dynamics followed by model based controller design, the policy iteration method relies on identification of the cost function associated with a given control policy followed by policy improvement in the sense of minimizing the identified cost. The infinite horizon optimal control (i.e. linear quadratic regulator (LQR)) problem which is the basis of adaptive critic scheme and policy iteration technique for continuous-time LTI systems is presented in this section [18,24,29,30].

Infinite horizon optimal control and continuous-time adaptive critic scheme

Consider the continuous-time linear time-invariant dynamical system described by

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{10}$$

where $x(t) \in R^n$, $u(t) \in R^m$ and (A, B) is stabilizable, subject to the optimal control problem

$$u^*(t) = \arg \min_{\substack{u(t) \\ to \leqslant t \leqslant \infty}} V(t_0, x(t_0), u(t)) \tag{11}$$

where the infinite horizon quadratic cost function to be minimized is expressed as

$$V(x(t_0), t_0) = \int_{t_0}^{\infty} (x^T(\tau)Qx(\tau) + u^T(\tau)Ru(\tau))d\tau \tag{12}$$

with $Q \geqslant 0$, $R > 0$ and $(Q^{1/2}, A)$ detectable.

The solution of this optimal control problem, determined by Bellman's optimality principle [15], is given by

$$u(t) = -Kx(t) \text{ with } K = R^{-1}B^TP \tag{13}$$

where the matrix P is the unique positive definite solution of the algebraic riccati equation (ARE)

$$A^TP + PA - PBR^{-1}B^TP + Q = 0 \tag{14}$$

Eq. (13) gives a stabilizing closed loop controller determined from the unique positive semi-definite solution of ARE under the detectability condition. Here to solve (14), both system matrix A and control input matrix B must be known i.e. complete knowledge of the system dynamics is required. Due to this reason, developing algorithms that will converge to the solution of the optimization problem without performing prior system identification and using explicit models of the system dynamics is of particular interest from the control systems point of view. In adaptive critic scheme (13) and (14) are represented by two parametric function approximation networks namely action network and critic network respectively. The online policy iteration technique which does not require the knowledge of the system internal dynamics gives optimal control solution of the LQR problem, and which gives an adaptive controller that converges to the state feedback optimal controller is presented in the following 'Policy iteration technique'.

Policy iteration technique

The policy iteration technique [22,25–30] is based on an actor–critic structure, consists of two-step iteration- critic update and actor update. For a given stabilizing controller critic computes the associated infinite horizon cost. The actor computes the control policy and is represented by its parameters (i.e. feedback controller gain) [29,30].

Let a stabilizing gain K for (10), under the assumption that (A, B) is stabilizable, such that $\dot{x} = (A - BK)x$ is a stable closed loop system. Then the corresponding infinite horizon quadratic cost is given by

$$V(x(t)) = \int_t^{\infty} x^T(\tau)(Q + K^TRK)x(\tau)d\tau = x^T(t)Px^T(t) \tag{15}$$

where P is the real symmetric positive definite solution of the Lyapunov matrix equation

$$(A - BK)^TP + P(A - BK) = -(K^TRK + Q) \tag{16}$$

and $V(x(t))$ serves as a Lyapunov function for (10) with controller gain K. The cost function (15) can be written as

$$V(x(t)) = \int_t^{t+T} x^T(\tau)(Q + K^TRK)x(\tau)d\tau + V(x(t+T)) \tag{17}$$

Based on (17), denoting $x(t)$ with x_t, with the parameterization $V(x_t) = x_t^TPx_t$ and considering an initial stabilizing control gain K_1, the following two-step online policy iteration algorithm can be implemented:

1. Policy evaluation

$$x_t^TP_ix_t = \int_t^{t+T} x_\tau^T(Q + K_i^TRK_i)x_\tau d\tau + x_{t+T}^TP_ix_{t+T} \tag{18}$$

2. Policy improvement

$$K_{i+1} = R^{-1}B^TP_i \tag{19}$$

Eqs. (18) and (19) formulate a new policy iteration algorithm. It is important to note that this algorithm does not require system matrix A for its solution, only control input matrix B must be known for updating K. The convergence of the PI algorithm is established by lemmas, remarks and theorems referred in [29,30].

Implementation of online policy iteration technique for adaptive optimal control design

In this section the implementation issues of online policy iteration algorithm for adaptive optimal control design is discussed. The complete knowledge of the system internal dynamics is not required to implement this control scheme. Only the knowledge of B matrix is required which explicitly appears in (19). Since the system matrix A does not appear explicitly in either (18) or (19) thus it is not required for the computation of either of two steps

of the PI algorithm. The information about the system internal dynamics which is represented by matrix A is embedded in the states $x(t)$ and $x(t+T)$ which are observed online.

To find the critic parameters (matrix P_i) of the cost function associated with the policy K_i, the left hand side term $x^T(t)P_ix(t)$ in (18) is written as

$$x^T(t)P_ix(t) = \bar{p}_i^T\bar{x}(t) \tag{20}$$

where $\bar{x}(t) = \{x_i(t)x_j(t)\}_{i=1,n;j=i,n}$ which denotes the Kronecker product quadratic polynomial basis vector and $p = \upsilon(P)$ with $\upsilon(\cdot)$ a vector valued matrix function that acts on symmetric matrices and returns a column vector by stacking the elements of the diagonal and upper triangular part of the symmetric matrix into a vector where the off-diagonal elements are taken as $2P_{ij}$ [29,30]. Using (20), (18) is rewritten as

$$\begin{aligned}\bar{p}_i^T(\bar{x}(t) - \bar{x}(t+T)) &= \int_t^{t+T} x^T(\tau)(Q + K_i^T R K_i)x(\tau)d\tau \\ &\equiv d(\bar{x}(t), K_i)\end{aligned} \tag{21}$$

where $\bar{p}_i$ is the vector of unknown parameters and $\bar{x}(t) - \bar{x}(t+T)$ acts as a regression vector. The right hand side target function is denoted by $d(\bar{x}(t), K_i)$ which is also known as the reinforcement on the time interval $[t, t+T]$, is measured based on the system states over the time interval $[t, t+T]$. Thus, $d(\bar{x}(t), K_i) = V(t+T) - V(t)$, where $V(t)$ is a new state introduced augmenting the system (10), is defined as $\dot{V}(t) = x^T(t)Qx(t) + u^T(t)Ru(t)$.

The parameter vector $\bar{p}_i$ of the function $V_i(x_t)$ (i.e. the Critic), which will then yield the matrix P_i, is found by minimizing, in the least-squares sense, the error between the target function, $d(\bar{x}(t), K_i)$, and the parameterized left hand side of (21). Evaluating the right hand side of (21) at $N \geqslant n(n+1)/2$ (the number of independent elements in the matrix P_i) points $\bar{x}^i$ in the state space, over the same time interval T, the least-squares solution is obtained as

$$\bar{p}_i = (XX^T)^{-1}XY \tag{22}$$

where

$$X = [\bar{x}_\Delta^1 \bar{x}_\Delta^2 \dots \bar{x}_\Delta^N]$$
$$\bar{x}_\Delta^i = \bar{x}^i(t) - \bar{x}^i(t+T)$$
$$Y = [d(\bar{x}^1, K_i)d(\bar{x}^2, K_i)\dots d(\bar{x}^N, K_i)]^T$$

The least-squares problem can be solved in real-time after a sufficient number of data points are collected along a single state trajectory, under the regular presence of an excitation requirement. Alternatively, (22) can be solved also using recursive estimation algorithms (e.g. gradient descent algorithms or the Recursive Least Squares (RLS) algorithm) in which case a persistence of excitation condition is required. Due to this reason there are no real issues related to the algorithm becoming computationally expensive with the increase of the state space dimension [29,30]. Even though the convergence property of the online algorithm is not affected by the value of sample time T; it affects the excitation condition necessary in the setup of a numerically well posed least squares problem and obtaining the least squares solution (22). From this point of view a minimal insight relative to the system dynamics would be required for choosing the sampling time T [29].

The online PI algorithm requires only measurements of the states at discrete moments in time, t and $t+T$, as well as knowledge of the observed cost over the time interval $[t, t+T]$, which is $d(\bar{x}(t), K_i)$. Therefore, knowledge of system matrix A is not required for the cost evaluation or the control policy update, and only the matrix B is required for the control policy update, using (19), which makes the tuning algorithm only partially model-free. The PI algorithm converges to the optimal control solution measuring cost along a single state trajectory, provided that there is enough initial excitation in the system. Since the algorithm iterates only on stabilizing policies which will make the system states go to zero, sufficient excitation in the initial state of the system is necessary. In the case that excitation is lost prior to obtaining the convergence (system reaches the equilibrium point) a new experiment needs to be conducted having as a starting point the last policy from the previous experiment. In this case, the control policy is updated at time $t+T$, after observing the state $x(t+T)$ and it is used for controlling the system during the time interval $[t+T, t+2T]$. The critic stops updating the control policy when the difference between the system performances evaluated at two consecutive steps crosses below a designer specified limit, i.e. the algorithm has converged to the optimal controller. Also in the case that this error is bigger than this specified limit the critic again starts tuning the actor parameters to obtain an optimal control policy. If there is a sudden change in system dynamics described by the matrix A as long as the present controller is stabilizing for the new matrix A, the algorithm will converge to the solution to the corresponding new ARE. Thus the algorithm is suitable for online implementation from the control theory point of view.

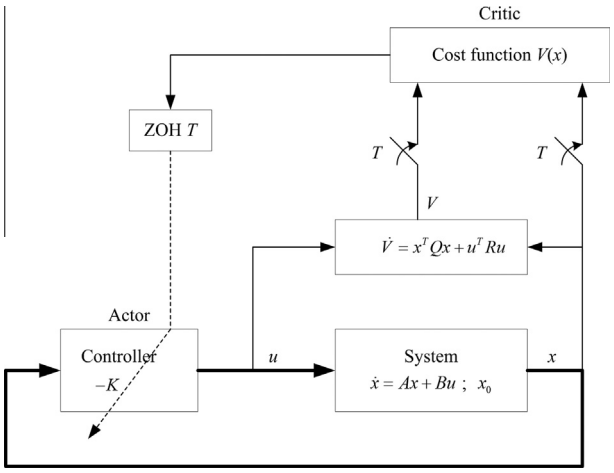

Fig. 3. Adaptive optimal controller with actor–critic structure.

Fig. 3 [29,30] shows the schematic block diagram of adaptive optimal control with actor–critic structure for LTI system. Since the system is augmented with an extra state $V(t)$ that is part of the adaptive critic control scheme thus this controller is actually a dynamic controller with the cost state. This adaptive optimal controller has a hybrid structure with a continuous-time internal state followed by a sampler and discrete-time update rule. The application of the proposed control scheme is implemented for automatic voltage regulator system of a single-area power system which is presented in the following 'Simulation results' of this paper.

Simulation results

Consider the AVR system parameters as [1–4,6,10]: $T_A = 0.1$ s, $T_E = 0.4$ s, $T_G = 1$ s, $T_s = 0.01$ s, $K_A = 10$, $K_E = 1$, $K_G = 1$, and $K_s = 1$.

Case 1: AVR system neglecting sensor model

For the given system parameters the closed loop transfer function of AVR system (6) is obtained as

$$G_{CL}(s) = \frac{250}{s^3 + 13.5s^2 + 37.5s + 275} \tag{23}$$

and the state space model of AVR system is obtained as

$$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -275 & -37.5 & -13.5\end{bmatrix},\quad B=\begin{bmatrix}0\\0\\1\end{bmatrix},\quad C=[250\ \ 0\ \ 0],\quad D=[0]$$

In implementation of PI algorithm for system (23), the initial conditions for states and cost function, and critic parameters are taken as $x_0=[0.1\quad 0.05\quad 0.04\quad 0]; P=[000;000;000]$. The length of the simulation in samples is taken 120, and sample time $T=0.05$ s. The cost function parameters Q and R are taken as identity matrices of appropriate dimensions. The unique positive definite solution of ARE (14), denoted here by matrix *RicP*, and adaptive optimal critic matrix P of adaptive critic scheme using PI in (18) & (19) with (22) respectively are obtained as

$$RicP=\begin{bmatrix}167.9165 & 22.5745 & 0.0018\\ 22.5745 & 11.3630 & 0.6104\\ 0.0018 & 0.6104 & 0.0820\end{bmatrix},\quad P=\begin{bmatrix}167.9185 & 22.5747 & 0.0018\\ 22.5747 & 11.3631 & 0.6104\\ 0.0018 & 0.6104 & 0.0820\end{bmatrix},$$

and the actor gains of LQR design by (13) & (14) denoted here by *RicK*, and actor K by adaptive critic scheme using PI in (19) respectively are obtained as

$$RicK=[0.0018\ \ 0.6104\ \ 0.0820],\quad K=[0.0018\ \ 0.6104\ \ 0.0820]$$

The eigenvalues of closed loop system are obtained as -12.3013, $-0.6404+4.6846$i, $-0.6404-4.6846$i.

The simulation responses using PI technique for LTI system (23) are shown in Figs. 4–8. Fig. 4 shows the system state trajectories which converge towards the equilibrium point. Each circle "o" on state trajectories shows the modification of initial conditions to new states values during simulation in each sample. Fig. 5 shows the control signal trajectory which also converges towards zero. Fig. 6 shows the evolution of closed loop poles of the system during simulation. Fig. 7 shows the convergence of critic parameters of matrix P towards optimal values. Fig. 8 shows P parameters updating with iteration, here * at one indicate update, and * at zero indicate no update.

Simulation is also done for the case of change in system parameter which changes the elements of system matrix A during simulation. Consider the change in system parameter as $T_G=1.25$ s at sample $k=41$; (i.e. $t=2.05$ s) then for this case system transfer function is obtained as

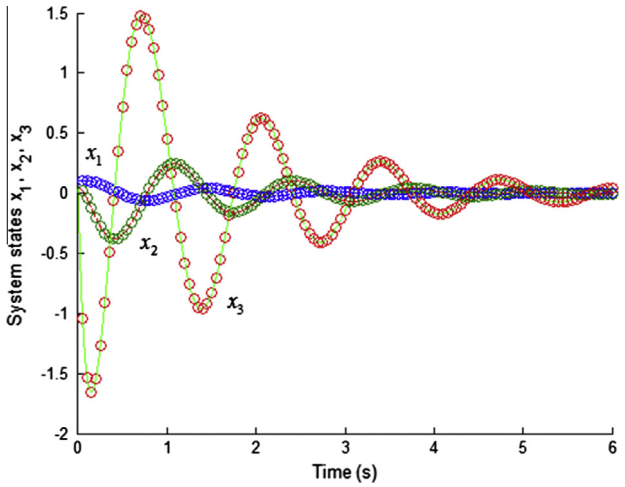

Fig. 4. System states.

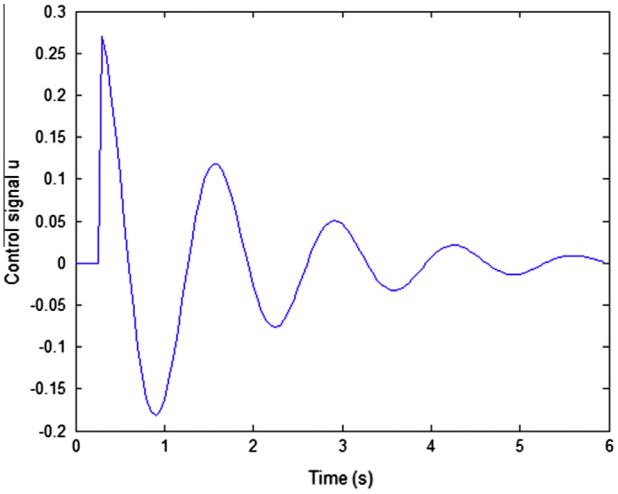

Fig. 5. Control signal.

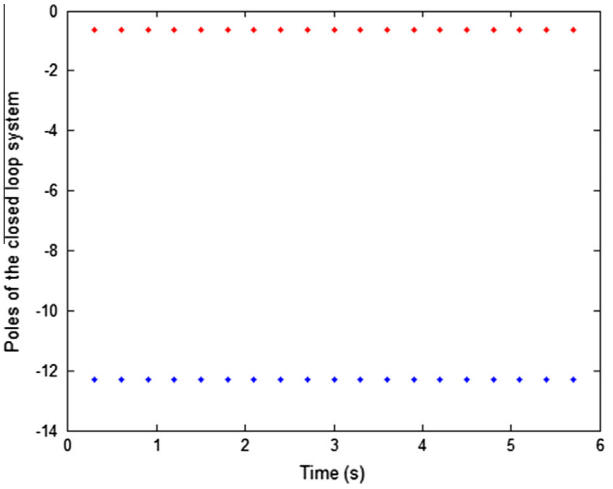

Fig. 6. Evolution of poles of closed loop system.

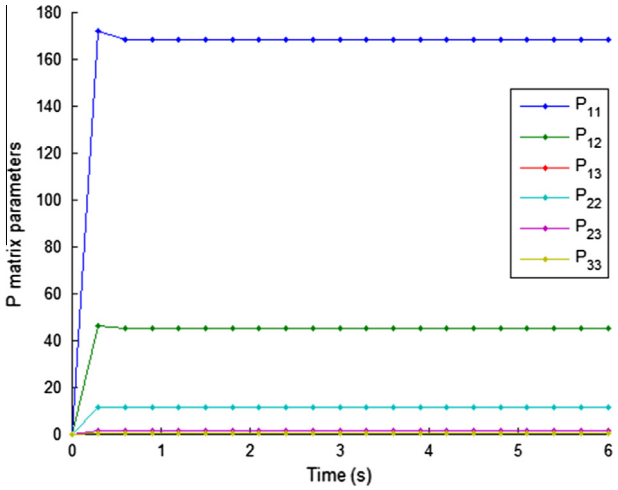

Fig. 7. Critic parameters.

$$G_{CL}(s)=\frac{200}{s^3+13.3s^2+35s+220}, \tag{24}$$

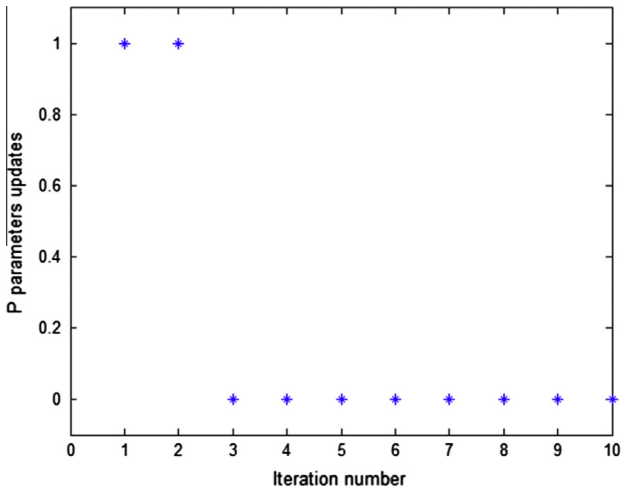

Fig. 8. Updating of critic parameters.

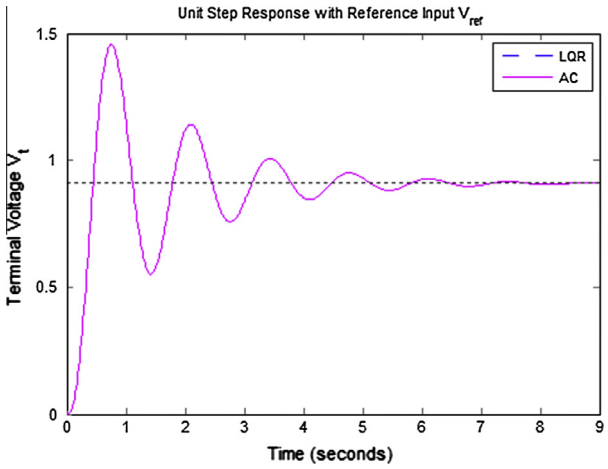

Fig. 10. Unit step response of closed loop system.

and thus the system state model is changed such that $A(3,1) = -220$; $A(3,2) = -35$; $A(3,3) = -13.3$; and $C = [200 \quad 0 \quad 0]$. In this case the unique positive definite solution of *RicP* and *P* using LQR and PI technique respectively are obtained as

$$RicP = \begin{bmatrix} 103.1486 & 16.0069 & 0.0023 \\ 16.0069 & 8.8044 & 0.4685 \\ 0.0023 & 0.4685 & 0.0726 \end{bmatrix}, \quad P = \begin{bmatrix} 167.9185 & 22.5747 & 0.0018 \\ 22.5747 & 11.3631 & 0.6104 \\ 0.0018 & 0.6104 & 0.0820 \end{bmatrix},$$

and the actor gains of LQR design *RicK*, and actor *K* by adaptive critic scheme using PI respectively are obtained as

$$RicK = [0.0023 \quad 0.4685 \quad 0.0726], \quad K = [0.0018 \quad 0.6104 \quad 0.0820]$$

The eigenvalues of closed loop system are obtained as -11.9427, $-0.7196 + 4.2313i$, $-0.7196 - 4.2313i$.

Fig. 9 shows the evolution of closed loop poles of the system during simulation with change in system parameter at sample $k = 41$. The remaining responses obtained are same as above in this case. Fig. 10 presents the closed loop response of LTI system (23) using both approaches of LQR and adaptive critic (AC) using PI technique by replacing $u = -Kx + r$ in state equations where $r = V_{ref}$ is a unit step input and *K* is actor gains *RicK* and *K* respectively. It remains exactly the same also for case with change in system parameters. It is observed here that the adaptive optimal controller using PI technique gives the similar response as of standard LQR.

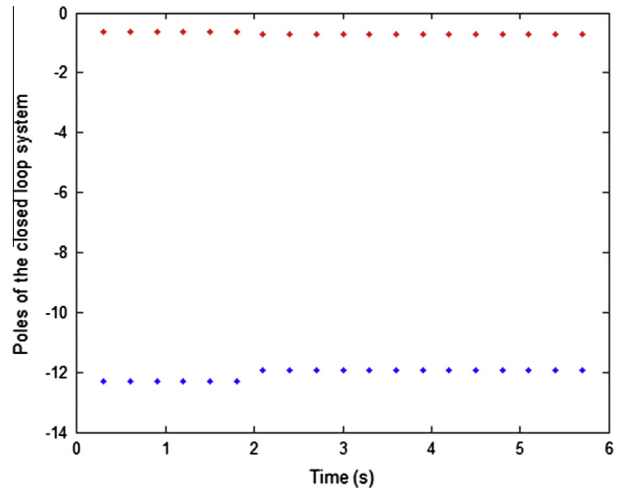

Fig. 9. Evolution of poles of closed loop system with change in system parameters at $k = 41$.

Case 2: AVR system including sensor model

For the given system parameters the transfer function of AVR system (9) is obtained as

$$G_{CL}(s) = \frac{250s + 25000}{s^4 + 113.5s^3 + 1387.5s^2 + 3775s + 27500}, \tag{25}$$

and the state space model of AVR system is obtained as

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -27500 & -3775 & -1387.5 & -113.5 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

$$C = [25000 \quad 250 \quad 0 \quad 0], \quad D = [0]$$

In implementation of PI algorithm for system (25), the initial conditions for states and cost function, and critic parameters are taken as $x_0 = [0.1 \quad 0.05 \quad 0.04 \quad 0.02 \quad 0]; P = [0000; 0000; 0000]$. The length of the simulation in samples is taken 120, and sample time $T = 0.05$ s. The cost function parameters *Q* and *R* are taken as identity matrices of appropriate dimensions. The unique positive definite solution of ARE (14), denoted here by matrix *RicP*, and adaptive optimal critic matrix *P* of adaptive critic scheme using PI in (18) & (19) with (22) respectively are obtained as

$$RicP = 1.0 \times 10^3 \begin{bmatrix} 6.5160 & 0.8940 & 0.1290 & 0.0000 \\ 0.8940 & 0.3225 & 0.0446 & 0.0002 \\ 0.1290 & 0.0446 & 0.0100 & 0.0000 \\ 0.0000 & 0.0002 & 0.0000 & 0.0000 \end{bmatrix},$$

$$P = 1.0 \times 10^3 \begin{bmatrix} 6.5164 & 0.8940 & 0.1290 & -0.0000 \\ 0.8940 & 0.3225 & 0.0446 & 0.0002 \\ 0.1290 & 0.0446 & 0.0100 & 0.0000 \\ -0.0000 & 0.0002 & 0.0000 & 0.0000 \end{bmatrix},$$

and the actor gains of LQR design by (13) & (14) denoted here by *RicK*, and actor *K* by adaptive critic scheme using PI in (19) respectively are obtained as

$$RicK = [0.0000 \quad 0.2369 \quad 0.0325 \quad 0.0047], \quad K = [-0.0001 \quad 0.2370 \quad 0.0325 \quad 0.0047]$$

The eigenvalues of closed loop system in this case are obtained as

−99.9763, −12.4887, −0.5199 + 4.6642i, and −0.5199 − 4.6642i.

The simulation responses using PI technique for LTI system (25) are shown in Figs. 11–15. Fig. 11 shows the system state trajectories which converge towards the equilibrium point. Each circle "o" on state trajectories shows the modification of initial conditions to new states values during simulation in each sample. Fig. 12 shows the control signal trajectory which also converges towards zero. Fig. 13 shows the evolution of closed loop poles of the system during simulation. Fig. 14 shows the convergence of critic parameters of matrix P towards optimal values. Fig. 15 shows P parameters updating with iteration, here * at one indicate update, and * at zero indicate no update.

Simulation is also done for the case of change in system parameter which changes the elements of system matrix A during simulation. Consider the change in system parameter as $T_G = 1.25$ s at sample $k = 41$; (i.e. $t = 2.05$ s) then for this case system transfer function is obtained as

$$G_{CL}(s) = \frac{200s + 20000}{s^4 + 113.3s^3 + 1365s^2 + 3520s + 22000} \tag{26}$$

the system state model is changed such that $A(4,1) = -22{,}000$; $A(4,2) = -3520$; $A(4,3) = -1365$; $A(4,4) = -113.3$; and $C = [20000\ 200\ 0\ 0]$. In this case also the unique positive definite solution of $RicP$ and P using LQR and PI technique respectively are obtained as

$$RicP = 1.0 \times 10^3 \begin{bmatrix} 3.5299 & 0.5643 & 0.1021 & 0.0000 \\ 0.5643 & 0.2072 & 0.0345 & 0.0002 \\ 0.1021 & 0.0345 & 0.0091 & 0.0000 \\ 0.0000 & 0.0002 & 0.0000 & 0.0000 \end{bmatrix},$$

$$P = 1.0 \times 10^3 \begin{bmatrix} 6.5164 & 0.8940 & 0.1290 & -0.0000 \\ 0.8940 & 0.3225 & 0.0446 & 0.0002 \\ 0.1290 & 0.0446 & 0.0100 & 0.0000 \\ -0.0000 & 0.0002 & 0.0000 & 0.0000 \end{bmatrix},$$

and the actor gains of LQR design $RicK$, and actor K by adaptive critic scheme using PI in respectively are obtained as

$$RicK = [0.0000\ \ 0.1604\ \ 0.0256\ \ 0.0046],\quad K = [-0.0001\ \ 0.2370\ \ 0.0325\ \ 0.0047]$$

The eigenvalues of closed loop system in this case are obtained as

−99.9821, −12.0977, −0.6125 + 4.2206i, and −0.6125 − 4.2206i.

Fig. 16 shows the evolution of closed loop poles of the system during simulation with change in system parameter at sample

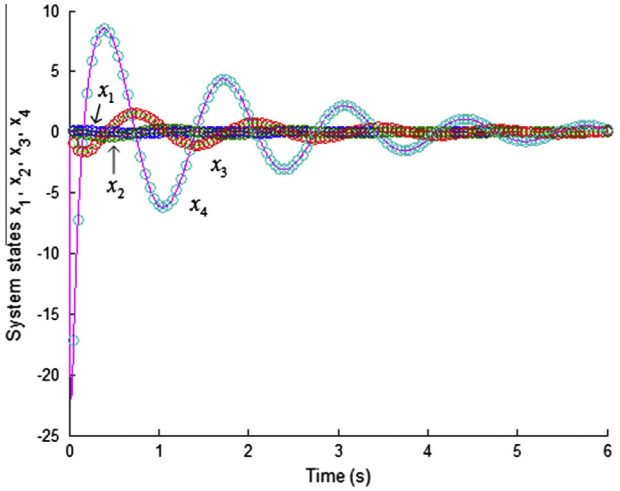

Fig. 11. System states.

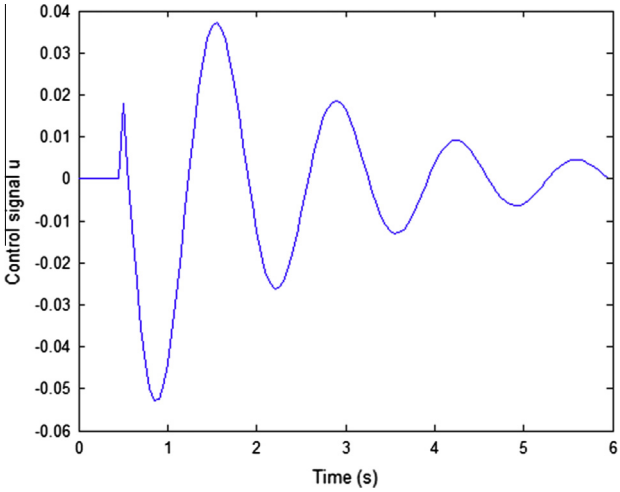

Fig. 12. Control signal.

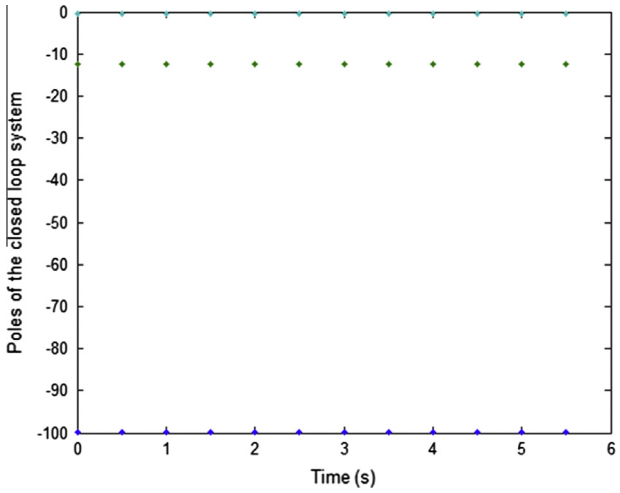

Fig. 13. Evolution of poles of closed loop system.

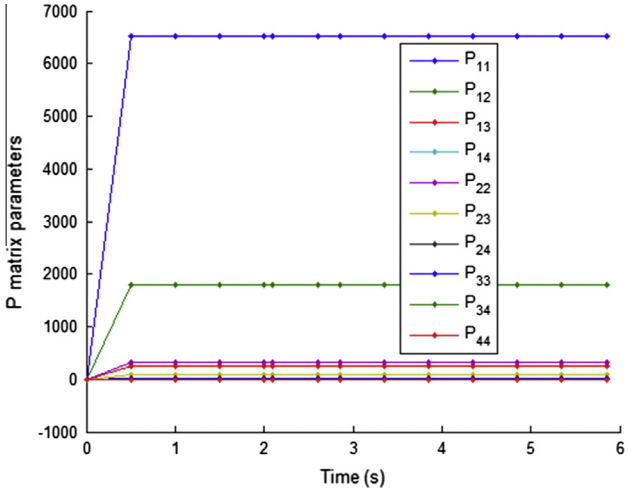

Fig. 14. Critic parameters.

$k = 41$. The remaining responses obtained are same as above in this case. Fig. 17 presents the closed loop response of LTI system (25) using both approaches of LQR and adaptive critic (AC) using PI

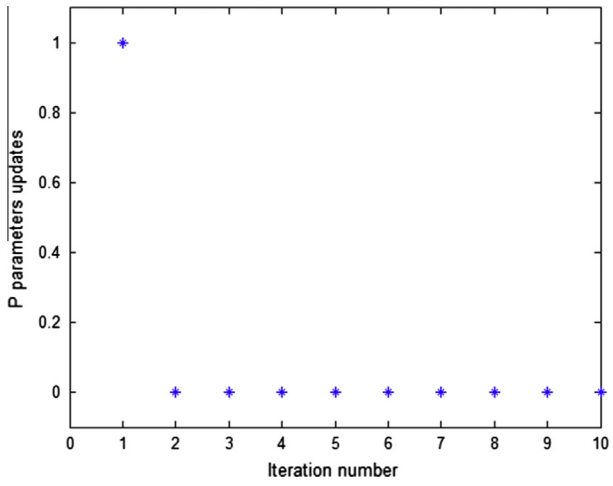

Fig. 15. Updating of critic parameters.

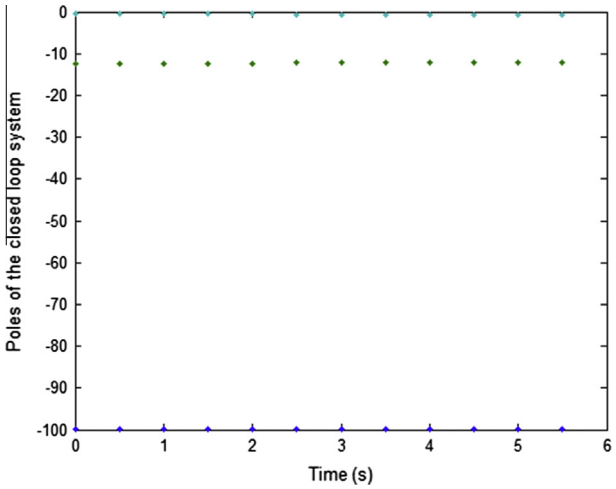

Fig. 16. Evolution of poles of closed loop system with change in system parameters at $k = 41$.

technique by replacing $u = -Kx + r$ in state equations where $r = V_{ref}$ is a unit step input and K is actor gains $RicK$ and K respectively. It remains exactly the same also for case with change in system parameters. It is observed here that the adaptive optimal controller using PI technique gives the similar response as of standard LQR.

It is observed from the above simulation results for AVR system for both cases that critic parameter matrix P and actor parameter K obtained from PI based adaptive critic scheme are converging adaptively to optimal values and are mostly of same values of $RicP$ and $RicK$ respectively that obtained from LQR approach. Also in case of change in the system parameter in real situation the controller adapts it and converges to same optimal values. Thus the actor K and critic P parameters remain unchanged.

Analyzing the simulation results obtained for AVR system in both the cases of system models neglecting and including sensor dynamics, and changes in system parameters applying adaptive critic design using online policy iteration technique, it is established that this proposed control scheme provide a promising adaptive optimal control solution for dynamical systems without complete knowledge of the system dynamics. The effect of system modeling uncertainties has been demonstrated by analyzing the simulation results for AVR system for both cases of neglecting and including sensor dynamics. It is observed that the simulation

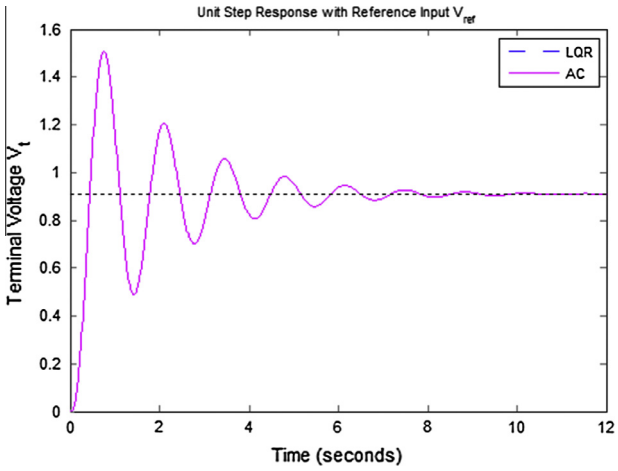

Fig. 17. Unit step response of closed loop system.

results obtained in both cases are similar. Thus it inference that the system model may be simplified by neglecting subsystems whose time constant is very small without considerably affecting the system characteristics and thus simplifying control design and its performance. The structural change introduced in system dynamics by including sensor dynamics is augmenting the system behaviour such as of its credit in closed loop response. The structural change in system will not be adapted by the proposed controller and in that case control algorithm has to be modified according to order of system dynamics. The proposed controller adapts the change in system parameters in real situation at any moment of time. Thus this technique is partially model-free, effective & robust.

Conclusion

In this paper adaptive optimal control design using online policy iteration technique for automatic voltage regulator system for single-area power system has been presented. Policy iteration technique which is based on actor–critic structure consists of two-step iteration: policy evaluation and policy improvement. The PI algorithm converges to the optimal control solution under the condition of initial stabilizing controller. The infinite horizon optimal design using LQR by solution of ARE requires the complete knowledge of the system dynamics. The online policy iteration technique based control scheme gives online the continuous-time adaptive optimal control solution without using the complete knowledge of the system's internal dynamics. The knowledge of systems internal dynamics (i.e. matrix A) is not needed for evaluation of cost or the update of control policy; only the knowledge of input matrix B is required for updating the control policy. Thus, adaptive optimal control scheme using online policy iteration technique is partially model-free. In this paper the application of control scheme is implemented for AVR system for both of system models neglecting and including sensor dynamics. Also a change in system parameter introduced during simulation to demonstrate the parametric changes in system in real situation. The simulation results and performance analysis justify the effectiveness & robustness of the proposed control scheme. The application of proposed control scheme for AVR system which represent a continuous-time LTI system demonstrate that online policy iteration technique based adaptive critic scheme gives an infinite horizon adaptive optimal control solution to the real-time dynamics of a continuous-time LTI system.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit organizations. First author is thankful to Madan Mohan Malaviya Engineering College Gorakhpur, QIP Centre, I.I.T. Roorkee, and AICTE, India for sponsoring & financing him for Ph.D work under QIP scheme.

References

[1] Elgerd OI. Electric energy systems theory: an introduction. 2nd ed. New Delhi: Tata McGraw-Hill Publishing Company Ltd.; 2004. ch. 9.
[2] Prabha Kundur. Power system stability and control. New Delhi: Tata McGraw-Hill Education Pvt. Ltd.; 2011. ch. 11.
[3] Kothari DP, Nagrath IJ. Modern power system analysis. New Delhi: Tata McGraw-Hill Publishing Company Ltd.; 2003. ch. 8.
[4] Zwe-Lee Gaing. A particle swarm optimization approach for optimum design of PID controller in AVR system. IEEE Trans Energy Convers 2004;19(2):84–391.
[5] Mukherjee V, Ghoshal SP. Comparison of intelligent fuzzy based AGC coordinated PID controlled and PSS controlled AVR system. Int J Electr Power Energy Syst 2007;29(9):679–89.
[6] Chatterjee A, Mukherjee V, Ghoshal SP. Velocity relaxed and craziness-based swarm optimized intelligent PID and PSS controlled AVR system. Int J Electr Power Energy Syst 2009;31(7–8):323–33.
[7] Pan I, Das S. Chaotic multi-objective optimization based design of fractional order $PI^{\lambda}D^{\mu}$ controller in AVR system. Int J Electr Power Energy Syst 2012;43(1):393–407.
[8] Pan I, Das S. Frequency domain design of fractional order PID controller for AVR system using chaotic multi-objective optimization. Int J Electr Power Energy Syst 2013;51:106–18.
[9] Razmi H, Shayanfar HA, Teshnehlab M. Steady state voltage stability with AVR voltage constraints. Int J Electr Power Energy Syst 2012;43(1):650–9.
[10] Saffet Ayasun, Ayetül Gelen. Stability analysis of a generator excitation control system with time delays. Electr Eng 2010;91(6):347–55.
[11] Joong-Moon Kim, Seung-Il Moon. A study on a new AVR parameter tuning concept using on-line measured data with the real-time simulator. Eur Trans Electr Power 2006;16(3):235–46.
[12] Lee SS, Park JK. Design of power system stabilizer using observer/sliding mode, observer/sliding mode model following and H-infinity sliding mode controllers for small signal stability study. Int J Electr Power Energy Syst 1998;20(8):543–53.
[13] Gurrala G, Sen I, Padhi R. Single network adaptive critic design for power system stabilizers. IET Gener Transm Distrib 2009;3(9):850–8.
[14] Jalili M, Yazdanpanah MJ. Transient stability enhancement of power systems via optimal nonlinear state feedback control. Electr Eng 2006;89(2):149–56.
[15] Lewis FL. Optimal control. New York: John Wiley & Sons, Inc.; 1986.
[16] Astrom KJ, Wittenmark B. Adaptive control. 2nd ed. New Delhi: Pearson Education, Inc.; 2006.
[17] Gopal M. Digital control and state variable methods: conventional and intelligent control systems. 4th ed. New Delhi: Tata McGraw Hill Education Pvt. Ltd.; 2012.
[18] Laxmidhar Behera, Indrani Kar. Intelligent systems and control: principles and applications. 1st ed. New Delhi: Oxford University Press; 2009.
[19] Bhuvaneswari NS, Uma G, Rangaswamy TR. Adaptive and optimal control of a non-linear process using intelligent controllers. Appl Soft Comput 2009;9(1):182–90.
[20] Murray JJ, Cox CJ, Lendaris GG, Saeks R. Adaptive dynamic programming. IEEE Trans Syst Man Cybernetics Part C Appl Rev 2002;32(2):140–53.
[21] Fei-Yue Wang, Huaguang Zhang, Derong Liu. Adaptive dynamic programming: an introduction. IEEE Comput Intell Magazine 2009;4(2):39–47.
[22] Lewis FL, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control. IEEE Circ Syst Magazine 2009;9(3):32–50.
[23] Prokhorov DV, Wunsch II DC. Adaptive critic designs. IEEE Trans Neural Nets 1997;8(5):997–1007.
[24] Hanselmann T, Noakes L, Zaknich A. Continuous-time adaptive critics. IEEE Trans Neural Nets 2007;18(3):631–47.
[25] Lewis FL, Vamvoudakis KG. Optimal adaptive control for unknown systems using output feedback by reinforcement learning methods. In: Proc. of 8th IEEE international conference on control and automation (ICCA), Xiamen, China, 9–11 June 2010: p. 2138–45.
[26] Draguna Vrabie, Frank Lewis. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. Neural Networks 2009;22(3):237–46.
[27] Vrabie D, Lewis FL. Adaptive optimal control algorithm for continuous-time nonlinear systems based on policy iteration. In: Proc. of 47th IEEE Conf. On Decision and Control, Cancun, Mexico, Dec. 9–11, 2008: p. 73–9.
[28] Vamvoudakis KG, Lewis FL. Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica 2010;46(5):878–88.
[29] Vrabie D, Pastravanu O, Abu-Khalaf M, Lewis FL. Adaptive optimal control for continuous-time linear systems based on policy iteration. Automatica 2009;45(2):477–84.
[30] Vrabie D, Pastravanu O, Lewis FL. Policy iteration for continuous-time systems with unknown internal dynamics. In: Proc. of 15th mediterranean conf. on control & automation, Athens-Greece, July 27–29, 2007: p. T01–010.
[31] Kumar S, Padhi R, Behera L. Direct adaptive control using single network adaptive critic. In: Proc. of 2007 IEEE international conference on system of systems engineering, SoSE '07, 16–18 April, 2007: p. 1–6.
[32] Ali SF, Padhi R. Optimal blood glucose regulation of diabetic patients using single network adaptive critics. Optim Contr Appl Methods 2011;32(2):196–214.
[33] Prem KP, Behera L, Siddique NH, Prasad G. A T-S fuzzy based adaptive critic for continuous-time input affine nonlinear systems. In: Proc. of IEEE international conference on systems, man and cybernetics, SMC 2009, 11–14 Oct. 2009: p. 4329–34.
[34] Lin C-K. Radial basis function neural network-based adaptive critic control of induction motors. Appl Soft Comput 2011;11(3):3066–74.
[35] Kulkarni RV, Venayagamoorthy GK. Adaptive critics for dynamic optimization. Neural Networks 2010;23(5):587–91.