2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Constrained Controller Design for a Class of Nonlinear Discrete-time Uncertain Systems

Dipak M. Adhyaru, I. N. Kar, Member, IEEE and M. Gopal

Abstract- **In this paper, constrained controller design is proposed for a class of nonlinear discrete-time uncertain system having matched type system uncertainties, using the solution of HJB (Hamilton-Jacobi-Bellman) equation. The discrete-time HJB equation is formulated using a suitable non-quadratic term in the performance functional to tackle constraints on the control input. Based on the non-quadratic functional, a greedy HDP algorithm is used to obtain the constrained robust-optimal controller. The constrained robust controller requires knowledge of the upper bound of system uncertainties. For facilitating the implementation of the iterative algorithm, two neural networks are used to approximate the value function and to compute the optimal control policy, respectively. Their weights have been tuned using least squares method. Proposed algorithm has been applied on a nonlinear discrete-time system with matched uncertainties.**

I. Introduction

In recent years, approximate dynamic programming (ADP) algorithm has been paid much attention by researchers [1–5] in order to obtain approximate optimal control law. ADP combines adaptive critic design and reinforcement learning technique with dynamic programming. ADP approaches were mostly classified into four main schemes: Heuristic Dynamic Programming (HDP), Dual Heuristic Dynamic Programming (DHP), Action Dependent Heuristic Dynamic Programming (ADHDP), also known as Q-learning, and Action Dependent Dual Heuristic Dynamic Programming (ADDHP). Liu et. al [15] proposed action dependent adaptive critic designs for nonlinear systems. Si and Wang [6] proposed two new ADP schemes known as Globalized-DHP (GDHP) and ADGDHP. In [8],[10] a greedy HDP iteration scheme was proposed to solve the optimal control problem for nonlinear discrete-time systems with known mathematical model. In all the above mentioned algorithms, constraints on the control input were not taken into account. However, in practical systems one should consider at least magnitude constraints on control input due to limitations of the actuators. Lysevski [14] suggested a constrained controller design for nonlinear discrete-time system. Khalaf et.al. [7] proposed similar approach for continuous-time nonlinear system with HJB equation based formulation. However, constrained optimal controller design for discrete-time system has not been much explored in the literature.

All the HJB based optimal controller designs need exact information about the system model. However, all practical control systems have to be robust with respect to model uncertainty such as unknown or partially known time-varying process parameters, exogenous disturbances etc. So model uncertainty needs to be considered during the time of controller design process to avoid the deterioration of nominal closed-loop performance. In other words, we need to design robust feedback control law to tackle the system uncertainty. Berger [12] proposed robust control of linear discrete-time systems by minimizing the frequency response of the perturbed plant at selected frequencies. Ho and Lu [13] proposed a method for robust stabilization of discrete nonlinear system with LMI (Linear Matrix Inequality) approach. Even though the above mentioned results provide systematic methods for robust controller design, they do not, in general, lead to controllers that are optimal with respect to a meaningful cost. An alternative realistic approach is to obtain approximate solution of HJB equation using NN, for feedback controller design of discrete-time nonlinear uncertain system having constrained input. However, constrained robust controller design using NN-based HJB solution has not been explored much in the literature.

In this paper, the main methodological contribution is the design of robust controller for nonlinear discrete-time uncertain system. In this approach, the robust control problem is formulated into an optimal control problem by properly choosing a cost functional. The cost functional is modified to account for matched system uncertainties and constraints on the input. Hence it can be referred to as a constrained robust-optimal control design approach. The proposed work is realized using iterative HDP algorithm, with necessary theoretical justifications. It is implemented using two neural networks to approximate the solution of discrete-HJB equation. The least squares method has been used to find tuning law of neural networks. Convergence proof of the present work is supplemented by necessary theoretical and simulation results.

The paper is organized as follows: In section 2, HJB based robust-optimal control framework has been developed to

Dipak M. Adhyaru is with the Department of Electrical Engineering, Indian Institute of Technology Delhi,Hauz Khas, New Delhi-110016, INDIA email: hetdip@rediffmail.com.

I.N. Kar is with the Department of Electrical Engineering, Indian Institute of Technology Delhi,Hauz Khas, New Delhi-110016, INDIA email: ink@ee.iitd.ac.in

M. Gopal is with the Department of Electrical Engineering, Indian Institute of Technology Delhi,Hauz Khas, New Delhi-110016, INDIA email: mgopal@ee.iitd.ac.in

design constrained robust controller for discrete-time matched uncertain systems. In the section 3, we propose HDP algorithm for solving constrained robust control problem. Implementation of HDP algorithm with neural networks is discussed in section 4. Proposed approach has been validated by simulating experiment on a nonlinear discrete-time uncertain system having matched type system uncertainties. Conclusions follow in section 6.

II. ROBUST–OPTIMAL CONTROL FRAMEWORK

Consider a nonlinear discrete-time system

$$x(k+1)=\bar{A}(x(k))+B(x(k))u(k)$$

where $x(k)\in\mathbb{R}^n$ is the state vector and $u(k)\in\mathbb{R}$ is the control input. It is assumed that $u(k)$ is bounded by a positive constant λ. i.e., $|u(k)|\le\lambda\in\mathbb{R}$ (1)

Suppose that the function $\bar{A}(x(k))$ is known only up to an additive perturbation which is bounded by a known function, and this perturbation is in the range of $B(x(k))$, i.e., $\bar{A}(x(k))$ can be written as $\bar{A}(x(k))=A(x(k))+B(x(k))f(x(k))$ with unknown $f(x(k))$. The condition that unknown perturbation be in the range space of $B(x(k))$ is called the matching condition, and can be incorporated by expressing the system as

$$x(k+1)=A(x(k))+B(x(k))u(k)+B(x(k))f(x(k)) \quad (2)$$

The function $B(x(k))f(x(k))$ models matched uncertainty in the system dynamics. The nominal model $A(x(k))$ and $B(x(k))$ are known with $A(0)=0$ and $f(0)=0$. This assumption ensures that origin is the equilibrium point of the system (2). It is assumed that the function $f(x(k))$ is bounded by a known function, $f_{\max}(x(k))$:

$$\|f(x(k))\|\le f_{\max}(x(k)) \quad \forall k. \quad (3)$$

In this section, design of a constrained optimal control law is proposed to ensure the asymptotic stability of the system (2).

(a) Robust Control Problem: For the open-loop system (2), find a feedback control law $u(k)$ such that the closed-loop system is globally asymptotically stable for all admissible uncertainties $f(x(k))$.

This problem can be formulated into an optimal control of the nominal system with appropriate cost functional.

(b) Optimal control problem:

For the nominal system

$$x(k+1)=A(x(k))+B(x(k))u(k) \quad (4)$$

find a feedback control $u(k)$ that minimizes the cost functional

$$J(x(k),u)=\sum_{i=k}^{\infty}\left(\rho f_{\max}^2(x(k))+x(i)^TQx(i)+M(u(i))\right) \quad (5)$$

where $M(u(i))=2\int_0^{u(i)}\left(\lambda\varphi^{-1}(\lambda^{-1}v)\right)Rdv$

$$=2\lambda Ru\varphi^{-1}(\lambda^{-1}u)+\lambda^2R\ln(1-(\lambda^{-1}u)^2)>0 \quad (6)$$

is non-quadratic term expressing cost related to constrained input. Q is positive definite matrix of appropriate dimensions, and it is assumed to be diagonal for simplicity of analysis. R is a positive constant. ρ is a positive constant and it is used as a design parameter. $\varphi(\cdot)$ is a bounded one-to-one function satisfying $|\varphi(\cdot)|\le 1$ and belonging to $C^p(p\ge 1)$ and $L_2(\Omega)$. Moreover, it is a monotonic increasing odd function with its first derivative bounded by a constant N. In this paper, we have considered $\varphi(\cdot)=\tanh(\cdot)$.

Assume that $A(x(k))+B(x(k))u(k)$ is lipschitz continuous on set Ω in $\mathbb{R}^n$ containing origin, and that the system (4) is controllable in the sense that there exists a continuous control on Ω that asymptotically stabilizes the system. For optimal control problem, state feedback control $u(k)$ must not only stabilize the system on Ω but also guarantee that (5) is finite, i.e., admissible control. From now on, we let $V^*(x(k))$ denote the minimum value of the performance functional $J(x(k),u)$, which is called value function or optimal cost function in the later parts.

Definition 1. (Admissible Control): A control $u(k)$ is defined to be admissible control with respect to (5) on Ω if u stabilizes (4) on Ω, $u(0)=0$, and for all $x(0)\in\Omega$, $J(x(0),u)$ is finite.

In this paper, we address the following problems:

(i) Solution of the robust control problem (a) and optimal control problem (b) are equivalent.

(ii) Solve the optimal control problem using HDP based iterative algorithm and implementation of it through NN.

According to Bellman optimality principle, we can obtain

$$V^*(x(k))=\min_{u(i)}\sum_{i=k}^{\infty}\left(\rho f_{\max}^2(x(i))+x(i)^TQx(i)+M(u(i))\right)$$
$$=\min_{u(i)}\left(\rho f_{\max}^2(x(k))+x(k)^TQx(k)+M(u(k))+V^*(x(k+1))\right) \quad (7)$$

According to the first-order necessary condition of the optimal control, the following equation holds:

$$\frac{\partial V^*(x(k))}{\partial u(k)}=2\lambda R\tanh^{-1}(\lambda^{-1}u(k))+\left(\frac{\partial x(k+1)}{\partial u(k)}\right)^T\frac{\partial V^*(x(k+1))}{\partial x(k+1)}=0 \quad (8)$$

It gives

$$u^*(k)=-\lambda\tanh\left(\frac{1}{2}(\lambda R)^{-1}B^T(x(k))V_x^*(x(k+1))\right) \quad (9)$$

where $V_x^*(x(k+1))=\dfrac{\partial V^*(x(k+1))}{\partial x(k+1)}$.

The resulting HJB equation is

$$V^*(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u^*(k)) + V^*(x(k+1)) \quad (10)$$

The optimal control $u^*(k)$ can be computed if the value function $V^*(x(k+1))$ can be solved from the HJB equation (7). In the next section we will discuss how to use the approximated dynamic programming algorithm named HDP (Heuristic Dynamic Programming) to solve the optimal control problem.

III. HDP ALGORITHM FOR CONSTRAINED ROBUST CONTROLLER DESIGN

In this section, we propose an iterative HDP algorithm to solve HJB equation having modified performance term related to constraint on the input and bound on the system uncertainties. Initially we find solution of HJB equation using HDP algorithm, which gives constrained optimal control and then we will prove that this control is the robust control for the system (2).

A. HDP algorithm for constrained optimal control problem

Let the initial cost function is $V_0(x(k)) = 0$. Now we can find the control $u_0(x(k))$ as follows:

$$u_0(x(k)) = \arg\min_{u(k)} \left(\rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u(k)) + V_0(x(k+1)) \right) \quad (11)$$

and then update the cost function as

$$V_1(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u_0(k)) + V_0(x(k+1)) \quad (12)$$

The HDP algorithm iterates between

$$\begin{aligned} u_i(x(k)) &= \arg\min_{u(k)} \left(\rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u(k)) + V_i(x(k+1)) \right) \\ &= -\lambda \tanh\left(\frac{1}{2}(\lambda R)^{-1} B^T(x(k)) \frac{\partial V_i(x(k+1))}{\partial x(k+1)} \right) \end{aligned} \quad (13)$$

and,

$$\begin{aligned} V_{i+1}(x(k)) &= \min_{u(k)} \left(\rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u(k)) + V_i(A(x(k)) + B(x(k))u(k)) \right) \\ &= \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u_i(k)) + V_i(A(x(k)) + B(x(k))u_i(k)) \end{aligned} \quad (14)$$

In this way, the cost function and control policy are updated by recurrent iteration until they converge to the optimal ones, with the iteration number i increasing from 0 to ∞. Asma et. al. [8],[10] presented theoretical results for unconstrained optimal control design using HDP based algorithm. In the following part, we state theoretical results to prove the convergence of the iteration between (13) and (14) with the cost function $V_i \to V^*$ and $u_i \to u^*$ as $i \to \infty$ with modified performance terms.

Lemma 1: Let μ_i be any arbitrary sequence of control policies and u_i be the policies as (13). Let V_i be as (14) and Λ_i as

$$\Lambda_{i+1}(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(\mu_i(k)) + \Lambda_i(x(k+1)) \quad (15)$$

If $V_0 = A_0 = 0$, then $V_i \le \Lambda_i, \forall i$.

Proof: Note that, V_{i+1} is a result of minimizing the right-hand side of (14) with respect to the control input u, while Λ_{i+1} is a result of any arbitrary control input. Since $V_0 = A_0 = 0$, it follows that $V_i \le \Lambda_i \ \forall i$, by induction. □

Lemma 2: Let the sequence $\{V_i\}$ be defined as (14). If the system is controllable, then following conditions hold:

(i) There exists an upper bound Y such that $0 \le V_i \le Y \ \ \forall i$.

(ii) If the optimal control problem (7) is solvable, then there exists a least upper bound $V^* \le Y$, where V^* solves (10) and that $0 \le V_i \le V^* \le Y \ \ \forall i$.

Proof: Let $\eta(x(k))$ be any stabilizing and admissible control input, and let $V_0(\cdot) = Z_0(\cdot) = 0$, where V_i is updated as (14) and Z_i is updated by

$$Z_{i+1}(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(\eta_i(k)) + Z_i(x(k+1)) \quad (16)$$

Using results of Lemma 1, one can prove similar to the lemma 2 of [10].

$$V_{i+1}(x(k)) \le Z_{i+1}(x(k)) \le Y, \ \ \forall i$$

This completes the proof of part 1. Moreover if $\eta(x(k)) = u^*(x(k))$, then

$$\sum_{j=0}^{\infty} \left(\rho f_{\max}^2(x(k+j)) + x(k+j)^T Qx(k+j) + M(u^*(x(k+j))) \right)$$
$$\le \sum_{j=0}^{\infty} \left(\rho f_{\max}^2(x(k+j)) + x(k+j)^T Qx(k+j) + M(\eta(x(k+j))) \right)$$

and hence, $V^* \le Y$, which proves part (2) and shows that $0 \le V_i \le V^* \le Y \ \ \forall i$ and for any Y determined by an admissible stabilizing policy $\eta(x(k))$. □

Using results of lemma 1 and lemma 2, in the following theorem it has been shown that the HDP algorithm converges to the value function of the DT HJB equation.

Theorem 1: Define the sequence $\{V_i\}$ as (14), with $V_0(\cdot) = 0$. Then $\{V_i\}$ is a non-decreasing sequence satisfying $V_{i+1}(x(k)) \ge V_i(x(k)), \ \forall i$ and converging to the value function of the discrete-time HJB equation (10), i.e, $V_i \to V^*$ as $i \to \infty$. Meanwhile, the control policy also converges to the optimal policy (9), i.e., $u_i \to u^*$ as $i \to \infty$.

Proof: For the convenience of analysis, define a new sequence ϕ_i as follows:

$$\phi_{i+1}(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u_{i+1}(k)) + \phi_i(x(k+1)) \quad (17)$$

with $\phi_0 = V_0 = 0$ and policies u_i defined as (13), the cost function V_i is updated by (14).

In the following, we prove $\phi_i(x(k)) \le V_{i+1}(x(k))$ by mathematical induction.

First, we prove that it holds for $i=0$. Noticing that

$$V_1(x(k)) - \phi_0(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u_0(k)) \ge 0 \quad (18)$$

Thus for $i=0$ we get,

$$V_1(x(k)) \ge \phi_0(x(k))$$

Second, we assume that it holds for $i-1$, i.e., $V_i(x(k)) \ge \phi_{i-1}(x(k)), \forall x(k)$.

Then for *i* since,

$$V_{i+1}(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u_i(k)) + V_i(x(k+1))$$

holds, we obtain

$$V_{i+1}(x(k)) - \phi_i(x(k)) = V_i(x(k+1)) - \phi_{i-1}(x(k+1)) \ge 0$$

i.e. following equation holds

$$\phi_i(x(k)) \le V_{i+1}(x(k))$$

Furthermore, from lemma 1 we know that $V_i(x(k)) \le \phi_i(x(k))$, therefore we have

$$V_i(x(k)) \le \phi_i(x(k)) \le V_{i+1}(x(k))$$

From part (1) in lemma 2 and the fact that V_i is non-decreasing sequence, it follows that $V_i \to V_\infty$ as $i \to \infty$.

From part (2) of lemma 2, it also follows that $V_\infty \le V^*$. It now remains to show that, in fact, V_∞ is V^*. To see this, note that, from (14), it follows that

$$V_\infty(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u_\infty(k)) + V_\infty(A(x(k)) + B(x(k))u_\infty(k)))$$

and hence,

$$V_\infty(A(x(k)) + B(x(k))u_\infty(k))) - V_\infty(x(k)) = -\rho f_{\max}^2(x(k)) - x(k)^T Qx(k) - M(u_\infty(k))$$

Therefore V_∞ is a Lyapunov function for a stabilizing and admissible policy $u_\infty(k) = \eta(k)$. By using part (2) of lemma 2, it follows that $V_\infty = Y \ge V^*$. This implies that $V^* \le V_\infty \le V^*$ and, hence $V_\infty = V^*$ and $u_\infty = u^*$. □

We have proven that the HDP algorithm converges to the value function of (10). In the next section we will prove that optimal control law design using (9) is the solution of the robust control problem discussed in section 2.

B. HDP algorithm for constrained robust control problem

For the nominal system (2), it has been shown in theorem 1 that if (13) and (14) are exactly solved then $V_\infty = V^*$ and $u_\infty = u^*$. Hence equation (14) can be written as

$$V^*(x(k)) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(u^*(k)) + V^*(A(x(k)) + B(x(k))u^*(k))) \quad (19)$$

It gives,

$$\Delta V^* = V^*(x(k+1)) - V^*(x(k)) = -\rho f_{\max}^2(x(k)) - x(k)^T Qx(k) - M(u^*(k)) \quad (20)$$

It means V^* is a Lyapunov function and u^*, the solution of constrained optimal control problem, is stabilizing and admissible. It is clear from the equation (20) that $x(k)$ has been derived from the nominal dynamics.

Now one can select the design parameter ρ such that

$$\Delta V^{un} = \left(V(A(x(k)) + B(x(k))u^*(k)) + B(x(k))f(k))) - V(x(k))\right) \le 0$$

and an optimal control law u^* will stabilize the uncertain system dynamics (2). It was shown with simulation experiment in section V. However, in that experiment ρ is selected using a trial and error method. Analytical proof related to the conditions for the selection of ρ can be considered as the future work.

To get an optimal control law it is required that the action and value update equations (13) and (14) can be exactly solved, at each iteration, which is a difficult problem for nonlinear system Therefore, for implementation purposes, one need to approximate u_i and V_i at each iteration, which gives the approximate solutions of (13) and (14). In the next section NN (Neural Network) based approximation has been used to solve equations (13) and (14).

IV. NN APPROXIMATION FOR HDP ALGORITHM

In this section, implementation of HDP algorithm using NN has been discussed. The important point stressed is that the use of two NNs, a critic for value function approximation and an action NN for the control, allows the implementation of HDP without knowing the system matrix *A(x)*. It is well known that NNs can be used to approximate smooth functions on prescribed sets [9]. Therefore, to solve (13) and (14), V_i is approximated at each step by a critic NN

$$\hat{V}_i(x) = \sum_{j=1}^{L} w_{vi}^j \Phi_j(x) = W_{Vi}^T \Phi(x) \quad (21)$$

and u_i by an action NN

$$\hat{u}_i(x) = \sum_{j=1}^{P} w_{ui}^j \sigma_j(x) = W_{ui}^T \sigma(x) \quad (22)$$

where the basis functions are $\Phi_j(x), \sigma_j(x) \in C^1(\Omega)$, respectively. Because it is required that $V_i(0)=0$ and $u_i(0)=0$, we select activation functions with $\Phi_j(0)=0$ and $\sigma_j(0)=0$. The NN weights in (21) are w_{vi}^j, L is the number of neurons. The vector $\Phi(x) \equiv [\Phi_1(x)\ \Phi_2(x) \cdots \Phi_L(x)]^T$ is the vector activation function, and $W_{Vi} \equiv [w_{Vi}^1\ w_{Vi}^2 \cdots w_{Vi}^L]^T$ is the weight vector at iteration *i*. Similarly, the weights of the NN in (22) are w_{ui}^j. P is the number of neurons. $\sigma(x) \equiv [\sigma_1(x)\ \sigma_2(x) \cdots \sigma_P(x)]^T$ is the vector activation function, and $W_{ui} \equiv \left[w_{ui}^1 \; w_{ui}^2 \cdots w_{ui}^P \right]^T$ is the weight vector. According to (14), the critic weights are tuned by minimizing the residual error between $\hat{V}_{i+1}(x(k))$ and the target function defined in (24), at each iteration of HDP, in a least squares sense for a set of states $x(k)$ sampled from a compact set $\Omega \subset \mathbb{R}^n$.

$$d(x(k),x(k+1),W_{Vi},W_{ui}) = \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(\hat{u}_i(k)) + \hat{V}_i(x(k+1))$$
$$= \rho f_{\max}^2(x(k)) + x(k)^T Qx(k) + M(\hat{u}_i(k)) + W_{Vi}^T \Phi(x(k+1)) \quad (23)$$

The residual error becomes

$$\left(W_{Vi}^T \Phi(x(k)) - d(x(k),x(k+1),W_{Vi},W_{ui}) \right) = e_L(x) \quad (24)$$

To find the least squares solution, the method of weighted residuals can be used [10]. The weights W_{Vi+1}^T are determined by projecting the residual error onto $\left(de_L(x)/dW_{Vi+1}^T \right)$ and setting the result to zero $\forall x \in \Omega$ using the inner product, i.e.,

$$\left\langle \frac{de_L(x)}{dW_{Vi+1}^T}, e_L(x) \right\rangle = 0 \quad (25)$$

where $\langle f, g \rangle = \int_\Omega f \cdot g^T dx$ is a Lebesgue integral. One has

$$\int_\Omega \Phi(x(k)) \left(\Phi^T(x(k)) W_{Vi+1} - d^T(x(k),x(k+1),W_{Vi},W_{ui}) \right) dx(k) = 0 \quad (26)$$

Therefore, a unique solution for W_{Vi+1} exists and is computed as

$$W_{Vi+1} = \left(\int_\Omega \Phi(x(k)) \Phi^T(x(k)) dx(k) \right)^{-1} \cdot \int_\Omega \Phi(x(k)) d^T(\Phi(x(k)), W_{Vi}, W_{ui}) dx(k) \quad (27)$$

The next assumption is standard in selecting the NN activation functions as a basis set.

Assumption 1: The selected activation functions $\left\{ \Phi_j(x(k)) \right\}^L$ are linearly independent on the compact set $\Omega \in \mathbb{R}^n$.

Assumption 1 guarantees that the excitation condition is satisfied, and hence, $\int_\Omega \Phi(x(k)) \Phi^T(x(k)) dx(k)$ is of full rank and invertible, and a unique solution for (27) exists. The weights of NN (22) are tuned at each iteration. Using $\hat{u}_i(x(k), W_{ui})$ (13) can be rewritten as,

$$W_{ui} = \arg\min_w \left(x(k)^T Qx(k) + M(\hat{u}(x(k),w)) + \hat{V}_i(x^i(k+1)) \right) \Big|_\Omega \quad (28)$$

where $x^i(k+1) = A(x(k)) + B(x(k)) \hat{u}_i(x(k), w)$ and the notation means minimization for a set of points $x(k)$ selected from the compact set $\Omega \in \mathbb{R}^n$.

Note that the control weights W_{ui} appear in (28) in an implicit fashion, i.e., it is difficult to solve explicitly for the weights because the current control weights determine $x(k+1)$. Therefore, one can use an LMS algorithm on a training set constructed from Ω. The weight update can be written as follows:

$$W_{ui}\big|_{m+1} = W_{ui}\big|_m - \alpha \left. \frac{\partial \left(x(k)^T Qx(k) + M(\hat{u}(x(k), W_{ui}|_m)) + \hat{V}_i(x(k+1)) \right)}{\partial W_{ui}} \right|_{W_{ui}|_m}$$
$$= W_{ui}\big|_m - \alpha \sigma(x(k)) \left(2\lambda R \tanh^{-1}(\lambda^{-1} u(k)) \right)^T$$
$$-\alpha \sigma(x(k)) \left(B^T(x(k)) \frac{\partial \Phi(x(k+1))}{\partial x(k+1)} W_{Vi} \right)^T \quad (29)$$

where α is a positive step size and m is the iteration number for the LMS algorithm. By a stochastic-approximation-type [6] argument, the weights $W_{ui}\big|_m \to W_{ui}$, as $m \to \infty$, and satisfy (28).

It can be observed from (27) that, neither $A(x(k))$ nor $B(x(k))$ is needed to update weights of the NN (21) used for value function approximation. Only the input coupling term $B(x(k))$ is needed to update weights of the NN (22). Therefore, the proposed algorithm works for a system with partially unknown dynamics—no knowledge of the internal feedback structure $A(x(k))$ is needed.

In the next section, simulation experiment carried out on a nonlinear discrete-time system to validate proposed algorithm, is described.

V. SIMULATION EXPERIMENT

Consider the following nonlinear discrete-time system:

$$x(k+1) = A(x(k)) + B(x(k))u(k) + B(x(k)) f(k)$$

where $A(x(k)) = \begin{bmatrix} 0.2 x_1(k) \exp(x_2^2(k)) \\ 0.3 x_2^3(k) \end{bmatrix}$,

$B(x(k)) = \begin{bmatrix} 0 \\ -0.2 \end{bmatrix}, f(x(k)) = \begin{bmatrix} 0 \\ p x_1(k) \sin(x_2(k)) \end{bmatrix}$;

p is the unknown parameter. For simplicity let us assume that $p \in [-1,1]$. This mathematical model is in the matched uncertainty form and

$$\left| f(x(k)) \right| = f(x(k))^T f(x(k)) \le p^2 \left| x_1(k) \right| \le \left| x_1(k) \right| = f_{\max}(x(k))$$

The performance function is defined as

$$J(x(k),u) = \sum_{i=k}^{\infty} \left(\rho f_{\max}^2(x(k)) + x(i)^T Qx(i) + 2 \int_0^{u(i)} \left(\tilde{\lambda} \varphi^{-1}(\tilde{\lambda}^{-1} v) \right) R dv \right)$$

where $Q = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ and $R = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ and control constraint is set to $\left| u(k) \right| \le 0.3$, i.e., $\lambda = 0.3$.

The approximation of value function is given as

$$\hat{V}_i(x) = W_{Vi}^T \Phi(x)$$
$$= w_{Vi}^1 x_1^2 + w_{Vi}^2 x_2^2 + w_{Vi}^3 x_1 x_2 + w_{Vi}^4 x_1^4 + w_{Vi}^5 x_2^4 + w_{Vi}^6 x_1^3 x_2 + w_{Vi}^7 x_1^2 x_2^2$$
$$+ w_{Vi}^8 x_1 x_2^3 + w_{Vi}^9 x_1^6 + w_{Vi}^{10} x_2^6 + w_{Vi}^{11} x_1^5 x_2 + w_{Vi}^{12} x_1^4 x_2^2 + w_{Vi}^{13} x_1^3 x_2^3$$

NN weights can be found from (27). The approximation of control input is given as $\hat{u}_i(x) = W_{ui}^T \sigma(x), \quad \sigma(x) = \Phi(x)$ is selected for simplicity. NN weights can be found from tuning law (29).

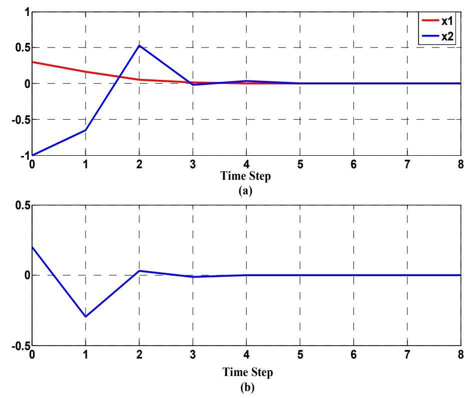

Figure 1: (a) System states Vs. Time step, (b) Control Vs. Time step

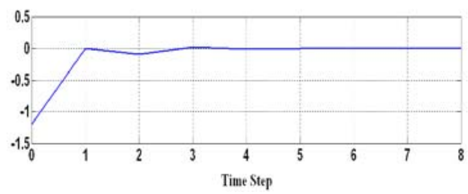

Figure 2: $\Delta \hat{V}^{n}$ Vs. Time step

It is to be noted that an optimal control law (9) has been derived for the nominal dynamics only. Then it was used with actual uncertain system. We have selected $\rho = 5$ and $\alpha = 0.4$ for simulation purpose. Results are shown in the figure 1. It can be observed from figure 1 (a) that, the system states converge to the equilibrium point. Also control input remains bounded, i.e., $|u(k)| \le 0.3$ as shown in figure 1(b).It has been also shown in the figure 2 that $\left(\hat{V}(A(x(k)) + B(x(k))u^*(k)) + B(x(k))f(k))) - \hat{V}(x(k))\right) \le 0.$ The boundedness of control input and convergence of the system state to the equilibrium point validates proposed algorithm.

VI. CONCLUSIONS

In this paper, an algorithm has been proposed to design a constrained robust control law for a class of the discrete-time systems having matched system uncertainties. A non-quadratic performance functional is proposed to tackle the constraints on the input. HDP based algorithm is developed to find the solution of DT-HJB equation with necessary theoretical results. Two neural networks have been used to approximate the value function and optimal control law, respectively. Least squares based method is used in NN based HDP algorithm to find the approximate solution of DT-HJB equation. Formulation we have developed in this paper is for single input case; but one can extend it for multiple input cases. However, the analytical proof related to the conditions for the selection of the design parameter mentioned in the section III-B is can be considered as the future work.

REFERENCES

[1] Saberi A, Lin Z, Teel A., "Control of linear systems with saturating actuators". IEEE Transaction Automatic Control, 1996; Vol.41, No. 3, pp.368–378.

[2] Widrow B, Gupta N, Maitra S., "Punish/reward: learning with a critic in adaptive threshold systems", IEEE Transaction on System Man Cybernetics-Part B ,1973;Vol.3, No.5, pp.455–65.

[3] Watkins C., "Learning from delayed rewards. Ph.D. thesis, Cambridge University, Cambridge, England; 1989.

[4] Bertsekas DP, Tsitsiklis JN. Neuro-dynamic programming. Massachusetts: Athena Scientific; 1996.

[5] Liu X, Balakrishnan SN. Convergence analysis of adaptive critic based optimal control. In: Proceedings of American control conference, Chicago, USA; 2000. pp. 1929–33.

[6] Si J, Wang YT. On-line learning control by association and reinforcement., IEEE Transaction on Neural Networks, 2001, Vol.12, No.2, pp.264–276.

[7] Abu-Khalaf M, Lewis FL. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach, Automatica 2005,Vol.41, No. 5, pp.779–791.

[8] Asma Al-Tamimi A, Lewis FL. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. In Proceedings of the IEEE symposium on approximate dynamic programming and reinforcement learning, USA; 2007. pp. 38–43.

[9] K. Hornik, M. Stinchcombe, and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," Neural Networks., 1990, vol. 3, no. 5, pp. 551–560.

[10] Asma Al-Tamimi A, Lewis F. L. and M.Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof", IEEE Transaction on System Man Cybernetics-Part B, vol.38 ,No. 4, pp.943-949, 2008.

[11] Shaosheng Zhou, Gang Feng, James Lam and Shengyuan Xu, "Robust H_∞control for discrete-time fuzzy systems via basis-dependent Lyapunov functions", Information Sciences, 2005,Vol. 174, No. 3, pp.197-217

[12] C.S. Berger, "Robust control of discrete systems", IEE Proceedings-Control Theory and Applications, 1989, Vol. 136, No. 4, pp. 165-170

[13] Ho DWC, Lu G. "Robust stabilization for a class of discrete-time non-linear systems via output feedback: the unified LMI approach". International Journal of Control, 2003, Vol. 76,pp. 105–15.

[14] S. E. Lyshevski, "Nonlinear discrete-time systems: Constrained optimization and application of nonquadratic costs", Proceedings of the American Control Conference,Philadelphia, 1998,pp.3699-3703.

[15] Derong Liu, Xiaoxu Xiong and Yi Zhang, "Action-dependent adaptive critic designs", Proceedings of the IJCNN'01,USA,pp.990-995.