Online Optimal Adaptive Control of a Class of Uncertain Nonlinear Discrete-time Systems

Rohollah Moghadam, Pappa Natarajan, Krishnan Raghavan*, Sarangapani Jagannathan
*Department of Electrical and Computer Engineering ,*Department of Mathematics and Computer Science*
*Missouri University of Science and Technology, *Argonne National Laboratory (This work was done while at MST)*
Rolla, MO, USA,*Chicago, IL, USA
{moghadamr@umsystem.edu, npappa@annauniv.edu, krm9c@umsystem.edu, sarangap@umsystem.edu}

***Abstract*—In this paper, a multi-layer neural network (MNN) based online optimal adaptive regulation of a class of nonlinear discrete-time systems in affine form with uncertain internal dynamics is introduced. The multi-layer neural networks (MNN)-based actor-critic framework is utilized to estimate the optimal control input and cost function. The temporal difference (TD) error is derived from the difference between actual and estimated cost function. The MNN weights of both critic and actor are tuned at every sampling instant as a function of the instantaneous temporal difference and control policy errors. The proposed approach does not require the selection of any basis function and its derivatives. The boundedness of the system state vector and actor and critic NN weights are shown through Lyapunov theory. Extension of the proposed approach to MNNs with more hidden layers is discussed. Simulation results are provided to illustrate the effectiveness of the proposed approach.**

***Index Terms*—Optimal adaptive control, Multi-layer neural network, Discrete-time systems.**

I. Introduction

Optimal control of linear and nonlinear discrete-time systems using neural networks has been a most sought out area in control for the past several decades [1], [2] given the system dynamics. However, in many practical applications, the system dynamics are normally uncertain and it is difficult to obtain accurate knowledge of the dynamics.

To overcome the need for system dynamics, value and or policy iteration and optimal adaptive-based approaches using neural networks (NNs) are introduced in the literature [3]–[10]. The value-iteration technique using adaptive dynamic programming (ADP) has been proven successful in the case of general nonlinear systems [3], [4]. Optimal adaptive control using ADP has been extensively studied for both discrete and continuous-time systems [5], [6], [7], [8] using actor-critic networks. The actor-critic framework uses two NNs, one for approximating the value function and the other for the control action. The value and action NNs in the actor-critic framework are tuned in an iterative manner. For convergence, a large number of iterations within a sampling interval is needed which is a bottleneck for real-time control.

The author would like to thank Fulbright Association for the Fellowship and Anna University for the support.

In contrast, the authors in [9] introduced an optimal adaptive approach in finding the optimal regulator for a class of nonlinear discrete-time systems in affine form with unknown internal dynamics. The state vector and its history is used to derive the NN weight update laws for the critic network and the weights are tuned at the sampling interval without any iterations. The proposed approach employed a single-layer actor-critic NN with proper selection of basis function. As an extension to [9], the authors in [4] employed an additional NN identifier to learn the unknown dynamics of the nonlinear system by using a time-based ADP method using temporal difference (TD) error on an actor–critic structure with two single layer NNs. To overcome the challenges associated with selection of proper basis function for single layer NNs [4], [9] and to relax the need for computing their derivatives, two-layer NNs using backpropagation weight tuning are utilized in [10]. However, convergence and stability analysis are not reported.

The gradient descent-based weight tuning schemes are also reported in [11] for deep NNs, but suffers from the vanishing gradient problem [12] due to chain rule. Though this issue is typically overcome using rectified linear units (RELU) activation functions, it has been shown in [13] that vanishing gradients can still occur. The deep NN has been widely used in the applications of data clustering [14] and image processing [15]. Most of the reported works [4], [9], [16] have employed a single-layer critic and actor NNs, and multi-layer NN based optimal adaptive control technique using instantaneous TD error and control policy error has not been investigated yet. In addition, the stability analysis and the effect of additional hidden layers for control applications is not studied.

Therefore, this paper aims at the MNN online adaptive optimal regulation of a class of uncertain nonlinear discrete-time systems in affine form. The proposed scheme uses the state vector to obtain optimal control without the knowledge of the internal dynamics. One MNN is used to approximate the cost function and another for approximating the control policy. The weight update laws for hidden and output layers in the critic and actor NN are proposed as a function of TD error instead of considering the time history of system state vector.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Since the hidden layer weight tuning utilizes the TD error directly, the proposed learning scheme appears to mitigate the vanishing gradient problem that is commonly found in the literature with gradient-based weight tuning.

The major contributions of the paper include the: 1) derivation of novel weight update laws for critic and actor NNs using TD and control input errors, 2) development of an optimal adaptive regulator using multilayer NNs, and 3) simulation results to confirm the effectiveness of the proposed approach. The major benefit observed due to this effort is the extension of deep NNs for control applications by relaxing the need for the selection of basis functions, and overcoming the vanishing gradient problem that is commonly observed with deep NNs.

II. Background

In this section, the optimal control of a nonlinear discrete-time system is formulated. Consider the nonlinear discrete-time system in affine form described by

$$x(k+1) = f(x(k)) + g(x(k))u(x(k)) \ , \tag{1}$$

where $x(k) \in \Re^n, f(x(k)) \in \Re^n$ and the nonlinear input function $g(x(k)) \in \Re^{n\times m}$ satisfies $\|g(x(k))\|_F \leqslant g_M$ with $u(x(k)) \in \Re^m$ being the control input. The internal dynamics of the system $f(x(k)$ is assumed to be unknown and the nonlinear control coefficient matrix $g(x(k))$ is considered known. The objective is to generate the control policy in order to minimize the infinite horizon cost function defined as [1]

$$J(x(k)) = \sum_{i=0}^{\infty} r(x(k+i), u(x(k+i))), \tag{2}$$

where $r(x(k), u(x(k))) = x(k)^T Q x(k) + u(x(k))^T R u(x(k))$ with $Q \in \Re^{n\times n}$ denotes a positive semi-definite matrix and $R \in \Re^{m\times m}$ is a positive definite matrix. The cost function (2) can be written as $J(x(k)) = r(x(k), u(x(k))) + J(x(k+1)) = r(x(k), u(x(k))) + J(f(x(k)) + g(x(k))u(x(k)))$. The initial control policy is required to be admissible in order to guarantee that the cost function (2) is finite while it stabilizes the system.

Using the Bellman's principle of optimality, it can be shown that the infinite horizon optimal cost function, $J^*(x(k))$ is time invariant and satisfies the discrete-time Hamiltonian-Jacobi-Bellman (HJB) equation. Then, for the optimal cost function one has $J^*(x(k)) = \min_{u(x(k))} (r(x(k), u(x(k)))) + J^*(f(x(k)) + g(x(k))u(x(k)))$. The optimal control $u^*(x(k))$ that minimizes $J^*(x(k))$ is found by applying the stationarity condition $\partial J^*(x(k))/\partial u(x(k)) = 2Ru(x(k)) + g(x(k))^T \partial J^*(x(k+1))/\partial(x(k+1)) = 0$, which yields [9]

$$u^*(x(k)) = -\frac{1}{2} R^{-1} g(x(k))^T \frac{\partial J^*(x(k+1))}{\partial(x(k+1))} \ . \tag{3}$$

The future state vector $x(k+1)$ is required to compute the optimal control (3), which is generally not available. To overcome this problem, online single layer NN-based optimal control is proposed in [9]. In contrast, a multi-layer NN-based optimal control is introduced in this paper. To proceed, the following fact is needed.

Fact. The closed-loop system is bounded above when the optimal control is asserted [9] such that, i.e., $\|f(x(k)) + g(x(k))u^*(x(k))\| \leqslant \overline{k}$ for a known constant $\overline{k}$. An upper bound for the optimal closed-loop system can be established using the Lyapunov theory.

Next, the optimal regulation control of nonlinear discrete-time systems using a MNN is presented.

III. Optimal Regulation of Nonlinear Discrete-Time Systems

In this section, the optimal control for nonlinear discrete-time systems is solved using a MNN. To this end, MNN-based actor and critic networks are used for approximating the cost function and optimal control policy. Due to the additional layer, estimation errors for generating the optimal control policy at every sampling instant is encountered, which needs to be considered in the design and analysis. The boundedness of the cost function and the closed-loop stability with optimal control is ensured by Lyapunov methods with proof of convergence.

The cost function as stated by (2) is approximated using a two layer NN called critic and represented as

$$J(x(k)) = w_c^T \sigma_c(v_c^T x(k)) + \varepsilon_{jk} \tag{4}$$

where v_c, w_c are the first and second layer weights of the critic NN, ε_{jk} is the bounded approximation error, and σ_c is the nonlinear activation function of the critic NN.

The optimal policy (3) is approximated using a two layer NN, called the actor, as

$$u(x(k)) = w_a^T \sigma_a(v_a^T x(k)) + \varepsilon_{uk} \tag{5}$$

where v_a, w_a are the first and second layer weights of the actor NN, ε_{uk} is the bounded approximation error, and σ_a is the nonlinear activation function of the actor NN. Next, the following assumption is stated.

Assumption 1. The NN weights and approximation errors are assumed to be bounded [2] such that $\|w_c\| \leqslant w_{cM}$, $\|v_c\| \leqslant v_{cM}$, $\|w_a\| \leqslant w_{aM}$, $\|v_a\| \leqslant v_{aM}$, $|\varepsilon_{jk}| \leqslant \varepsilon_{jM}$, $|\varepsilon_{uk}| \leqslant \varepsilon_{uM}$ where $w_{cM}, v_{cM}, w_{aM}, v_{aM}, \varepsilon_{jM}, \varepsilon_{uM}$ are positive constants. In addition, the gradient of the approximation error is assumed to be bounded from above as $\|\partial\varepsilon_{jK}/\partial x(k+1)\|_F \leqslant \varepsilon_{jM}^{'}$ where $\varepsilon_{jM}^{'}$ is also a positive constant [9].

A. Cost Function Approximation

In this subsection, the cost function is approximated by a two layer NN. The weights update laws are derived using TD error and, finally, the boundedness of the cost function is proven.

To proceed, let the cost function be estimated using a two layer critic NN as

$$\hat{J}_k(x(k)) = \hat{w}_c^T \sigma_c(\hat{v}_c^T(k) x(k)) \tag{6}$$

where $\hat{w}_c^T$ and $\hat{v}_c^T$ are the estimated NN target weights. Using the delayed values of system state vector and current values of the weights, the cost function (2) can be written as $\hat{J}_k(x(k-1)) = r(x(k-1), u(x(k-1))) + \hat{J}_k(x(k))$. Further using

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

(6), one has $\hat{J}_k(x(k)) - \hat{J}_k(x(k-1)) + r(x(k-1), u(x(k-1))) = \hat{w}_c^T \sigma_c(\hat{v}_c^T(k)x(k)) - \hat{w}_c^T \sigma_c(\hat{v}_c^T(k)x(k-1)) + r(x(k-1), u(x(k-1)) = e_{jk}$ which leads to

$$e_{jk} = r(x(k-1), u(x(k-1))) + \hat{w}_c^T(k)\Delta\sigma_c(x(k-1)) \quad (7)$$

with $\Delta\sigma_c(x(k-1)) = \sigma_c(\hat{v}_c^T(k)x(k)) - \sigma_c(\hat{v}_c^T(k)x(k-1))$.

Adding and subtracting $w_c^T \sigma_c(\hat{v}_c^T(k)(x(k))$ and $w_c^T \sigma_c(\hat{v}_c^T(k)(x(k-1))$ and, after some simplification, (7) becomes

$$\begin{aligned} e_{jk} = &-\tilde{w}_c^T(k)\sigma_c(\hat{v}_c^T(k)x(k)) \\ &+ w_c^T\left[\tilde{\sigma}_c(k) + \tilde{\sigma}_c(k-1)\right] - \Delta\varepsilon_{jk} \end{aligned} \quad (8)$$

where $\tilde{\sigma}_c(k) = \sigma_c(\hat{v}_c^T(k)(x(k)) - \sigma_c(v_c^T(x(k))$. Substituting $\Delta\sigma_c(x(k-1))$, and $\Pi(k) = \tilde{\sigma}_c(k) + \tilde{\sigma}_c(k-1)$, equation (8) reduces to

$$e_{jk} = -\tilde{w}_c^T(k)\Delta\hat{\sigma}_c(k-1) + w_c^T\Pi(k) - \Delta\varepsilon_{jk} \quad (9)$$

Remark 1. The temporal difference (TD) error e_{jk} in (9) depends on NN weight estimation errors, the activation function outputs from past sampling instants and their accumulated values. It does not depend on the future state of the system, i.e., $x(k+1)$.

In the following theorem, the boundedness of the approximated cost function is demonstrated.

Theorem 1: (Boundedness of the cost function) Let $u_0(x_k)$ be the admissible control policy for the controllable system (1) with the cost function (2). Let the critic NN second layer weight update law be given by

$$\hat{w}_c(k+1) = \hat{w}_c(k) - \frac{\alpha_J \Delta\sigma_c(\hat{v}_c^T(k)x(k))e_{jk}}{\Delta\sigma_C^T(\hat{v}_c^T(k)x(k))\Delta\sigma_c(\hat{v}_c^T(k)x(k)) + 1} \quad (10)$$

with the first layer weights tuned by

$$\hat{v}_c(k+1) = \hat{v}_c(k) + x(k)(\hat{v}_c^T(k)x(k) + B_1 k_v e_{jk})^T \quad (11)$$

with B_1 being a known design matrix of appropriate dimension. There exists a positive constant α_J such that the critic NN weights estimation error is uniformly ultimately bounded (UUB), with ultimate bound given by $\|\tilde{w}_c\| \leq b'_{w_c}$ and $\|\tilde{v}_c\| \leq b'_{v_c}$ for a small positive constant b'_{w_c} and b'_{v_c}, respectively.

Proof: Consider the Lyapunov candidate function as

$$V_J(\tilde{w}_c, \tilde{v}_c) = tr\{\tilde{w}_c^T(k)\tilde{w}_c(k)\} + tr\{\tilde{v}_c^T(k)\tilde{v}_c(k)\} \quad (12)$$

The proof of boundedness can be done by taking the first difference of (12) and showing that under some bound conditions $\Delta V_J < 0$. The detailed proof is omitted due to the space limitation. ■

B. Estimated Optimal Control Policy

In this subsection, a NN-based adaptive optimal control is presented using the temporal difference error e_{jk} and control input error. The optimal control input is generated by an actor NN that minimizes the cost function. A Two layer NN is considered for the actor to generate optimal control based on the cost function approximated by the critic NN.

To this end, let the optimal control input be approximated by a two layer NN as

$$\hat{u}(x(k)) = \hat{w}_a^T \sigma_a(\hat{v}_a^T(k)x(k)) \quad (13)$$

where $\hat{v}_a$ and $\hat{w}_a$ are the estimated values of actor NN weights of the first and second layer, respectively, and σ_a is the nonlinear activation function chosen for the hidden layer neurons. Using the optimal control policy (3) the control input error is defined as

$$\begin{aligned} \tilde{u}(k) = &\hat{w}_a^T(k)\sigma_a\left(\hat{v}_a^T(k)x(k)\right) \\ &+ \frac{1}{2}R^{-1}g^T(x(k))\frac{\partial\sigma_c\left(\hat{v}_c^T(k)x(k+1)\right)^T}{\partial x(k+1)}\hat{w}_c(k) \end{aligned}. \quad (14)$$

Adding and subtracting $w^T{}_a\sigma_a\left(\hat{v}_a^T(k)x(k)\right)$ and, after some simplifications (14), one has

$$\begin{aligned} \tilde{u}(k) = &-\tilde{w}_a^T(k)\sigma_a(k) - w_a^T(k)\tilde{\sigma}_a(k) \\ &- \frac{1}{2}R^{-1}g^T(x(k))\frac{\partial\sigma_c\left(\hat{v}_c^T(k)x(k)\right)^T}{\partial x(k+1)}\tilde{w}_c(k) \\ &- \frac{1}{2}R^{-1}g^T(x_k)\frac{\partial\tilde{\sigma}_c(k+1)}{\partial x(k+1)}w_c(k) - \tilde{\varepsilon}_{uk} \end{aligned} \quad (15)$$

where $\sigma_a(k) = \sigma_a\left(\hat{v}_a^T(k)x(k)\right)$, $\tilde{\sigma}_a(k) = \sigma_a\left(v_a^T(k)x(k)\right) - \sigma_a\left(\hat{v}_a^T(k)x(k)\right)$ and $\tilde{\varepsilon}_{uk} = \varepsilon_{uk} + \frac{1}{2}R^{-1}g^T(x(k))\left(\partial\varepsilon_{jk+1}/\partial x(k+1)\right)$.

The NN weight update law for the action network by employing the control input error (15) is defined as

$$\hat{w}_a(k+1) = \hat{w}_a(k) - \frac{\alpha_u\sigma_a\left(\hat{v}_a^T(k)x(k)\right)\tilde{u}(k)^T}{\left(\sigma_a{}^T\left(\hat{v}_a^T(k)x(k)\right)\sigma_a\left(\hat{v}_a^T(k)x(k)\right)+1\right)} \quad (16)$$

where $0 < \alpha_u < 1$ is a positive design parameter. The weight update law for the first layer of control policy is given by

$$\hat{v}_a(k+1) = \hat{v}_a(k) + x(k)(\hat{v}_a^T(k)x(k) + B_2 k_v\tilde{u}(k))^T \ , \quad (17)$$

where B_2 is a design matrix of appropriate dimension and k_v is a scaling factor.

The weight estimation error dynamics is given by

$$\tilde{w}_a(k+1) = \tilde{w}_a(k) + \frac{\alpha_u\sigma_a\left(\hat{v}_a^T(k)x(k)\right)\tilde{u}(k)^T}{\sigma_a^T\left(\hat{v}_a^T(k)x(k)\right)\sigma_a\left(\hat{v}_a^T(k)x(k)\right)+1} \quad (18)$$

The closed-loop nonlinear system dynamics can be written in terms of $u^*(k)$ and the $\tilde{w}_a$ and $\tilde{v}_a$ as

$$\begin{aligned} &x(k+1) = f(x(k)) + g(x(k))\hat{u}(x(k)) \\ &= f(x(k)) + g(x(k))\hat{w}_a^T(k)\sigma_a\left(\hat{v}_a^T(k)x(k)\right) \\ &= f(x(k)) + g(x(k))\hat{w}_a^T(k)\sigma_a\left(\hat{v}_a^T(k)x(k)\right) \\ &-g(x(k))[w_a^T(k)\sigma_a\left(v_a^T(k)x(k)\right) + \varepsilon_{uk}] \\ &+g(x(k))[w_a^T(k)\sigma_a\left(v_a^T(k)x(k)\right) + \varepsilon_{uk}] \\ &= f(x(k)) + g(x(k))u*(x(k)) \\ &-g(x(k))\tilde{w}_a^T(k)\sigma_a\left(\hat{v}_a^T(k)x(k)\right) - g(x(k))\varepsilon_{uk} \end{aligned}$$

In the following theorem, the boundedness of the overall closed-loop system states and parameters are provided.

Theorem 2: (Boundedness of the the optimal control) Let $u_0(x_k)$ be the initial admissible control policy for the controllable system (1) with cost function (2). The cost and control NN weights update laws for the two layers are given by (10), (11), (16) and (17).There exist a positive constant α_J, α_U and positive constants such that the system states $x(k)$, cost and action network NN weight estimation errors $\tilde{w}_c$,$\tilde{v}_c$,$\tilde{w}_a$, and $\tilde{v}_a$ are all UUB for all k with ultimate bounds given by $\|\tilde{w}_c\| \leq b'_{w_c}$,$\|\tilde{v}_c\| \leq b'_{v_c}$, $\|\tilde{w}_a\| \leq b'_{w_a}$, $\|\tilde{v}_a\| \leq b'_{v_a}$ for small positive constants b'_{w_c}, b'_{v_c}, b'_{w_a} and b'_{v_a}.

Proof: Consider the Lyapunov candidate function as

$$V = V_D(x(k)) + V_U(\tilde{w}_a(k)) + V_U(\tilde{v}_a(k)) + V_J(\tilde{w}_c(k)) + V_J(\tilde{v}_c(k)) \tag{19}$$

where $V_D(x) = x(k)^T x(k)$, $V_U(\tilde{w}_a(k)) = tr\{\tilde{w}_a^T(k)\tilde{w}_a(k)\}$, $V_U(\tilde{v}_a(k)) = tr\{\tilde{v}_a^T(k)\tilde{v}_a(k)\} V_J(\tilde{w}_c(k)) = tr\{\tilde{w}_c^T(k)\tilde{w}_c(k)\}$ and $V_J(\tilde{v}_c(k)) = tr\{\tilde{v}_c^T(k)\tilde{v}_c(k)\}$. Taking the first derivative of (19), one can show that under certain conditions $\Delta V < 0$ which shows the boundedness of the overall closed-loop system states and parameters. The rest of the proof is omitted. ■

After demonstrating that the closed-loop system will be bounded in the presence of proposed optimal control policy using a MNN, next we show that the proposed learning approach, if extended to more number of hidden layers, does not result in the vanishing gradient problem [17] that is found with standard backpropagation-based gradient learning scheme.

Theorem 3: Consider the nonlinear discrete-time system (1) along with the infinite horizon cost function (2). Let the control policy (13) be utilized with the weight update law (10), (11), (16) and (17). Then, as the number of layers increases, the vanishing gradient does not happen.

Proof: In the proposed scheme, the TD error for critic network and control input error for actor network are employed directly at each layer in the the weight update law, instead of propagating the errors through the hidden layers in the case of Backpropagation. Consequently, the vanishing gradient problem does not occur in the proposed approach. ■

In the next section, simulation results of the proposed approach are presented.

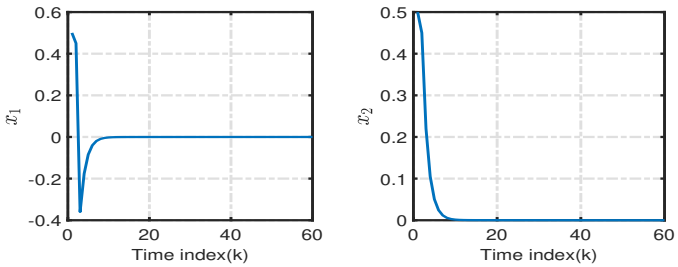
IV. Simulation Results

In this section, two examples are provided for both linear and nonlinear discrete-time systems to show the effectiveness of the proposed optimal control approach.

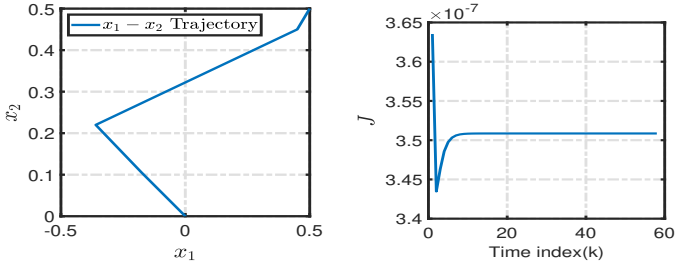
A. Linear system

Consider the linear system stated in [9] whose dynamics are given by

$$x_{k+1} = \begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} = \begin{bmatrix} 0 & -0.8 \\ 0.8 & 1.8 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u(x_k) \tag{20}$$



(a) States x_1 and x_2



(b) State Trajectory (x_1 vs x_2) and cost function J

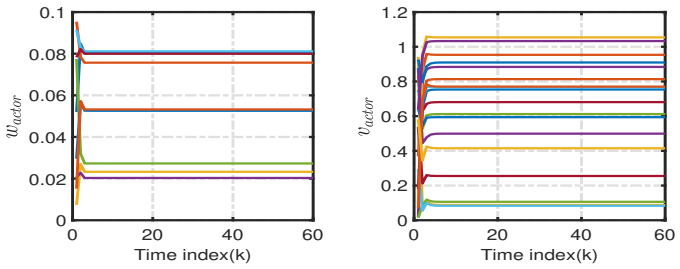
Fig. 1: Performance of the proposed MNN based optimal control for the linear discrete-time system given by (20) from [9].

The initial stabilizing policy was selected as $u(x_0) = \begin{bmatrix} 0.5 & 1.4 \end{bmatrix} x_0$ and the initial values of states are considered as $x_0 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^T$.

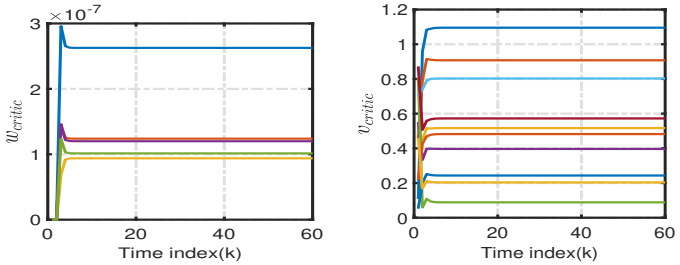
The architecture of the critic MNN chosen in this work for the linear system has 2 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. The actor MNN has 2 neurons in the input layer, 9 neurons in the hidden layer and 1 neuron in the output layer with $a_J = 10^{-6}$ and $a_u = 0.1$. The nonlinear hyperbolic tangent activation function is used for the hidden layer neurons, and linear activation function is used for the output layer. The initial weights of the two layer critic NN are set to zero and the initial weights of first layer of actor NN are set at random in the range of $[-1, 1]$. The initial weights of second layer of the actor NN are set to $[-0.1, 0.1]$. The design matrices Q and R of the cost function are chosen as identity matrix and 1, respectively.

Fig. 1 shows that the control input generated by the proposed MNN converges within a couple of iterations from different initial conditions on the state vector, further demonstrating the effectiveness of the new MNN weight update laws. To illustrate the improved performance of the proposed MNN based optimal regulator, the single layer NN based optimal regulator reported in [9] is simulated. The basis functions for the single layer NN based critic network are constructed using the fourth-order polynomial given by $\begin{bmatrix} x_1^2, & x_1x_2, & x_2^2, & x_1^4, & x_1^3x_2, & x_2^3x_1, & x_1^2x_2^2, & x_2^4 \end{bmatrix}$ and the single layer NN based action network basis functions are constructed using the gradient of the fourth order polynomial. The basis functions are constructed using the expansion of the polynomial from [18]. Hence, the critic network is a single layer NN with 8 neurons in the input layer that receives the 8

978-1-7281-6926-2/20/$31.00 ©2020 IEEE



(a) Actor weight changes



(b) Critic weight changes

Fig. 2: Actor and critic weight variations for the linear discrete-time system (20).

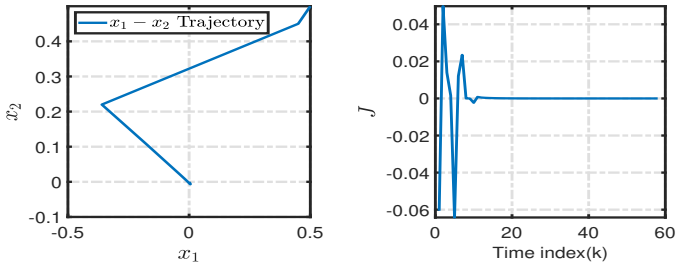


Fig. 3: Performance of the single layer NN based optimal control for the linear discrete-time system given by (20) .

polynomial functions and one neuron in the output layer. The actor network is also a single layer NN with 16 inputs which are the derivations of the polynomial functions. The initial weights of critic are chosen as zero and actor are set at random in the range of $[-1, 1]$. The design parameters of the critic and actor are $a_J = 10^{-6}$ and $a_u = 0.1$. The simulation is run for 60 time steps and reported in Fig. 3. It is observed that the performance depends on the proper selection of polynomial function which will be challenging for complex system.

It is observed from Figs. 1 , 2, 3 and 4 that the MNN-based optimal adaptive control converges within a few iterations and generates an optimal control input when compared with single layer NN based optimal control. However, the performance depends on the selection of initial weights of the neural network for admissible control. Since this is a linear system example, a few neurons in the critic and actor NNs appear to be sufficient. The convergence of the functional approximation error also appears to be faster.

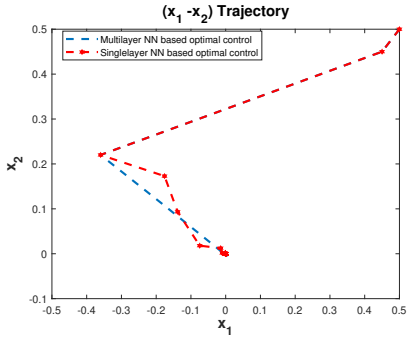


Fig. 4: Performance comparison of the proposed MNN with single layer NN based optimal control for (20)

B. Nonlinear system

Consider the nonlinear example reported in [4] given by

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} -\sin(0.5x_2(k)) \\ -\cos(1.4x_2(k))\sin(0.9x_1(k)) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k) \tag{21}$$

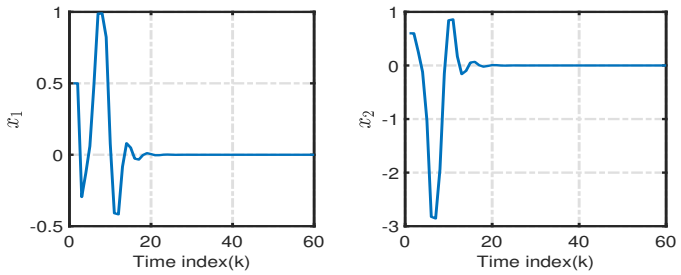
The initial state is taken as $x_0 = \begin{bmatrix} 0.5 & 0.6 \end{bmatrix}^T$. The architecture of actor and critic MNN chosen for the nonlinear system has 2 neurons in the input layer, 8 neurons in the hidden layer and 1 neuron in the output layer with $\alpha_J = 0.8$ and $\alpha_u = 0.3$. The nonlinear hyperbolic tangent activation functions are used for the hidden layer neurons and linear activation function are used for the output layer. The initial weights of the two layers of critic NN are set to zero and the initial weights of first layer of actor NN are randomly set to values in the range of $[-1, 1]$.

It is observed from Fig. 5 and 6 that the adaptive NN approach with the proposed weight update law generates the optimal control which leads to the convergence of state vector and NN weights within few iterations. The simulations are run with different initial conditions. The MNN based critic and actor networks learn the nonlinearity associated with the cost function and control action online without issues related to basis function selection.

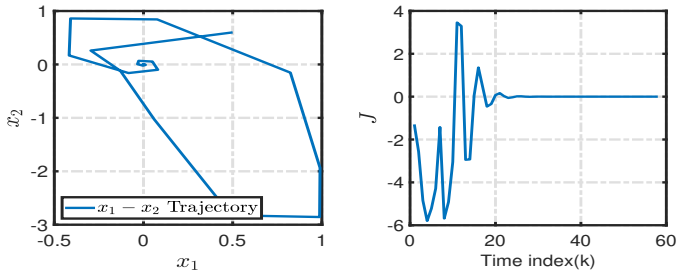
V. Conclusion

In this paper, the problem of infinite horizon optimal regulation of nonlinear discrete-time systems with uncertain internal dynamics is addressed using two layer NNs. It has been demonstrated that additional layers enhance approximation error and result in better regulation provided with the weight tuning laws developed. For the case of optimal regulation, TD error is utilized for tuning the weights and it appears to result in acceptable performance. Future work can be focused on the design of the controller with both uncertain input matrix and internal dynamics with implementation to some practical applications. Moreover, efficient tracking control of complex systems can also be aimed at using appropriate augmented systems.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

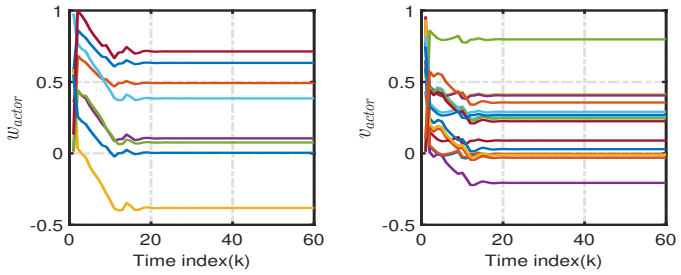


(a) States x_1 and x_2

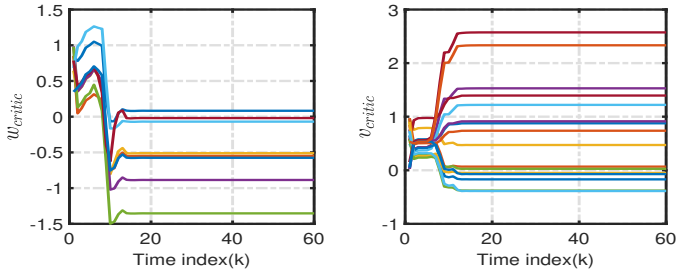


(b) State Trajectory (x_1 vs x_2) and cost function J

Fig. 5: Performance of the proposed MNN based optimal control for the Nonlinear discrete-time system (21).



(a) Actor weight changes



(b) Critic weight changes

Fig. 6: Actor and critic weight variations for the nonlinear discrete-time system (21).

REFERENCES

[1] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. John Wiley & Sons, 2012.

[2] J. Sarangapani, *Neural network control of nonlinear discrete-time systems*. CRC press, 2006.

[3] H. Zargarzadeh, Q. Yang, and S. Jagannathan, "Online optimal control of nonaffine nonlinear discrete-time systems without using value and policy iterations," *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, pp. 221–257, 2013.

[4] Y. Geyang Xiao, Huaguang Zhang and Luo, "Online optimal control of unknown discrete-time nonlinear systems by using time-based adaptive dynamic programming," *Neurocomputing*, vol. 165, pp. 163–170, 2015.

[5] B. Kiumarsi, F. L. Lewis, and D. S. Levine, "Optimal control of nonlinear discrete time-varying systems using a new neural network approximation structure," *Neurocomputing*, vol. 156, pp. 157–165, 2015.

[6] E. N. Sanchez and F. Ornelas-Tellez, *Discrete-time inverse optimal control for nonlinear systems*. CRC Press, 2017.

[7] F. Al-Tamimi and M. Abu-Khalaf, "Discrete-time nonlinear hjb solution using approximate dynamic programming: Convergence proof," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 943–949, 2008.

[8] G. Xiao, H. Zhang, and Y. Luo, "Online optimal control of unknown discrete-time nonlinear systems by using time-based adaptive dynamic programming," *Neurocomputing*, vol. 165, pp. 163–170, 2015.

[9] T. Dierks and S. Jagannathan, "Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1118–1129, 2012.

[10] Q. Zhang Huaguang,, Ruizhuo Song, and Tieyan Zhang, "Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming," *IEEE Transactions on Neural networks*, vol. 22, no. 12, pp. 1851–1862, 2011.

[11] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on neural networks*, vol. 1, no. 1, pp. 4–27, 1990.

[12] R. Krishnan, V. Samaranayake, and S. Jagannathan, "A multi-step nonlinear dimension-reduction approach with applications to bigdata," *Procedia computer science*, vol. 144, pp. 81–88, 2018.

[13] B. Hanin, "Which neural net architectures give rise to exploding and vanishing gradients?," in *Advances in Neural Information Processing Systems*, pp. 582–591, 2018.

[14] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[15] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, 2019.

[16] Y. Huang and D. Liu, "Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative adp algorithm," *Neurocomputing*, vol. 125, pp. 46–56, 2014.

[17] B. Hanin, "Which neural net architectures give rise to exploding and vanishing gradients?," in *Advances in Neural Information Processing Systems*, pp. 582–591, 2018.

[18] R. Beard, "Improving the closed-loop performance of nonlinear systems," *Rensselaer Polytechnic Institute*, 1995.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE