A reinforcement learning adaptive fuzzy controller for robots

Chuan-Kai Lin *

Department of Electrical Engineering, Chinese Naval Academy, 669 Chun Hsiao Road, Kaohsiung 813, Taiwan

Received 2 January 2001; received in revised form 27 November 2001; accepted 5 June 2002

Abstract

In this paper, a new reinforcement learning scheme is developed for a class of serial-link robot arms. Traditional reinforcement learning is the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment. In the proposed reinforcement learning scheme, an agent is employed to collect signals from a fixed gain controller, an adaptive critic element and a fuzzy action-generating element. The action generating element is a fuzzy approximator with a set of tunable parameters, and the performance measurement mechanism sends an error metric to the adaptive critic element for generating and transferring a reinforcement learning signal to the agent. Moreover, a tuning algorithm of the proposed scheme that can guarantee both tracking performance and stability is derived from the Lyapunov stability theory. Therefore, a combination of adaptive fuzzy control and reinforcement learning scheme is also concerned with algorithms for eliminating a sequence of decisions from experience. Simulations of the proposed reinforcement adaptive fuzzy control scheme on the cart-pole balancing problem and a two-degree-of freedom (2DOF) manipulator, SCARA robot arm verify the effectiveness of our approach.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Reinforcement learning; Adaptive fuzzy controller

1. Introduction

The reinforcement learning systems has enjoyed considerable success in application to nonlinear control 1,13,3. Most of the successful results, however, are obtained when the desired trajectory is set-point. The performance of the reinforcement learning systems for tracking control is less satisfactory. Recently, reinforcement learning systems used parameterized function approximators such as neural networks 13 in order to generalize between similar situations and actions. In these cases there are no strong theoretical results on the accuracy of convergence and stability. Although the learning speed can be improved by temporal difference learning 16 and Q-learning methods 14, the tracking performance and the stability cannot be guaranteed.

* Tel.: 886-7-5834700 ext. 1312; fax: 886-7-5829681.
E-mail address: cklin@mail.cna.edu.tw (Chuan-Kai Lin).
Function approximation, which can be implemented by neural networks and fuzzy systems, is essential to reinforcement learning. In [5], the proposed fuzzy actor-critic learning (FACL) and fuzzy Q-learning (FQL) are reinforcement learning methods with function approximation implemented by fuzzy inference systems. However, both FACL and FQL based on Dynamic Programming principles cannot guarantee the system stability and performance. Furthermore, in spite of the optimal policy, the stability is a severe topic when the reinforcement learning controller is applied to the real system. Therefore, it needs control theory to provide a mathematical framework in which the design of reinforcement learning controller can be formulated. The linear quadratic regulation (LQR) can be combined with neural Q-learning to control nonlinear systems; however, LQR is not suitable for tracking problems.

For reinforcement learning systems, an agent must learn behavior through trial-and-error interactions with a dynamic environment. Adaptive control and fuzzy self-organizing control are also concerned with algorithms for improving a sequence of decisions from experience. The proposed reinforcement learning systems combined with adaptive control and fuzzy control will be capable of avoiding too many trial-and-error interactions according to the stability and performance of reinforcement learning systems can be resorted to the powerful mathematical analysis of adaptive control algorithms. The architecture of the proposed reinforcement learning systems is similar to the fuzzy self-organizing control, which has a critic element, performance measurement, fuzzy controller, and an agent. For guaranteeing the stability and performance, the feedback linearization control law will introduce additional elements and the adaptation law of the parameters of adaptive critic element and fuzzy approximator will be derived from Lyapunov theory.

This paper consists of five sections. The background for development of the reinforcement learning systems is provided in Section 2, based largely on the system presented in 1. In Section 3, the new scheme of the reinforcement learning systems and the corresponding adaptation laws are introduced. Stability of the closed-loop system is analyzed and the tracking performance is guaranteed. Section 4 illustrates the operation of the proposed reinforcement learning systems using SCARA robot arm simulations for demonstrating the effectiveness. The last section, Section 5, gives the conclusions.

2. Preliminaries

2.1. Mathematical notation

The trace of a square matrix $S = [s_{ij}] \in \mathbb{R}^{n \times n}$, $tr(S)$, is defined as the sum of diagonal elements of $S$. By the definition, it is obvious that the trace of a matrix and the trace of the transpose of a matrix are the same, that is, $tr(S) = tr(S^T)$. Given $B \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{n \times m}$, it can be easily proved that

$$tr(BC) = tr(CB).$$

(2.1)

The norm of a vector and the Frobenius norm used in the following sections are defined previously. Let $\mathbb{R}$ denote the real numbers, $\mathbb{R}^n$ denote the real $n$-vectors, and $\mathbb{R}^{m \times n}$ be the real $m \times n$ matrices.
The norm of a vector is defined as
\[ \| x \| = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \] (2.2)
and the Frobenius norm of a matrix \( A = [a_{ij}] \in \mathbb{R}^{n \times n} \) is defined by
\[ \| A \|_F = \left[ \text{tr}(A^T A) \right]^{1/2} = \sqrt{\sum_{i,j} a_{ij}^2}. \] (2.3)
The Frobenius norm is also compatible with the two-norm so that \( \| Ax \| \leq \| A \|_F \| x \| \).

2.2. Stability

Consider the nonlinear dynamical system given by  \( \dot{x} = f(x, u, t) \),  \( y = g(x, t) \), where \( x(t) \in \mathbb{R}^n \), \( u(t) \) is the input vector and \( y(t) \) is the output. If there exists a compact set \( V \subset \mathbb{R}^n \) such that for all \( x(t_0) = x_0 \in V \), there exists an \( \varepsilon > 0 \) and a number \( T(e, x_0) \) such that \( \|x(t)\| < \varepsilon \) for all \( t \geq t_0 + T \), then the solution is uniformly ultimately bounded (UUB).

2.3. Fuzzy systems

General fuzzy systems consist of fuzzification unit, fuzzy rule base, inference engine and defuzzification unit. The fuzzy system is a nonlinear system performing a real (nonfuzzy) mapping from \( U \in \mathbb{R}^n \) to \( \mathbb{R}^m \). Inside the fuzzy system, fuzzifier maps nonfuzzy input space to fuzzy sets and the defuzzifier works in the opposite way to map from fuzzy sets to nonfuzzy output space. Inference engine is the brain of the fuzzy system deciding how to use the fuzzy rule base. As to the rule base, it determines the computation accuracy, memory storage and computational efficiency. Therefore, the main part of fuzzy systems is the fuzzy rule base, which may be expressed in the form of linguistic IF-THEN rules from experts. Assume the fuzzy rule consists of \( N \) rules as follows:
\[ R_j \ (j \text{th rule}): \text{If } x_1 \text{ is } A_{j1} \text{ and } x_2 \text{ is } A_{j2} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then } y_1 \text{ is } B_{j1} \text{ and } y_2 \text{ is } B_{j2} \text{ and } \ldots \text{ and } y_m \text{ is } B_{jm}, \]\nwhere \( j = 1, 2, \ldots, N \), \( x_i \ (i=1, 2, \ldots, n) \) are the input variables to the fuzzy system, \( y_k \ (k=1, 2, \ldots, m) \) are the output variable of the fuzzy system, and \( A_{ji} \) and \( B_{jk} \) are linguistic terms characterized by their corresponding fuzzy membership functions \( \mu_{A_{ji}}(x_i) \) and \( \mu_{B_{jk}}(y_k) \), respectively. Each rule \( R_j \) can be viewed as a fuzzy implication.

The fuzzy logic system with center-average defuzzifier, product inference, and singleton fuzzifier is of the following form:
\[ y_k = \frac{\sum_{j=1}^{N} c_{jk} \left( \prod_{i=1}^{n} \mu_{A_{ji}}(x_i) \right)}{\sum_{j=1}^{N} \left( \prod_{i=1}^{n} \mu_{A_{ji}}(x_i) \right)}, \quad k = 1, \ldots, m, \] (2.4)
where \( \mu_{B_{jk}}(c_{jk}) = 1 \). Eq. (2.4) can be written as sum of firing strength of rules:
\[ y_k = \sum_{j=1}^{N} c_{jk} \phi_j, \quad k = 1, \ldots, m, \] (2.5)
where the firing strength \( \phi_j = \prod_{i=1}^{n} \mu_{A_i}(x_i)/\sum_{j=1}^{N} (\prod_{i=1}^{n} \mu_{A_i}(x_i)) \). The shape of the membership function can be chosen as triangle, bell (gaussian function). Since Wang \cite{17} proved that the fuzzy system can be viewed as a universal approximator to any arbitrary accuracy, fuzzy control law has been transformed to be linear feedback control combined with fuzzy systems approximating nonlinear part. Hence, by Lyapunov theory and other traditional control theories, the stability and system performance can be analyzed and guaranteed, respectively.

2.4. Dynamics of robot manipulators

Consider a rigid robot manipulator with \( n \) serial links described by the equation

\[
M(\theta)\ddot{\theta} + V_m(\theta, \dot{\theta})\dot{\theta} + G(\theta) + F\dot{\theta} + \tau_d = \tau
\]

with vector \( \theta \in \mathbb{R}^n \) being the joint position vector; \( M(\theta) \in \mathbb{R}^{n \times n} \) being a symmetric positive definite inertia matrix; \( V_m(\theta, \dot{\theta}) \) being a vector of Coriolis and centripetal torques; \( G(\theta) \in \mathbb{R}^n \) representing the gravitational torques; \( F=K_\omega + V_f \in \mathbb{R}^{n \times n} \) being a diagonal matrix consisting of the back emf coefficient matrix \( K_\omega \) and the viscous friction coefficients matrix \( V_f \); \( \tau_d \in \mathbb{R}^{n \times 1} \) being the unmodeled disturbances vector; and \( \tau \in \mathbb{R}^{n \times 1} \) being the vector of control input torques. The structural properties of the robot manipulator are hold such as skew-symmetry of matrix \( M - 2V_m \) and boundedness of \( M(\theta), V_m(\theta, \dot{\theta}) \) and \( \tau_d \).

3. Controller design

Based on the architecture of reinforcement learning control systems \cite{1}, we employ the fuzzy system corresponding to the associate search element (ASE) and modify the adaptive critic element (ACE). In the spirit of actor-critic reinforcement learning control, the weights of fuzzy approximator (ASE) will be tuned by the signal from ACE. Fig. 1 shows the architecture of the proposed reinforcement learning adaptive fuzzy controller (RLAFC) and the details of the overall system will be subsequently derived.

As in \cite{1}, the action-generating element is a fuzzy logic system tuned by the ACE, however, the fuzzy system is used to approximate the nonlinear part of the nonlinear system instead of generating control input directly for guaranteeing stability and performance. The behavior of the RLAFC is that the performance measurement unit measures the system states to generate an error metric signal vector \( s \) providing for the ACE to tune the fuzzy system. The ACE collecting the error metric signal generates a reinforcement signal vector \( r \) to tune the fuzzy system. In the following, the control law, weight updating law and stability analysis will be discussed.

In practical robotic systems, the load may vary while performing different tasks, the friction coefficients may change in different configurations and some neglected nonlinearities as backlash may appear as disturbances at control inputs, that is, the robot manipulator may receive unpredictable interference from the environment where it resides. Therefore, the control objective is to design a robust RLAFC so that the movement of robot arms follow the desired trajectory and all signals in the closed-loop system are bounded even when exogenous and endogenous perturbations are present.
Denote the tracking error vector $e(t)$ and error metric $s(t)$ as

$$e(t) = \theta_d(t) - \theta(t),$$
$$s(t) = \dot{e}(t) + \Lambda e(t),$$

(3.1)

where $\theta_d(t)$ is the desired robot manipulator trajectory vector and $\Lambda = \Lambda^T > 0$. It is typical to define an error metric $s(t)$ to be a performance measure. When the error metric $s(t)$ is smaller, the system performance is better. Therefore, differentiating $s(t)$ and using (3.1), the dynamics of robot arms can be expressed in terms of $s(t)$ as

$$M \dot{s} = -V_m s + f + \tau_d - \tau,$$

(3.2)

where the unknown function $f$ is given by

$$f = M(\theta)(\dot{\theta}_d^T + \Lambda \dot{e}) + V_m(\theta, \dot{\theta})(\dot{\theta}_d^T + \Lambda e) + G(\theta) + F(\dot{\theta})$$

(3.3)

and the disturbance $\tau_d(t)$ is assumed to be bounded by $b_d$. It is also assumed that the complex nonlinear function $f$ can be represented by an ideal fuzzy approximator $C^T \phi(x)$ as follows:

$$f(x) = C^T \phi(x) + \varepsilon(x),$$

(3.4)

where $\varepsilon(x)$ is the approximation error, $C = [c_{jk}] \in \mathbb{R}^{N \times m}$ is the ideal weight matrix, $\phi(x) = [\phi_1 \ \phi_2 \ \ldots \ \phi_N]^T$, and $x(t)$ is given by

$$x(t) = [\dot{\theta}_d^T \ \dot{\theta}_d^T \ \dot{\theta}_d^T \ \dot{\theta}_d^T]^T.$$  

(3.5)
As we know, the function can be estimated by a fuzzy approximator
\[ \hat{f}(x) = C^T \phi(x). \] (3.6)

Define the control law as
\[ \tau = Ks + \hat{f} + \hat{d}, \] (3.7)

where the positive matrix \( K = K^T > 0 \), the output vector of the fuzzy approximator \( \hat{f} \) estimates \( f \) and \( \hat{d} \) is the robustifying term to attenuate disturbances. The architecture of the closed-loop system is shown in Fig. 1. Using the control input (3.7), we get closed-loop dynamics as
\[ Ms = -(K + V_m)s + \hat{f} + \tau_d - \hat{d}, \] (3.8)

where the approximation error \( \hat{f} \) is denoted as
\[ \hat{f} = f - \hat{f} = \hat{C}^T \phi(x) + \varepsilon(x) \] (3.9)

with \( \hat{C} = C - \check{C} \) and the robustifying term is given by
\[ \hat{d}(t) = k_d s(t)/\|s(t)\| \] (3.10)

with \( k_d \geq b_\tau \). Therefore, the rewritten closed-loop error dynamics (3.8) implies that the overall system is driven by the functional estimation error \( \hat{f} \). From Fig. 1, we can obtain the reinforcement signal from ACE as
\[ r(t) = s(t) + \|s\| \hat{W}^T \phi. \] (3.11)

The reinforcement signal is used to update the weight matrix \( \hat{C} \). Before the stability analysis, there are some assumptions that should be made.

**Assumption 1.** The norms of optimal weights, \( \|W\| \) and \( \|C\| \), are bounded by known positive real values, i.e., \( \|W\| \leq W_m \) and \( \|C\| \leq C_m \) with some known \( W_m \) and \( C_m \) (the norm is Frobenius norm).

**Assumption 2.** The approximation errors and disturbances are bounded, i.e., specified \( b_\varepsilon \) and \( b_\tau \) satisfying \( \|\varepsilon\| \leq b_\varepsilon \) and \( \|\tau_d\| \leq b_\tau \), respectively.

**Theorem 1.** Let the desired trajectory \( \theta_d \) be bounded. Suppose that Assumption 1 and 2 are hold. If the weight tuning laws provided for the fuzzy approximator and ACE are as follows:
\[ \dot{\hat{W}} = -K_W \|s\| \phi(\hat{C}^T \phi) - \eta K_W \|s\| \hat{W}, \] (3.12)
\[ \dot{\hat{C}} = K_C \phi r^T - \eta K_C \|s\| \hat{C}, \] (3.13)
where constant matrices $K_W$ and $K_C$ are positive and diagonal, $\eta_1$ and $\eta_2$ are positive constants, and the reinforcement signal $r(t)$ is provided by (3.11), then the control input $\tau(t)$ provided by (3.7) guarantees that the error metric $s(t)$, $\hat{W}$ and $\hat{C}$ are uniformly ultimately bounded.

**Proof.** Define the Lyapunov function candidate

$$V(t) = \frac{1}{2} s^T M s + \frac{1}{2} tr\{\hat{C}^T K_C^{-1} \hat{C}\} + \frac{1}{2} tr\{\hat{W}^T K_W^{-1} \hat{W}\}$$

(3.14)

with $\hat{W} = W - \hat{W}$. Evaluating the time derivative of $V(t)$ along the trajectories of the weight tuning laws (3.12) and (3.13) yield

$$\dot{V}(t) = \frac{1}{2} s^T M s + s^T M \dot{s} - tr\{\hat{C}^T K_C^{-1} \dot{C}\} - tr\{\hat{W}^T K_W^{-1} \dot{W}\}.$$ 

(3.15)

Substituting (3.8), and (3.10)–(3.13) into (3.15) we get

$$\dot{V}(t) \leq -s^T K s + s^T \varepsilon + \|s\|tr\{-\hat{C}^T \phi\hat{W}^T \phi^T + \eta \hat{C}^T \hat{C} + \hat{W}^T \phi \hat{C}^T \phi^T + \eta \hat{W}^T \hat{W}\}$$

(3.16)

by applying the property of skew-symmetry of matrix $\dot{M} - 2V_m$ and the assumption $k_d \geq b_t$. After some manipulation, we can obtain the following:

$$\dot{V}(t) \leq -s^T K s + s^T \varepsilon + \|s\|tr\{-\hat{C}^T \phi\hat{W}^T \phi^T + \eta \hat{C}^T \hat{C} + \hat{W}^T \phi \hat{C}^T \phi^T + \eta \hat{W}^T \hat{W}\}.$$ 

(3.17)

Applying the matrix inequalities $\hat{C}^T \hat{C} \leq \|\hat{C}\|C_m - \|\hat{C}\|^2$ and $\hat{W}^T \hat{W} \leq \|\hat{W}\|W_m - \|\hat{W}\|^2$, (3.17) becomes

$$\dot{V}(t) \leq -s^T K s + \|s\|\varepsilon + \|s\|\left\{(W_m\|\phi\|^2 + C_m)\|\hat{C}\| + (C_m\|\phi\|^2 + W_m)\|\hat{W}\| - \eta\|\hat{C}\|^2 - \eta\|\hat{W}\|^2\right\}.$$ 

(3.18)

Furthermore, since $\|\phi\|^2 \leq N$, there results

$$\dot{V}(t) \leq -K_{\min}\|s\|^2 - \eta\|s\|\left\{\left(\|\hat{C}\| - \frac{NW_m + C_m}{2\eta}\right)^2 + \left(\|\hat{W}\| - \frac{NC_m + W_m}{2\eta}\right)^2\right\}$$

$$- \left[\left(\frac{NW_m + C_m}{2\eta}\right)^2 + \left(\frac{NC_m + W_m}{2\eta}\right)^2 + \frac{b_t}{\eta}\right],$$

(3.19)

where $K_{\min}$ is the minimum singular value of $K$. Thus, $\dot{V}$ is negative as long as

$$\|s\| > \eta\left[\left(\frac{NW_m + C_m}{2\eta}\right)^2 + \left(\frac{NC_m + W_m}{2\eta}\right)^2 + \frac{b_t}{\eta}\right]/K_{\min} = \delta_t \quad \text{or}$$

(3.20)
Clearly, from inequalities (3.20)–(3.22), we can define compact set \( \Omega \) as 
\[
\Omega = \left\{ (\|s\|, \|\hat{C}\|, \|\hat{W}\|) : 0 \leq \|s\| \leq \delta_s, \|\hat{C}\| \leq \delta_C, \|\hat{W}\| \leq \delta_W \right\}. 
\]
By Lyapunov theory, for \( s(t_0) \), \( \hat{C}(t_0) \) and \( \hat{W}(t_0) \) in \( \Omega \), there exist numbers \( T(\delta_s, \delta_C, \delta_W, s(t_0), \hat{C}(t_0), \hat{W}(t_0)) \) such that \( 0 \leq \|s\| \leq \delta_s, \|\hat{C}\| \leq \delta_C \) and \( \|\hat{W}\| \leq \delta_W \) for all \( t \geq t_0 + T \). In other words, \( \hat{V}(s, \hat{C}, \hat{W}) \) is negative outside the compact set \( \Omega \) and then \( s, \hat{C} \) and \( \hat{W} \) are UUB. □

**Remark 1.** The updating law for weights of ASE (fuzzy approximator) is mainly determined by a reinforcement signal with magnitude modification (the first term of (3.13)). As to the first term of (3.12), it consists of three signals: \( \|s\|, \phi \) and \( \hat{C}^T\phi \). It is reasonable according to ACE generating reinforcement signal to consider the performance \( \|s\|, \) firing strength \( \phi \), and the affected output \( \hat{C}^T\phi \). This indicates how the outputs of fuzzy approximator influence the changes in the adaptive critic element. In adaptive control, the phenomenon of the possible unboundedness of weight estimates will occur when the persistency of excitation (PE) condition fails to hold. Without the last terms of (3.12) and (3.13), the \( \|s\|, \phi, r \) and \( \hat{C}^T\phi \) should be persistently exciting signals. In other words, positive numbers \( T_i, \delta_i, \epsilon_i (i = 1, 2, 3) \) exist such that given \( t \geq t_0 \), there exists \( t_i \in [t, t + T_i] \) such that \( [t_i, t_i + \delta_i] \subset [t, t + T_i] \) and

\[
\frac{1}{T_i} \int_{t_i}^{t_i+\delta_i} g_s(\tau)g_3(\tau)^T \mathrm{d}\tau \geq \epsilon_i I \quad \forall t \geq t_0,
\] (3.23)

where \( g_1 = \|s\|, g_2 = r \) and \( g_3 = \hat{C}^T\phi \). The last terms of the weight tuning rules (3.12) and (3.13) for RLAFC, which are similar to the e-modification of adaptive control theory, are employed to guarantee the boundedness of weight estimates even though PE does not hold. Therefore, the proposed control scheme with actor-critic reinforcement learning rules can guarantee the boundedness of all signals generated in the closed-loop system without making any assumptions of PE conditions.

**Remark 2.** It can be found that an implicit parameter \( \eta \) in (3.20)–(3.22) determines the magnitudes of \( \|s\|, \|\hat{C}\| \) and \( \|\hat{W}\| \). A larger \( \eta \) will result in smaller convergence regions of \( \|s\|, \|\hat{C}\| \) and \( \|\hat{W}\| \), and vice versa. Another factor in (3.20) is \( K_{\min} \), which can be increased to reduce \( \|s\| \).

### 4. Simulation results

There are two examples in this section: the first is the cart-pole problem used for comparing RLAFC with other reinforcement learning controllers, and the second is controlling the SCARA robot for demonstrating the effectiveness of the proposed RLAFC.
Example 1. Controlling the cart-pole system can be viewed as the benchmark for reinforcement learning controllers. If the position of the cart on the track is not taken into consideration, then balancing a rigid pole mounted on a cart is similar to controlling a one-link robot. The dynamics of the cart-pole system can be described as follows:

\[
\dot{\theta} = \frac{g \sin \theta + \cos \theta \left[(-u - ml\dot{\theta}^2 \sin \theta + \mu_c \text{sgn}(\dot{x}))/(m_c + m)\right] - \mu_p \dot{\theta}/ml}{l[(4/3) - m \cos^2 \theta/(m_c + m)]},
\]

\[
\dot{x} = \frac{u + ml[\dot{\theta}^2 \sin \theta - \dot{\theta} \cos \theta] - \mu_c \text{sgn}(\dot{x})}{m_c + m}
\]

with \(g = -9.8 \text{ m/s}^2\) the acceleration due to gravity, \(m = 0.1 \text{ kg}\) the mass of pole, \(m_c = 1 \text{ kg}\) the mass of cart, \(l = 0.5 \text{ m}\) the half-pole length on cart, \(\mu_c = 0.0005\) the coefficient of friction of cart on track, \(x\) the position of the cart on the track, \(\theta\) the angle of the pole with the vertical and \(u\) the force applied to cart 1. Therefore, structural property \(\dot{M} - 2\dot{V}_n = 0\) is satisfied according to \(M(\dot{\theta})=(4/3) - (m \cos^2 \theta)/(m_c + m)\) and \(V_n(\theta, \dot{\theta})=ml\dot{\theta} \cos \theta \sin \theta/(m_c + m)\). The inputs of fuzzy system are chosen as \(x_1 = \theta, x_2 = \dot{\theta}, x_3 = x\) and \(x_4 = \dot{x}\). The format of jth rule (\(j=27(l_4 - 1) + 9(l_3 - 1) + 3(l_2 - 1) + (l_1 - 1)\)) is as follows:

If \(x_1 \) is \(A_{j1}^{l_1}\) and \(x_2 \) is \(A_{j2}^{l_2}\) and \(x_3 \) is \(A_{j3}^{l_3}\) and \(x_4 \) is \(A_{j4}^{l_4}\) then \(y_1 \) is \(B_{j1}\), where \(j=1, \ldots, 81, l_i = 1, 2, 3\) for \(i=1, 2, 3, 4\), \(y_1 \) is the output of fuzzy approximator. The membership functions for inputs are selected as

\[
\mu_{A_{ji}^{l_i}}(x_i) = e^{-((x_i-x_i^j)/\sigma_i)},
\]

where \(x_i^j = a_i + b_i(l_i - 1)\) with \(a_1=a_2=-1, a_3=a_4=-10, b_1=b_3=1, b_2=b_4=10\), \(\sigma_1=\sigma_3=0.5\), and \(\sigma_2=\sigma_4=8\). The membership functions for outputs are singletons. As other experiments of reinforcement learning controllers for the cart-pole system, the simulation will only concern about
Table 1
Reinforcement learning methods comparison on the cart-pole problem

<table>
<thead>
<tr>
<th>Learning method</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHC</td>
<td>26</td>
</tr>
<tr>
<td>Lee’s</td>
<td>8</td>
</tr>
<tr>
<td>FACL</td>
<td>6</td>
</tr>
<tr>
<td>RLAFC</td>
<td>1</td>
</tr>
</tbody>
</table>

balancing the pole on the cart 1,5,7. The initial weights for $\hat{C}$ and $\hat{W}$ are small random numbers and the design parameters are chosen as $K=50$, $\eta=1$, $K_W=50$ and $K_C=50$. Suppose each trial has 6 s. The initial conditions are $\theta=45^\circ$, $\dot{\theta}=0$, $x=0$ and $\dot{x}=0$. Fig. 2 shows simulation results for balancing angle of the pole. The comparison with AHC 1, Lee’s 7, FACL approaches 5 and proposed RLAFC are listed in Table 1. The proposed RLAFC needs only 4 s to complete the control task, however, the other three approaches need at least 6 trials (each trial has 500 000 time steps). This demonstrates the faster convergence rate of the RLAFC.

Example 2. In this example, computer simulations were conducted on the 2DOF SCARA robot manipulator to verify the availability and tracking performance of the proposed controller. The SCARA robot manipulator, which consists of two parallel revolute axis and two rigid links, is depicted in Fig. 3. The axes of the revolute joints are directed upwards in positive $z$ direction. Therefore, the angles of the first and second links were considered to be $\theta_1$ and $\theta_2$, respectively. The numerical values of parameters of the robot model were specified as that in 11. For demonstrating the tracking performance of our proposed controller, the desired trajectories for $\theta_1$ and $\theta_2$ were set as

$$\theta_{d1} = 0.5 + 0.2(\sin t + \sin 2t) \text{ (rad)} \quad \text{for } \theta_1$$

and

$$\theta_{d1} = 13 - 0.1(\sin t + \sin 2t) \text{ (rad)} \quad \text{for } \theta_2,$$

respectively.

The inputs of fuzzy system are chosen as $x_1=\theta_1$, $x_2=\dot{\theta}_1$, $x_3=\theta_2$ and $x_4=\dot{\theta}$. The format of $j$th rule ($j=27(l_4-1)+9(l_3-1)+3(l_2-1)+(l_1-1)$) is as follows:
If $x_1$ is $A_{j_1}^{l_1}$ and $x_2$ is $A_{j_2}^{l_2}$ and $x_3$ is $A_{j_3}^{l_3}$ and $x_4$ is $A_{j_4}^{l_4}$ then $\hat{f}_1$ is $B_{j_1}$ and $\hat{f}_2$ is $B_{j_2}$, where $j = 1, \ldots, 81, l_i = 1, 2, 3$ for $i = 1, 2, 3, 4$, $\hat{f}_1$ and $\hat{f}_2$ are the outputs of fuzzy approximator. The membership functions for inputs are chosen as

$$
\mu_{A_{j_i}^{l_i}}(x_i) = e^{-((x_i - x_i^{l_i})/\sigma_i)},
$$

where $x_i = a_i + b_i(l_i - 1)$ with $a_1 = a_3 = -2$, $a_2 = a_4 = -10$, $b_1 = b_3 = 2$, $b_2 = b_4 = 10$, $\sigma_1 = \sigma_3 = 1.5$, and $\sigma_2 = \sigma_4 = 8$. The membership functions for inputs are singletons. There are 81 rules in the rule base and each rule has four inputs and two outputs, i.e., there are 162 weights for ACE and 162 weights for fuzzy rule base should be tuned.

Fig. 4. Simulations for (a) $\theta_1(t)$ and (b) $\theta_2(t)$ using RLAFC controller with different $K$. 
The weights for adaptive critic element and fuzzy approximator are small random numbers. Fig. 4 shows that the tracking errors are decreased as the gain matrix $K$ is increased as expected. Although we can only guarantee $s$ is UUB, the tracking errors can be kept very small. The reinforcement learning signals for $K=diag\{500,500\}$, $\eta=1$, $K_W=diag\{500,500\}$ and $K_C=diag\{50,50\}$ are shown in Fig. 5. It can be found that $r(t)$ are bounded. Moreover, random noises presented at joint 1 and joint 2 with bounds 10 and 1 Nm, respectively. Fig. 6 shows the RLAFC can suppress the external disturbances with the same parameters $K=diag\{500,500\}$, $\eta=1$, $K_W=diag\{500,500\}$ and $K_C=diag\{50,50\}$. All these simulations were carried out using Delphi programs on a PC with AMD-Duron 700 MHz CPU and 128 MB RAM, and the running time is about 20 s. For
traditional reinforcement learning control, the controller can only deal with set-point regulation and it takes many trials to achieve an acceptable performance. Therefore, the proposed RLAFC is more effective.

5. Conclusions

In this paper, the development of a reinforcement learning adaptive fuzzy control system has been presented for tracking control of a robot with uncertainties. By the weight updating rules of RLAFC developed from Lyapunov stability theorem, we can show that all signals in the closed-loop system are bounded without any assumptions of PE conditions to make the controller robust even in
the presence of approximation errors and external disturbances. All simulation results show that the proposed RLAFC converges fast to a very small error region that improves the convergence rate of traditional reinforcement learning controller.

References