Reinforcement Learning-based Output Feedback Control of Nonlinear Systems with Input Constraints

P. He and S. Jagannathan

Abstract—**A novel neural network (NN) -based output feedback controller with magnitude constraints is designed to deliver a desired tracking performance for a class of multi-input-multi-output (MIMO) discrete-time strict feedback nonlinear systems. Reinforcement learning in discrete time is proposed for the output feedback controller, which uses three NNs: 1) a NN observer to estimate the system states with the input-output data; 2) a critic NN to approximate certain *strategic* utility function; and 3) an action NN to minimize both the *strategic* utility function and the unknown dynamics estimation errors. The magnitude constraints are manifested as saturation nonlinearities in the output feedback controller design. Using the Lyapunov approach, the uniformly ultimate boundedness (UUB) of the state estimation errors, the tracking errors and weight estimates is shown.**

I. Introduction

THE output feedback controller schemes are necessary when certain states of the plant become unavailable for measurement. Moreover, the separation principle does not hold for nonlinear systems, even when an exponentially decaying state estimation error can lead to instability at finite escape time [1]. Consequently, the output feedback control design is quite difficult.

Several output feedback controller designs in discrete time are proposed for the signal-input-single-out (SISO) nonlinear systems [2-4]. In particular, a backstepping-based adaptive output feedback controller scheme is presented [2] for the control of a class of strict feedback nonlinear systems, where a rank condition is required to ensure the boundedness of all signals. In [3], a discrete-time NN output feedback controller is designed for a class of nonlinear systems expressed in input-output fashion, where the system is assumed to be minimum phase. A deadzone algorithm is used to develop a well-defined controller. In [4], two discrete output feedback control schemes are given based on a causal input-output representation and an adaptive NN observer, respectively. The semi-globally *UUB* of the closed-loop systems is shown.

In this paper, the output feedback design using adaptive critic neural network (NN) architecture is considered for an unknown MIMO nonlinear discrete system. The reinforcement learning-based adaptive critic NN approach [5-9] has emerged as a promising tool to develop optimal NN controllers due to its potential to find approximate solutions to dynamic programming, where a *strategic* utility function, which is considered as the long-term system performance measure, can be optimized. The adaptive critic output feedback NN controller consists of: 1) a NN observer to estimate the system states with the input-output data, 2) an action NN to drive the output to track the reference signal and to minimize both the strategic utility function and the unknown dynamics estimation errors, and 3) an adaptive critic NN to approximate certain *strategic* utility function and to tune the weights of the action NN. With incomplete information of the system states and dynamics, an approximate optimization is accomplished using the proposed controller. Further, the actuator constraints are manifested as saturation nonlinearities during the controller development in contrast to other works where no explicit magnitude constraints are treated [1-9].

Besides optimization, contributions of this paper can be summarized as follows: 1) the demonstration of the UUB of the overall system is shown even in the presence of NN approximation errors and bounded unknown disturbances unlike in the existing adaptive critic works [6-9] where the convergence is given under ideal circumstances; 2) the NN weights are tuned online instead of offline training that is commonly employed in adaptive critic design [5]; and 3) the LIP assumption is overcome along with the persistent excitation (PE) condition requirement [4] both in NN observer and controller designs.

II. Background

A. Nonlinear System Description

Consider the following nonlinear system, to be controlled, given in the following form

$$\begin{aligned} x_1(k+1) &= x_2(k) \\ &\cdot \\ &\cdot \\ x_n(k+1) &= f(x(k)) + g(x(k))u(k) + d'(k) \end{aligned}, \quad (1)$$

$$y(k) = x_1(k), \quad (2)$$

The authors are with the Department of Electrical and Computer Engineering, The University of Missouri-Rolla, 1870 Miner Circle, Rolla, MO 65409. Contact author's email address: ph8p5@umr.edu.

Research supported in part by a NSF grant ECS #0296191.

with state $x(k)=\left[x_1^T(k),x_2^T(k),\cdots,x_n^T(k)\right]^T\in R^{nm}$, and each $x_i(k)\in R^m$, $i=1,\cdots,n$ is the state at time instant k, $f(x(k))\in R^m$ is the unknown nonlinear function vector, $g(x(k))\in R^{m\times m}$ is a diagonal matrix of unknown nonlinear functions, $u(k)\in R^m$ is the control input vector and $d'(k)\in R^m$ is the unknown but bounded disturbance vector, whose bound is assumed to be a known constant, $\|d'(k)\|\le d_m'$. The Frobenius norm [10] is used through this paper. It is assumed that the output, $y(k)\in R^m$, is known at the *kth* instant and the state vector $x_i(k)\in R^m$, $i=2,\cdots,n$ is considered unavailable at *kth* step.

Assumption 1: Let the diagonal matrix $g(x(k))\in R^{m\times m}$ be a positive definite matrix for each $x(k)\in R^{nm}$, let $g_{\min}\in R$ and $g_{\max}\in R$ be the minimum and maximum eigenvalues of the matrix $g(x(k))\in R^{m\times m}$ respectively, with $0<g_{\min}<g_{\max}$.

III. NN Observer Design

A. Observer Structure

For the system (1) & (2), we use the following state observer to estimate the state $x(k)$.

$$\begin{aligned}\hat{x}_1(k)&=\hat{x}_2(k-1)\\&\vdots\\\hat{x}_n(k)&=\hat{w}_1^T(k-1)\phi_1(v_1^T\hat{z}_1(k-1))=\hat{w}_1^T(k-1)\phi_1(\hat{z}_1(k-1))\end{aligned}\tag{3}$$

where $\hat{x}_i(k)\in R^m$ is the estimated state of $x_i(k)\in R^m$ with $i=1,\cdots,n$ and $\hat{z}_1(k-1)=\left[\hat{x}_1^T(k-1),\ldots,\hat{x}_n^T(k-1),u^T(k-1)\right]^T\in R^{(n+1)m}$ is the input vector to the observer NN at *kth* instant, $\hat{w}_1(k-1)\in R^{n_1\times m}$ and $v_1\in R^{(n+1)m\times n_1}$ denote the output and hidden layer weights, the hidden layer activation function $\phi_1(\hat{z}_1(k-1))\in R^{n_1}$ represents $\phi_1(v_1^T\hat{z}_1(k-1))$, and n_1 is the number of the nodes in the hidden layer. It is demonstrated in [11] that, if the hidden layer weights, v_1, is chosen initially at random and kept constant and the number of hidden layer nodes is sufficiently large, the NN approximation error can be made arbitrarily small since the hidden layer NN activation function vector forms a basis.

B. Observer Error Dynamics

Define the state estimation error by

$$\tilde{x}_i(k)=\hat{x}_i(k)-x_i(k),\quad i=1,\ldots,n,\tag{4}$$

where $\tilde{x}_i(k)\in R^m$, $i=1,\ldots,n$, is the state estimation error. In fact, the observer NN approximates the nonlinear function given by $f(x(k-1))+g(x(k-1))u(k-1)$. This nonlinear function can be expressed as

$$\begin{aligned}f(x(k-1))+g(x(k-1))u(k-1)&=w_1^T\phi_1(v_1^Tz_1(k-1))+\varepsilon_1(z_1(k-1))\\&=w_1^T\phi_1(k-1)+\varepsilon_1(z_1(k-1)),\end{aligned}\tag{5}$$

where $w_1\in R^{n_1\times m}$ is the target NN weight matrix, $\varepsilon_1(z_1(k-1))$ is the NN approximation error, and the NN input is given by $z_1(k-1)=\left[x_1^T(k-1),\ldots,x_n^T(k-1),u^T(k-1)\right]^T\in R^{(n+1)m}$.

Combining (3), (4) and (5) to get

$$\tilde{x}_n(k)=\hat{x}_n(k)-x_n(k)=\xi_1(k-1)+d_1(k-1),\tag{6}$$

where

$$\tilde{w}_1(k-1)=\hat{w}_1(k-1)-w_1,\tag{7}$$

$$\xi_1(k-1)=\tilde{w}_1^T(k-1)\phi_1(\hat{z}_1(k-1)),\tag{8}$$

$$\phi_1(\tilde{z}_1(k-1))=\phi_1(\hat{z}_1(k-1)-\phi_1(z_1(k-1)),\tag{9}$$

$$d_1(k-1)=w_1^T\phi_1(\tilde{z}_1(k-1))-(\varepsilon_1(z_1(k-1))+d'(k-1)).\tag{10}$$

The dynamics of the estimation error using (4) and (6) is obtained as

$$\begin{aligned}\tilde{x}_1(k)&=\tilde{x}_2(k-1)\\&\vdots\\\tilde{x}_n(k)&=\xi_1(k-1)+d_1(k-1)\end{aligned}\tag{11}$$

IV. Output Feedback Controller Design

Our objective is to design an adaptive critic NN output feedback controller for the system (1) and (2) such that 1) all the signals in the closed-loop system remain *UUB*; 2) the state $x(k)$ follows a desired trajectory $Y_d(k)=[y_d^T(k),\cdots,y_d^T(k+n-1)]^T\in R^{nm}$, with $y_d(k)\in R^m$ and $y_d(k+i)$ represent the future value of $y_d(k)$, $i=1,\ldots,n-1$; and 3) a long-term system performance index is optimized.

Assumption 2: The desired trajectory, $Y_d(k)$, is a smooth function and it is bounded over the compact subset of R^{nm}.

A. Auxiliary Controller Design

Define the tracking error between actual and desired trajectory as

$$e_i(k+1)=x_i(k+1)-y_d(k+i),\ i=1,\ldots,n,\tag{12}$$

Equation (1) can be rewritten as

$$e_n(k+1)=f(x(k))+g(x(k))u(k)+d'(k)-y_d(k+n).\tag{13}$$

Define the desired auxiliary control signal as

$$v_d(k)=g^{-1}(x(k))(-f(x(k)+y_d(k+n)+l_1e_n(k)),\tag{14}$$

where $l_1\in R^{m\times m}$ is a design matrix selected such that the tracking error, $e_n(k)$, is bounded.

Since $f(x(k))$ and $g(x(k))$ are unknown smooth functions, the desired auxiliary feedback control input $v_d(k)$ cannot be implemented. From (14) and the Assumptions 1 and 2, $v_d(k)$ can be approximated by the action NN

$$v_d(k) = w_2^T \phi_2(v_2^T s(k)) + \varepsilon_2(s(k)) = w_2^T \phi_2(s(k)) + \varepsilon_2(s(k)), \quad (15)$$

where $s(k) = \left[x^T(k), e_n^T(k)\right]^T \in R^{(n+1)m}$ is the NN input, $w_2 \in R^{n_2 \times m}$ and $v_2 \in R^{(n+1)m \times n_2}$ denote the output and hidden layer target weights, the hidden layer activation function $\phi_2(s(k)) \in R^{n_2}$ represents $\phi_2\left(v_2^T s(k)\right)$, $\varepsilon_2(s(k))$ is the action NN approximation error, and n_2 is the number of the nodes in the hidden layer.

Replacing the actual states with their estimated values, (15) can be expressed as

$$v(k) = \hat{w}_2^T(k)\phi_2(v_2^T \hat{s}(k)) = \hat{w}_2^T(k)\phi_2(\hat{s}(k)), \quad (16)$$

where $\hat{w}_2(k) \in R^{n_2 \times m}$ is the actual weight matrix, the action NN input is given by $\hat{s}(k) = \left[\hat{x}^T(k), \hat{e}_n^T(k)\right]^T \in R^{(n+1)m}$, where $\hat{e}_n(k) \in R^m$ is referred as the modified tracking error, which is defined between the estimated state and the desired trajectory as

$$\hat{e}_i(k+1) = \hat{x}_i(k+1) - y_d(k+i), \quad i = 1, \ldots, n-1, \quad (17)$$

and

$$\hat{e}(k) = \begin{bmatrix} \hat{x}_1(k) - y_d(k) \\ \vdots \\ \hat{x}_n(k) - y_d(k+n-1) \end{bmatrix}, \quad (18)$$

B. Controller Design with Magnitude Constraints

By applying the magnitude constraints, the actual control input $u(k) \in R^m$ is now given by

$$u(k) = \begin{cases} v(k) & if \|v(k)\| \le u_{\max} \\ u_{\max}\,\mathrm{sgn}(v(k)) & if \|v(k)\| \ge u_{\max} \end{cases}, \quad (19)$$

where $u_{\max}$ is the actuator limit.

Case 1: $\|v(k)\| \le u_{\max}$

In this case, the control input $u(k) = v(k)$. Substituting (14), (15) and (16) into (13) yields

$$e_n(k+1) = l_1 e_n(k) + g(x(k))\zeta_2(k) + d_2(k), \quad (20)$$

where

$$\tilde{w}_2(k) = \hat{w}_2(k) - w_2, \quad (21)$$

$$\xi_2(k) = \tilde{w}_2^T(k)\phi_2(\hat{s}(k)), \quad (22)$$

$$\phi_2(\tilde{s}(k)) = \phi_2(\hat{s}(k)) - \phi_2(s(k)), \quad (23)$$

$$d_2(k) = g(x(k))\left(w_2^T \phi_2(\tilde{s}(k)) - \varepsilon_2(s(k))\right) + d'(k). \quad (24)$$

Thus, the tracking error dynamics is given by

$$\begin{aligned} e_1(k+1) &= e_2(k) \\ &\vdots \\ e_n(k+1) &= l_1 e_n(k) + g(x(k))\zeta_2(k) + d_2(k) \end{aligned}, \quad (25)$$

Case 2: $\|v(k)\| \ge u_{\max}$

In this case, the control input $u(k) = u_{\max}\,\mathrm{sgn}(v(k))$. Combining with (13), (14), (15) and (16) to get

$$\begin{aligned} e_n(k+1) &= f(x(k)) + g(x(k))u(k) + d'(k) - y_d(k+n) \\ &= f(x(k)) + g(x(k))\left(u(k) + v_d(k) - v_d(k)\right) + d'(k) - y_d(k+n) \\ &= l_1 e_n(k) + g(x(k))(u_{\max}\,\mathrm{sgn}(v(k)) - w_2^T \phi_2(s(k)) - \varepsilon_2(s(k))) + d'(k) \\ &= l_1 e_n(k) + d_2'(k), \end{aligned} \quad (26)$$

where

$$d_2'(k) = g(x(k))(u_{\max}\,\mathrm{sgn}(v(k)) - w_2^T \phi_2(s(k)) - \varepsilon_2(s(k))) + d'(k), \quad (27)$$

Therefore, for the *Case 2*, the tracking error dynamics can be written as

$$\begin{aligned} e_1(k+1) &= e_2(k) \\ &\vdots \\ e_n(k+1) &= l_1 e_n(k) + d_2'(k) \end{aligned}, \quad (28)$$

V. Weight Updates for Guaranteed Performance

A. Weights Updating Rule for the Observer NN

The observer NN weight update is driven by the state estimation error $\tilde{x}_1(k)$, i.e.,

$$\hat{w}_1(k+1) = \hat{w}_1(k) - \alpha_1 \phi_1(\hat{z}_1(k))\left(\hat{w}_1^T(k)\phi_1(\hat{z}_1(k)) + l_2 \tilde{x}_1(k)\right)^T, \quad (29)$$

where $l_2 \in R^{m \times m}$ is a design matrix, and $\alpha_1 \in R$ is the observer NN adaptation gain.

B. Strategic Utility Function

The utility function $p(k) = [p_i(k)]_{i=1}^m \in R^m$ is defined based on the modified tracking error $\hat{e}(k)$ and it is given by

$$p_i(k) = \begin{cases} 0, & if \|\hat{e}_i(k)\| \le c \\ 1, & otherwise \end{cases}, \quad i = 1, 2, \ldots, m \quad (30)$$

where $c \in R^+$ is a pre-defined threshold. The utility function $p(k)$ is viewed as the current system performance index: $p_i(k) = 0$ and $p_i(k) = 1$ refers to the good and unacceptable tracking performance respectively.

The *strategic* utility function $Q(k) \in R^m$, is defined as

$$Q(k) = \alpha^N p(k+1) + \alpha^{N-1} p(k+2) + \cdots + \alpha^{k+1} p(N), \quad (31)$$

where $\alpha \in R$ and $0 < \alpha < 1$, and N is the final time instant. The term $Q(k)$ is viewed here as the future system performance measure.

C. Design of the Critic NN

The critic NN is used to approximate the *strategic* utility function $Q(k)$. The prediction error is defined as

$$e_c(k) = \hat{Q}(k) - \alpha\left(\hat{Q}(k-1) - \alpha^N p(k)\right), \quad (32)$$

where the subscript "c" stands for the "critic" and

$$\hat{Q}(k) = \hat{w}_3^T(k)\phi_3\left(v_3^T\hat{x}(k)\right) = \hat{w}_3^T(k)\phi_3(\hat{x}(k)), \quad (33)$$

and $\hat{Q}(k) \in R^m$ is the critic signal, $\hat{w}_3(k) \in R^{n_3 \times m}$ and $v_3 \in \Re^{nm \times n_3}$ represent the matrix of weight estimates, $\phi_3(\hat{x}(k)) \in \Re^{n_3}$ is the activation function vector in the hidden layer, n_3 is the number of the nodes in the hidden layer, and the critic NN input is the system state estimate $\hat{x}(k) = \left[\hat{x}_1^T(k), \ldots, \hat{x}_n^T(k)\right]^T \in \Re^{nm}$. The objective function to be minimized by the critic NN is defined as

$$E_c(k) = \frac{1}{2}e_c^T(k)e_c(k). \quad (34)$$

The weight update rule for the critic NN is a gradient-based adaptation, which is given by

$$\hat{w}_3(k+1) = \hat{w}_3(k) + \Delta\hat{w}_3(k), \quad (35)$$

where

$$\Delta\hat{w}_3(k) = \alpha_3\left[-\frac{\partial E_c(k)}{\partial \hat{w}_3(k)}\right]. \quad (36)$$

Before we proceed further, the following Lemma is needed.

Lemma 1: Given the matrices $A \in R^{m\times m}$, $X \in R^{n\times m}$ and vectors $b \in R^n$ and $q \in R^m$, the derivative of the following scalar with respect to the matrix X is given by

$$\frac{\partial\left(\left(AX^Tb+q\right)^T\left(AX^Tb+q\right)\right)}{\partial X} = 2b\left(A^T\left(AX^Tb+q\right)\right)^T. \quad (37)$$

Using *Lemma 1* and (36), the weight updating rule for the adaptive critic NN is given by

$$\hat{w}_3(k+1) = \hat{w}_3(k) - \alpha_3\phi_3(\hat{x}(k))\left(\hat{Q}(k) + \alpha^{N+1}p(k) - \alpha\hat{Q}(k-1)\right)^T, \quad (38)$$

where $\alpha_3 \in R$ is the critic NN adaptation gain.

D. Weight Updating Rule for the Action NN

The action NN weights $\hat{w}_2^T(k)$ are tuned by using the functional estimation error, $\zeta_2(k)$, and the error between the desired *strategic* utility function $Q_d(k) \in R^m$ and the critic signal $\hat{Q}(k)$. Define

$$e_a(k) = \sqrt{g(x(k))}\zeta_2(k) + \left(\sqrt{g(x(k))}\right)^{-1}\left(\hat{Q}(k) - Q_d(k)\right), \quad (39)$$

where $\zeta_2(k)$ is defined in (22), $\sqrt{g(x(k))} \in R^{m\times m}$ is the principle square root of the diagonal positive definite matrix $g(x(k))$, i.e., $\left(\sqrt{g(x(k))}\right)^2 = g(x(k))$, and $\left(\sqrt{g(x(k))}\right)^T = \left(\sqrt{g(x(k))}\right)$, $e_a(k) \in R^m$, and the subscript "a" stands for the "action NN".

The desired *strategic* utility function $Q_d(k)$ is taken as "0" [8], to indicate that at every step, the nonlinear system can track the reference signal well. Thus, (39) becomes

$$e_a(k) = \sqrt{g(x(k))}\zeta_2(k) + \left(\sqrt{g(x(k))}\right)^{-1}\hat{Q}(k), \quad (40)$$

The objective function to be minimized is given by

$$E_a(k) = \frac{1}{2}e_a^T(k)e_a(k), \quad (41)$$

Using Lemma 1 and (25), the gradient-based weight updating rule is given by

$$\hat{w}_2(k+1) = \hat{w}_2(k) - \alpha_2\phi_2(\hat{s}(k))\left(e_n(k+1) - l_1e_n(k) - d_2(k) + \hat{Q}(k)\right)^T, \quad (42)$$

where $\alpha_2 \in R$ is the action NN adaptation gain. Since $e_n(k+1)$ and $e_n(k)$ are unavailable, the modified tracking errors $\hat{e}_n(k+1)$ and $\hat{e}_n(k)$ respectively are used instead. In the ideal case, we take the $d_2(k)$ as zero to obtain the action NN $\hat{w}_2^T(k)\phi_2(\hat{s}(k))$ weight updating rule.

$$\hat{w}_2(k+1) = \hat{w}_2(k) - \alpha_2\phi_2(\hat{s}(k))\left(\hat{e}_n(k+1) - l_1\hat{e}_n(k) + \hat{Q}(k)\right)^T, \quad (43)$$

VI. Main Result

Assumption 3: Let w_1, w_2 and w_3 be the unknown output layer target weights for the observer, action and critic NNs, and assume that they are bounded above so that

$$\|w_1\| \le w_{1m}, \|w_2\| \le w_{2m}, \text{ and } \|w_3\| \le w_{3m}, \quad (44)$$

where $w_{1m} \in R$, $w_{2m} \in R$ and $w_{3m} \in R$ represent the bounds on the unknown target weights.

Fact 1: The activation functions are bounded by known positive values so that

$$\|\phi_i(k)\| \le \phi_{im}, \quad i = 1,2,3, \quad (45)$$

where $\phi_{im} \in R, i = 1,2,3$ is the upper bound for $\phi_i(k), i = 1,2,3$.

Assumption 4: The NN approximation errors $\varepsilon_1(z_1(k))$ and $\varepsilon_2(s(k))$ are bounded over the compact set $S \subset R^m$ by ε_{1m} and ε_{2m}, respectively [11].

Fact 2: The terms $d_1(k) \in R^m$, $d_2(k) \in R^m$ and $d_2'(k) \in R^m$ are bounded over the compact set $S \subset R^m$ by

$$\|d_1(k)\| \le d_{1m} = 2w_{1m}\phi_{1m} + d_m' + \varepsilon_{1m}, \quad (46)$$

where $d_{1m} \in R^+$ is the upper bound for $d_1(k)$,

$$\|d_2(k)\| \le d_{2m} = 2g_{\max}w_{1m}\phi_{1m} + \varepsilon_{2m} + d_m', \quad (47)$$

and

$$\|d_2'(k)\| \le d_{2m}' = g_{\max}\left(u_{\max} + w_{2m}\phi_{2m} + \varepsilon_{2m}\right) + d_m', \quad (48)$$

Theorem 1: Consider the system given by (1) and (2). Let the Assumptions 1 through 4 hold with the disturbance bound d_m' a known constant. Let the state estimates are provided by the observer (3), and the control input be given by (19). Let the observer NN $\hat{w}_1^T(k)\phi_1(\hat{z}_1(k))$, action NN $\hat{w}_2^T(k)\phi_2(\hat{s}(k))$, and the critic NN $\hat{w}_3^T(k)\phi_3(\hat{x}(k))$ weight tuning be given by (29), (43) and (38) respectively. Then the state estimation error $\tilde{x}_i(k)$, the tracking error $e_i(k)$, and the NN weight estimates, $\hat{w}_1(k)$, $\hat{w}_2(k)$ and $\hat{w}_3(k)$ are *UUB*, with the bounds specifically given by (A.4) through

(A.8) provided the controller design parameters are selected as:

$$\text{(a)}\ \ 0<\alpha_1\left\|\phi_1\left(\hat{z}_1(k)\right)\right\|^2<1, \tag{49}$$

$$\text{(b)}\ \ 0<\alpha_2\left\|\phi_2\left(\hat{s}(k)\right)\right\|^2<\frac{g_{\min}}{g_{\max}^2}, \tag{50}$$

$$\text{(c)}\ \ 0<\alpha_3\left\|\phi_3\left(\hat{x}(k)\right)\right\|^2<1, \tag{51}$$

$$\text{(d)}\ \ 0<\alpha<\frac{\sqrt{2}}{2}, \tag{52}$$

where α_1, α_2 and α_3 are NN adaptation gains, and α is employed to define the *strategic* utility function.

Proof: See Appendix. ■

Remark 1: The proposed scheme results in a well-defined controller by avoiding the problem of $\hat{g}(x(k))$ becoming zero.

Remark 2: The weights of the observer, action and critic NNs can be initialized at zero or random. This means that there is no explicit off-line learning phase needed.

VII. Simulation

The MIMO nonlinear system is described by

$$x_1(k+1)=x_3(k), \tag{53}$$

$$x_2(k+1)=x_4(k), \tag{54}$$

$$x_3(k+1)=-\frac{5}{8}\frac{x_1(k)}{\left(1+x_3^2(k)\right)}+x_3(k)-\frac{3}{\left(1+x_1^2(k)+x_3^2(k)\right)}u_1(k), \tag{55}$$

$$x_4(k+1)=-\frac{9}{16}\frac{x_2(k)}{\left(1+x_4^2(k)\right)}+x_4(k)-\frac{2}{\left(1+x_2^2(k)+x_4^2(k)\right)}u_2(k), \tag{56}$$

$$y(k)=\left[x_1(k),x_2(k)\right]^T. \tag{57}$$

where $x_i(k)\in R, i=1,\cdots,4$ is the state, $u_1(k)\in R$ and $u_2(k)\in R$ are the control inputs and $y(k)\in R^2$ is the system output. The objective is to track a reference signal using the proposed adaptive NN output feedback controller. The reference signal used was selected as $y_d(k)=\left[\sin(\omega kT+\xi),\sin(\omega kT+\xi+\pi)\right]^T$, $\omega=0.1$, $\xi=\frac{\pi}{2}$, with a sampling interval of *T=25msec*. The total simulation time is taken as *250* seconds. The actuator constraint is taken as 0.6. All the three NNs have 8 nodes in the hidden layer. For weight updating, the learning rate is selected as $\alpha_1=\alpha_3=0.001$ and $\alpha_2=0.01$. The parameter α is taken as 0.5. Both l_1 and l_2 are selected as $0.5I$, where I is a 2-by-2 identity matrix. All the initial weights are chosen at random in the interval [0,1] and all the activation functions are hyperbolic tangent sigmoid functions.

Figs 1 and 2 illustrate the good tracking performance of the adaptive output feedback NN controller with saturation. Fig. 3 depicts the bounded control inputs.

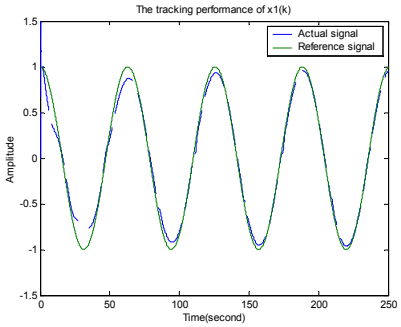

Fig. 1. Tracking performance of the state x1(k).

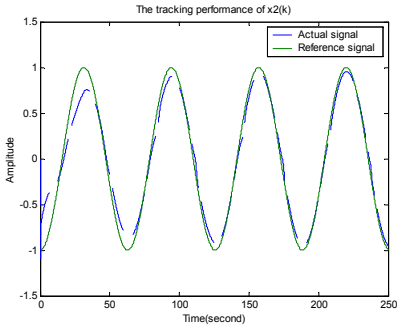

Fig. 2. Tracking performance of the state x2(k).

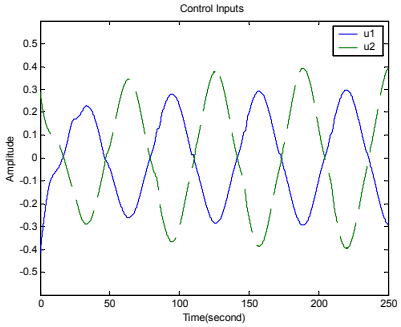

Fig. 3. NN control inputs with saturation.

VIII. Conclusion

A novel adaptive critic NN based output feedback controller with magnitude constraints is designed to deliver a desired tracking performance for a class of MIMO strict feedback nonlinear discrete-time systems. The adaptive critic NN structure optimizes certain *strategic* utility function. Magnitude constraints on the control input allow the designer to meet the physical limits of the actuator while meeting the closed-loop stability and tracking performance. The *UUB* of the closed-loop tracking and the estimation errors and NN weight estimates was demonstrated. Simulation results justify theoretical conclusions.

Appendix

Proof of Theorem 1

Case 1: $\left\|v(k)\right\|\le u_{\max}$. Define the Lyapunov function as

$$J(k)=\frac{\gamma_1}{2}\sum_{i=1}^{n}\left\|\tilde{x}_i(k-1)\right\|^2+\frac{\gamma_2}{2}\sum_{i=1}^{n}\left\|\tilde{x}_i(k)\right\|^2+\frac{\gamma_3}{3}\sum_{i=1}^{n}\left\|e_i(k)\right\|^2+\frac{\gamma_4}{3}\sum_{i=1}^{n}\left\|e_n(k)\right\|^2$$
$$+\frac{\gamma_5}{\alpha_1}tr\left(\tilde{w}_1^T(k-1)\tilde{w}_1(k-1)\right)+\frac{\gamma_6}{\alpha_1}tr\left(\tilde{w}_1^T(k)\tilde{w}_1(k)\right)$$
$$+\frac{\gamma_7}{\alpha_2}tr\left(\tilde{w}_2^T(k)\tilde{w}_2(k)\right)+\frac{\gamma_8}{\alpha_3}tr\left(\tilde{w}_3^T(k)\tilde{w}_3(k)\right)+\gamma_9\left\|\zeta_3(k)\right\|^2 \quad (A.1)$$

where $\gamma_i\in R^+, i=1,\ldots,9$ are design parameters. By using the observer (3), the control input (19) and the weight updating rules (29), (43) and (38), to get

$$\Delta J=-\frac{1}{2}\left(\gamma_1-4\gamma_5 l_{2\max}^2\right)\left\|\tilde{x}_1(k-1)\right\|^2-\frac{1}{2}\left(\gamma_2-4\gamma_6 l_{2\max}^2\right)\left\|\tilde{x}_1(k)\right\|^2$$
$$-\frac{\gamma_3}{3}\left\|e_1(k)\right\|^2-\frac{1}{3}\left(\gamma_4-3(\gamma_3+\gamma_4)l_{1\max}^2\right)\left\|e_n(k)\right\|^2$$
$$-\gamma_6\left(1-\alpha_1\left\|\phi_1(k)\right\|^2\right)\left\|\zeta_1(k)+l_2\tilde{x}_1(k)+w_1^T\phi_1(k)\right\|^2-\left(\gamma_6-\gamma_2-2\gamma_7'\right)\left\|\zeta_1(k)\right\|^2$$
$$-\gamma_5\left(1-\alpha_1\left\|\phi_1(k-1)\right\|^2\right)\left\|\zeta_1(k-1)+l_2\tilde{x}_1(k-1)+w_1^T\phi_1(k-1)\right\|^2$$
$$-\left(\gamma_5-\gamma_1-2\gamma_7' l_{2\max}^2\right)\left\|\zeta_1(k-1)\right\|^2$$
$$-\gamma_7\left(g_{\min}-\alpha_2\left\|\phi_2(k)\right\|^2 g_{\max}^2\right)\left\|\zeta_2(k)+\frac{\left(I-\alpha_2\left\|\phi_2(k)\right\|^2 g(x(k))\beta(k)\right)}{g_{\min}-\alpha_2\left\|\phi_2(k)\right\|^2 g_{\max}^2}\right\|^2$$
$$-\left(\gamma_7 g_{\min}-\gamma_3 g_{\max}^2-\gamma_4 g_{\max}^2\right)\left\|\zeta_2(k)\right\|^2$$
$$-\gamma_8\left(1-\alpha_3\left\|\phi_3(k)\right\|^2\right)\left\|\zeta_3(k)+w_3^T\phi_3(k)+\alpha^{N+1}p(k)-\alpha\hat{Q}(k-1)\right\|^2$$
$$-\left(\gamma_8-2\gamma_8\alpha^2-\gamma_7'\right)\left\|\zeta_3(k)\right\|^2+D_M^2, \quad (A.2)$$

where

$$D_M^2=\left(\gamma_1+\gamma_2+2\left(1+l_{2\max}^2\right)\gamma_7'\right)d_{1m}^2+\left(\gamma_3+\gamma_4+2\gamma_7'\right)d_{2m}^2+2\gamma_5 w_{1m}^2\phi_{1m}^2$$
$$+2\gamma_6 w_{1m}^2\phi_{1m}^2+6\gamma_8+2\left(\gamma_7'+3\gamma_8\left(1+\alpha^2\right)\right)w_{3m}^2\phi_{3m}^2. \quad (A.3)$$

This implies that $\Delta J(k)\le 0$ as long as (49) through (52) hold and the following conditions hold

$$\left\|\tilde{x}_1(k)\right\|\ge\max\left(\frac{\sqrt{2}D_M}{\sqrt{\gamma_1-4\gamma_5 l_{2\max}^2}},\frac{\sqrt{2}D_M}{\sqrt{\gamma_2-4\gamma_6 l_{2\max}^2}}\right) \quad (A.4)$$

or

$$\left\|e_1(k)\right\|\ge\max\left(\frac{\sqrt{3}D_M}{\sqrt{\gamma_3}},\frac{\sqrt{3}D_M}{\sqrt{\gamma_4-3(\gamma_3+\gamma_4)l_{1\max}^2}}\right) \quad (A.5)$$

or

$$\left\|\zeta_1(k)\right\|\ge\max\left(\frac{D_M}{\sqrt{\gamma_6-\gamma_2-2\gamma_7'}},\frac{D_M}{\sqrt{\gamma_5-\gamma_1-2\gamma_7' l_{2\max}^2}}\right) \quad (A.6)$$

or

$$\left\|\zeta_2(k)\right\|\ge\frac{D_M}{\sqrt{\gamma_7 g_{\min}-\gamma_3 g_{\max}^2-\gamma_4 g_{\max}^2}} \quad (A.7)$$

or

$$\left\|\zeta_3(k)\right\|\ge\frac{D_M}{\sqrt{\gamma_8-2\gamma_8\alpha^2-\gamma_7'}}. \quad (A.8)$$

Case 2: $\left\|v(k)\right\|>u_{\max}$.

The proof is similar to that in *Case 1* and it is omitted.

For both Case 1 and Case 2, $\Delta J(k)\le 0$ for all k is greater than zero. According to the standard Lyapunov extension theorem [10], this demonstrates that $\tilde{x}_1(k), e_1(k)$ and the weight estimation errors are *UUB*. The boundedness of $\left\|\zeta_1(k)\right\|$, $\left\|\zeta_2(k)\right\|$ and $\left\|\zeta_3(k)\right\|$ implies that $\left\|\tilde{w}_1(k)\right\|$, $\left\|\tilde{w}_2(k)\right\|$ and $\left\|\tilde{w}_3(k)\right\|$ and weight estimates $\hat{w}_1(k)$, $\hat{w}_2(k)$ and $\hat{w}_3(k)$ are bounded. Since $\tilde{x}_1(k)$ is bounded, from the estimation errors system given by (11), it implies that all the estimation errors are bounded. Similarly, based on the tracking error system (25) and (28), bounded $e_1(k)$ implies that all the tracking errors are bounded. Therefore all the signals in the observer-controller system are bounded.

ACKNOWLEDGMENT

The authors thank Mr. Zheng Chen for the help during the manuscript preparation.

REFERENCE

[1] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adaptive Control Design*, John Wiley & Sons, Inc., 1995.

[2] P. C. Yeh and P. V. Kokotovic, "Adaptive output feedback design for a class of nonlinear discrete-time systems", *IEEE Trans. Automat. Contr.*, vol. 40, no. 9, pp. 1663-1668, 1995.

[3] F. C. Chen and H. K. Khalil, "Adaptive control of a class of nonlinear discrete-time systems using neural networks", *IEEE Trans. Automat. Contr.*, vol. 40, no. 5, pp. 791-801, 1995.

[4] S. S. Ge, T. H. Lee, G. Y. Li, and J. Zhang, "Adaptive NN control for a class of discrete-time nonlinear systems, ", *Int. J. Contr.*, vol. 76, no. 4, pp. 334-354, 2003.

[5] P. J. Werbos, "Neurocontrol and supervised learning: An overview and evaluation", *Handbook of Intelligent Control*, edited by D. A. White and D. A. Sofge, Van Nostrand Reinhold, New York, 1992, pp. 65-90.

[6] J. J. Murray, C. Cox, G. G. Lendaris, and R. Saeks , "Adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern.,* vol. 32, no. 2, pp 140-153, 2002.

[7] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.

[8] J. Si and Y. T. Wang, "On-line learning control by association and reinforcement", *IEEE Trans. on Neural Networks*, vol. 12, no. 2, pp. 264 – 276, 2001.

[9] X. Lin and S. N. Balakrishnan, "Convergence analysis of adaptive critic based optimal control," *Proc. Amer. Contr. Conf.* pp. 1929 – 1933, 2000.

[10] F. L. Lewis, S. Jagannathan, and A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*, Taylor & Francis, PA, 1999.

[11] B. Igelnik and Y. H. Pao, "Stochastic choice of basis functions in adaptive function approximation and the functional-link net," *IEEE Trans. Neural Networks*, vol. 6, no. 6, pp. 1320 – 1329, 1995.