51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

A Self-tuning Optimal Controller for Affine Nonlinear Continuous-time Systems with Unknown Internal Dynamics

Travis Dierks[1], *Member, IEEE*, and S. Jagannathan[2], *Senior Member, IEEE*

Abstract—**This paper presents a novel neural network (NN) - based self-tuning controller for the optimal regulation of affine nonlinear continuous-time systems. Knowledge of the internal system dynamics is not required whereas the control coefficient matrix is considered to be available. The proposed nonlinear optimal regulator tunes itself in order to simultaneously learn the optimal control input, optimal cost function, and the system internal dynamics using a single NN. A novel NN weight tuning algorithm is derived which ensures the internal system dynamics are learned while simultaneously minimizing a predefined cost function. An initial stabilizing controller is not required. Lyapunov methods are used to show that all signals are uniformly ultimately bounded (UUB). In the absence of NN reconstruction errors, the approximated control input is shown to converge to the optimal control asymptotically for the regulator design, and simulation results illustrate the effectiveness of the approach.**

I. Introduction

Optimality, in addition to stability, is a desirable characteristic of both linear and nonlinear controller designs since optimal controllers minimize predefined cost functions in order to meet design specifications [1]. In contrast to linear systems with quadratic cost functions, the optimal control of nonlinear system poses a significant challenge since the optimal controller designs require finding the solutions to nonlinear Bellman equations [1], and closed form solutions are generally unavailable. As a result, many research efforts in both offline [2]-[4] and online [5]-[9] dynamic programming and optimal control for nonlinear systems have recently occurred.

Online optimal controllers for nonlinear affine systems are a special class of self-tuning controllers which minimize a predefined cost function [5]-[8]. Most current approaches require two or more [5]-[7] online approximators (OLA's), such as neural networks (NN), to learn the cost function and control input whereas recent developments illustrate how both the optimal cost function and control input can be achieved by using a single OLA (SOLA) [6]. However, in all current online optimal controller designs [5]-[8], the persistency of excitation (PE) condition is required in order to ensure the optimal cost function learned, and complete knowledge of the system dynamics is assumed to be available [5], [8]. To relax the requirement of known internal dynamics, an additional OLA may be introduced [6] at the cost of more computational demand.

Other efforts in online optimal control have been proposed which do not require the internal system dynamics to be known. In [7], two NN's are used to learn the optimal control and cost function for nonlinear discrete-time systems by exploiting the discrete nature of the system dynamics. However, the methods in [7] are only applicable for the discrete-time regulation problem. In [9], a policy iteration-based approach was proposed to control partially unknown nonlinear systems using iterative solutions of HJB equation. Besides requiring an initial stabilizing controller, iterative methods are often unsuitable for hardware implementation since the number of iterations required for convergence between consecutive samples is unknown.

In this work, the optimal regulation of nonlinear systems with unknown internal dynamics is undertaken through the definition of a specific cost function. As a result, the proposed optimal regulation scheme simultaneously learns the optimal control input, cost function and internal dynamics using a SOLA, such as a NN. Novel parameter tuning laws are derived under two cases: an ideal case where approximator reconstruction errors are negligible, as would be the case for standard adaptive control [13], and a non-ideal case where non-zero reconstruction errors are considered, as would be the case when NN are considered [10]. Lyapunov methods demonstrate the closed loop stability in both cases, and simulation results are presented to confirm the theoretical results.

The contributions of this work include the development of a novel online adaptive optimal controller that minimizes a predefined cost function without using knowledge of the system internal dynamics. The proposed adaptive controller does not require an initial stabilizing control since the NN weight tuning law tunes itself in order to ensure system stability in addition to tuning to learn the optimal control input. Although a specific form of the traditional cost function is selected in this work, the proposed method still allows for freedom of design to penalize the system states or control input as required to meet design specifications. Further, the proposed solution is less computationally demanding than previous results in online adaptive optimal control [5], since only one NN is required in this work whereas three NNs would be required to extend traditional actor-critic architectures like [5] to include uncertainties in the system dynamics.

II. Background

Consider the affine nonlinear continuous-time system

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

[1] Travis Dierks is with DRS Sustainment Systems, Inc., 201 Evans Lane, St Louis, MO 63121. Contact author Email: tdierks@drs.com.

[2] S. Jagannathan is with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology (formerly University of Missouri Rolla), 1870 Miner Circle, Rolla, MO 65409.

Research supported in part by NSF ECCS #0901562 and #1128281.

978-1-4673-2066-5/12/$31.00 ©2012 IEEE

where $x \in \Re^n$, $f(x) \in \Re^n$ and $g(x) \in \Re^{nxm}$ are smooth nonlinear functions with $g(x)$ satisfying $g_m \leq \| g(x) \|_F \leq g_M$ where g_m and g_M are known positive constants when the Frobenius norm [10] is applied, and $u \in \Re^m$ is the control input. In this work, it is assumed that $f(x)$ is unknown while $g(x)$ is known. Without loss of generality, assume that the system controllable with $x = 0$ a unique equilibrium point on a compact set $\Omega \in \Re^n$ with $f(0) = 0$. Under these conditions, the optimal control input for the nonlinear system (1) can be calculated [1].

The infinite horizon HJB cost function for (1) is given by

$$V(x(t)) = \int_t^\infty r(x(\tau), u(\tau))\, d\tau \tag{2}$$

where $r(x,u) = Q(x) + u^T Ru$, $Q(x) \geq 0$ is the positive semi - definite with penalty on the states, and $R \in \Re^{mxm}$ is a positive definite matrix. The control input, u, must be admissible so that the cost function (2) is finite [3].

Next, the Hamiltonian for the cost function (2) with an associated admissible control input u is written as [1]

$$H(x,u) = r(x,u) + V_x^T(x)(f(x) + g(x)u) \tag{3}$$

where $V_x(x)$ is the gradient of the cost or value function $V(x)$ with respect to x. It is well known that the optimal control input that minimizes the cost function (2) also minimizes the Hamiltonian (3); therefore, the optimal control is found by solving the stationarity condition $\partial H(x,u)/\partial u = 0$ and revealed to be [1]

$$u^*(x) = -R^{-1} g(x)^T V_x^*(x)/2\ . \tag{4}$$

Substituting the optimal control (4) into the Hamiltonian (3) while observing $H(x,u^*) = 0$ reveals the Hamilton-Jacobi-Bellman (HJB) equation and the necessary and sufficient condition for optimal control to be [1]

$$0 = Q(x) + V_x^{*T}(x) f(x) - V_x^{*T}(x) g(x) R^{-1} g(x)^T V_x^*(x)/4 \tag{5}$$

with the value function $V^*(0) = 0$. It is observed that both V^* and V_x^* are generally unavailable thus spawning active research efforts in online dynamic programming and optimal control for nonlinear systems.

In this work, a specific form of the cost function (2) is utilized in the development of the novel self-tuning optimal controller. That is, in this work, we define

$$Q(x) = V_x^{*T}(x) Q V_x^*(x) \tag{6}$$

where $Q \in \Re^{mxm}$ is a constant positive definite matrix. Then, the Hamiltonian (5) can be written as

$$0 = V_x^{*T}(x) Q V_x^*(x) + V_x^{*T}(x) g(x) R^{-1} g(x)^T V_x^*(x)/4 + V_x^T(x)(f(x) + g(x)u^*)$$

or

$$V_x^T(x)\dot{x}^* = -V_x^{*T}(x) Q V_x^*(x) - V_x^{*T}(x) g(x) R^{-1} g(x)^T V_x^*(x)/4 \tag{7}$$

where $\dot{x}^* = f(x) + g(x)u^*$ is the optimal trajectory and $V_x^{*T}(x)\dot{x}^* = \dot{V}^*(x)$.

Remark 1: By choosing $Q(x)$ as shown in (6), it will be shown that the requirement for knowledge of the internal system dynamics can be relaxed, and an optimal control input is achieved by using only one NN. In addition, the proposed choice of $Q(x)$ in (6) still allows for freedom of design to penalize the system states as required in order to meet design specifications.

Before proceeding, the following technical lemmas are required.

Lemma 1. Given the nonlinear system (1) and associated cost function (2) with $Q(x)$ given by (6), and optimal control (4), let $J(x)$ be a continuously differentiable, radially unbounded Lyapunov candidate such that $\dot{J}(x) = J_x^T(x)\dot{x} = J_x^T(x)(f(x) + g(x)u^*) < 0$ with $J_x(x)$ being the radially unbounded and continuously differentiable partial derivative of $J(x)$. Moreover, let $\overline{Q}(x) \in \Re^{nxn}$ be a differentiable positive definite matrix satisfying

$$V_x^{*T}(qI + \overline{Q}(x))J_x = r(x,u^*) = Q(x) + u^{*T} Ru^* \tag{8}$$

where q is a positive constant. Then, the following relation holds

$$J_x^T(f(x) + g(x)u^*) = -J_x^T(qI + \overline{Q}(x))J_x\ . \tag{9}$$

Proof: When the optimal control (4) is applied to the nonlinear system (1), the cost function (2) becomes a Lyapunov function rendering

$$\begin{aligned} \dot{V}^*(x) &= V_x^{*T}(x)\dot{x} = V_x^{*T}(x)(f(x) + g(x)u^*) \\ &= -Q(x) - u^{*T} Ru^* \end{aligned} \tag{10}$$

from (7). After manipulation and substitution of (8), equation (10) is rewritten as

$$\begin{aligned} f(x) + g(x)u^* &= -(V_x^* V_x^{*T})^{-1} V_x^*(Q(x) + u^{*T} Ru^*) \\ &= -(V_x^* V_x^{*T})^{-1} V_x^* V_x^{*T}(qI + \overline{Q}(x))J_x \\ &= -(qI + \overline{Q}(x))J_x \end{aligned} \tag{11}$$

Now, multiply both sides of (11) by J_x^T yields the desired relationship in (9). ■

Lemma 2. Given the nonlinear system (1) and associated cost function (2) with $Q(x)$ given by (6) and optimal control (4), the internal dynamics $f(x)$ can be rewritten in a parameterized form given by

$$f(x) = -\left(Q - \frac{1}{4} g(x) R^{-1} g(x)^T \right) V_x^*(x)\ . \tag{12}$$

Proof: When the optimal control (4) is applied to the nonlinear system (1), the cost function (2) becomes a Lyapunov function rendering

$$V_x^T(x)(f(x) - g(x)g(x)^T V_x^*(x)/2) = \\ -V_x^{*T}(x) Q V_x^*(x) - \frac{1}{4} V_x^{*T}(x) g(x)) R^{-1} g(x)^T V_x^*(x)$$

After further manipulation

$$V_x^T(x)f(x) = -V_x^{*T}(x)QV_x^*(x) + \frac{1}{4}V_x^{*T}(x)g(x))R^{-1}g(x)^T V_x^*(x)$$

or

$$V_x^T(x)\left(f(x) + QV_x^*(x) - \frac{1}{4}g(x)g)R^{-1}(x)^T V_x^*(x) \right) = 0\,. \tag{13}$$

Observing that $V_x^{*T}(x) \neq 0$ for $\| x \| > 0$, the relation in (12) is obtained. ■

Remark 2: The general relationship between the optimal cost function and the internal system dynamics has been noted in [11] where converse optimal control was considered. However, the specific relationship in (12) is a result of selecting the penalty matrix $Q(x)$ in (6).

In the next section, the novel NN-based self-tuning regulation scheme will be introduced.

III. Neural Network-based Self Tuning optimal Regulator

To begin the development, we express the cost function (2) using a NN as

$$V(x) = W^T\phi(x) + \varepsilon(x) \tag{14}$$

where $W \in \Re^L$ is the constant target NN weight vector, $\phi(\cdot) : \Re^n \to \Re^L$ is a linearly independent basis vector which satisfies $\phi(0) = 0$, and $\varepsilon(x)$ is the NN reconstruction error. The target NN vector and reconstruction errors are assumed to be upper bounded according to $\|W\| \leq W_M$ and $\|\varepsilon(x)\| \leq \varepsilon_M$, respectively [10]. In addition, it will be assumed that the gradient of the NN reconstruction error with respect to x is upper bounded according to $\|\partial\varepsilon(x)/\partial x\| = \|\nabla_x \varepsilon(x)\| \leq \varepsilon'_M$ [12], [5], [8].

Moving on, the gradient of the NN cost function (14) is

$$\partial V(x)/\partial x = V_x(x) = \nabla_x^T\phi(x)W + \nabla_x\varepsilon(x) \tag{15}$$

Now, using (15), the control (4) is rewritten as

$$u^*(x) = -R^{-1}g(x)^T\nabla_x^T\phi(x)W/2 - R^{-1}g(x)^T\nabla_x\varepsilon(x)/2 \tag{16}$$

Moving on, the NN estimate of (14) is now written as

$$\hat{V}(x) = \hat{W}^T\phi(x) \tag{17}$$

where $\hat{W}$ is the NN estimate of the target weight vector W. Similarly, the estimate of the target control input (16) is written in terms of $\hat{W}$ as

$$\hat{u}(x) = -R^{-1}g(x)^T\nabla_x^T\phi(x)\hat{W}/2 \tag{18}$$

Substituting (18) into the nonlinear system (1) reveals

$$\dot{x} = f(x) - g(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\hat{W}/2\,,$$

and after manipulation

$$\begin{aligned} \dot{x} &= f(x) - g(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\hat{W}/2 \pm g(x)u^*(x) \\ &= f(x) + g(x)u^*(x) + \frac{g(x)R^{-1}g(x)^T}{2}\nabla_x^T\phi(x)\tilde{W} + \frac{g(x)R^{-1}g(x)^T}{2}\nabla_x\varepsilon(x) \end{aligned} \tag{19}$$

where $\tilde{W} = W - \hat{W}$ is the weight estimation error. To ensure that the parameter estimate, $\hat{W}$, converges to its target optimal value, W, previous works [5]-[6] select the tuning law to minimize an auxiliary error function defined as the square of the estimated Hamiltonian. In this work, convergence to the optimal value is achieved by identifying the systems internal dynamics. That is, the results of *Lemma 2* allow a state estimator to be defined such that by learning the system internal dynamics, the Hamiltonian is also minimized.

To begin, define $\hat{x}$ to be the state of the system estimator defined as

$$\dot{\hat{x}} = \hat{f}(x) + g(x)u + K\tilde{x} \tag{20}$$

where $K > 0$ is a constant design matrix, $\tilde{x} = x - \hat{x}$ is the state estimation error, and the estimate of the internal dynamics is

$$\begin{aligned} \hat{f}(x) &= -(Q - g(x)R^{-1}g(x)^T/4)\hat{V}_x(x) \\ &= -(Q - g(x)R^{-1}g(x)^T/4)\nabla_x^T\phi(x)\hat{W} \end{aligned}\,. \tag{21}$$

Note that introduction of the state estimator does not introduce an additional NN. Instead, it uses the cost function approximator and leverages the properties of the system highlighted in *Lemma 2*. Moving on and subtracting (20) from (1), the estimation error dynamics are

$$\begin{aligned} \dot{\tilde{x}} &= f(x) - \hat{f}(x) - K\tilde{x} \\ &= \tilde{f}(x) - K\tilde{x} - (Q - g(x)R^{-1}g(x)^T/4)\nabla_x\varepsilon(x) \end{aligned} \tag{22}$$

where $\tilde{f}(x) = -(Q - g(x)R^{-1}g(x)^T/4)\nabla_x^T\phi(x)\tilde{W}$. It is observed that the convergence of $\tilde{f}(x)$ ensures that the relations in (12) and (13) have been satisfied. That is, $\hat{W}$ has converged to W and the estimated control input (18) has converged to optimal control input (4).

Next, *Theorem 1* illustrates the stability of the nonlinear system (19) in the ideal case of $\varepsilon(x) = \| \nabla_x\varepsilon(x) \| = 0$ when the standard adaptive control is utilized [13]. Then, *Theorem 2* shows the stability in the non-ideal case of a NN-based OLA using a modified parameter tuning law. In both cases, asymptotic convergence of the system states can be shown when there are no OLA approximation errors. First, the following definition is required.

Definition 1 [10]: An equilibrium point x_e is said to be uniformly ultimately bounded (UUB) if there exists a compact set $S \subset \Re^n$ so that for all $x_0 \in S$ there exists a bound B and a time $T(B, x_o)$ such that $\|x(t) - x_e\| \leq B$ for all $t \geq t_0 + T$.

Theorem 1 (Self-tuning Optimal Regulator in an Ideal Case)*:* Given the nonlinear system (1) and the cost function (2) with $Q(x)$ given by (6), let the control input be given by (18) with the NN weight tuning law as

$$\begin{aligned} \dot{\hat{W}} &= \frac{\alpha}{2}\nabla_x\phi(x)g(x)R^{-1}g(x)^T J_x(x) \\ &\quad -\alpha v\nabla_x\phi(x)(Q - g(x)R^{-1}g(x)^T/4)\tilde{x} \end{aligned} \tag{23}$$

where $\alpha > 0$ and $v > 0$ are design parameters and $J_x(x)$ is described in *Lemma 1*. Then, the closed loop systems (19) and (22) with $\| \nabla_x \varepsilon(x) \| = 0$ as well as the NN weight estimation errors are asymptotically stable. That is, $\hat{V}(x) \to V(x)$ and $\hat{u}(x) \to u^*(x)$.

Proof: Considered the positive definite Lyapunov candidate

$$L(x,\tilde{x},\tilde{W}) = J(x) + \frac{\nu}{2}\tilde{x}^T\tilde{x} + \frac{1}{2\alpha}\tilde{W}^T\tilde{W}$$

where $J(x)$ is given by *Lemma 1*. Taking the derivative of $L(x,\tilde{x},\tilde{W})$ with respect to time and substituting (19) and (22) with $\| \nabla_x \varepsilon(x) \| = 0$ reveals

$$\begin{aligned}\dot{L}(x,\tilde{x},\tilde{W}) = {} & J_x^T(x)(f(x) + g(x)u^*(x)) - \nu\tilde{x}K\tilde{x} \\ & + \frac{1}{\alpha}(\tilde{W}^T(\alpha\nabla_x\phi(x)g(x)R^{-1}g(x)^T J_x(x)/2 \\ & - \nu\alpha\nabla_x\phi(x)(Q - g(x)R^{-1}g(x)^T/4)\tilde{x} + \dot{\tilde{W}}))\end{aligned}$$

Then, observing $\dot{\tilde{W}} = -\dot{\hat{W}}$, using the tuning law (23), and applying the results of *Lemma 1* gives

$$\begin{aligned}\dot{L}(x,\tilde{x},\tilde{W}) &= J_x^T(x)(f(x) + g(x)u^*(x)) - \nu\tilde{x}K\tilde{x} \\ &= -J_x^T(qI + \overline{Q}(x))J_x - \nu\tilde{x}K\tilde{x} \le 0\end{aligned}.$$

Finally, using the properties of J_x and $\overline{Q}(x)$ described in *Lemma 1*, $\ddot{L}(x,\tilde{x},\tilde{W})$ is shown to be bounded, and Barbalat's Lemma [10] can be invoked to conclude the asymptotic stability of the system states, parameter estimation error, and the NN weight estimation error. Note that the convergence of the parameter estimation error ensures $\hat{f}(x) \to f(x)$ and thus $\hat{W} \to W$. As a result, it is straight forward to show that $\hat{V}(x) \to V(x)$ and $\hat{u}(x) \to u^*(x)$. ■

Remark 3: Traditional applications of Barbalat's Lemma to analyze the stability of NN controllers typically yield asymptotic convergence of the system states whereas the NN weight estimation errors are shown to be bounded [10]. In contrast, this work is able to show the asymptotic convergence of both the system states and NN estimation error when applying Barbalat's Lemma through the introduction of the state estimator (20) and (21).

Theorem 2 (Self-tuning Optimal Regulator with Augmented Tuning): Given the nonlinear system (1) and the cost function (2) with $Q(x)$ given by (6), let the control input be given by (18) with the NN weight tuning law as

$$\begin{aligned}\dot{\hat{W}} = {} & \alpha\nabla_x\phi(x)g(x)R^{-1}g(x)^T J_x(x)/2 \\ & -\alpha\kappa\nabla_x\phi(x)g(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\hat{W} \\ & -\alpha\nu\nabla_x\phi(x)(Q - g(x)R^{-1}g(x)^T/4)\tilde{x}\end{aligned} \tag{24}$$

where $\alpha > 0$ and $\kappa > 0$ are a design parameters and $J_x(x)$ is described in *Lemma 1*. Then, the closed loop system (19) and the NN weight estimation error are UUB. That is, $|\hat{V}(x) - V(x)| \le \varepsilon_V$ and $\| \hat{u}(x) - u^*(x) \| \le \varepsilon_u$ where ε_u and ε_V are small positive constants.

Proof: Considered the Lyapunov candidate

$$L(x,\tilde{x},\tilde{W}) = V(x) + J(x) + \frac{\nu}{2}\tilde{x}^T\tilde{x} + \frac{1}{2\alpha}\tilde{W}^T\tilde{W} \tag{25}$$

where $V(x)$ is given by (2) and $J(x)$ is given in *Lemma 1*. Taking the first derivative of (25) and substituting the closed loop system dynamics (19) and identifier error dynamics (22), observing $\dot{\tilde{W}} = -\dot{\hat{W}}$ and using the tuning law (24), and substituting (7) and (9) gives

$$\begin{aligned}\dot{L}(x,\tilde{x},\tilde{W}) = {} & -V_x^T(x)(Q + g(x)R^{-1}g(x)^T/4)V_x(x) - J_x^T(qI + \overline{Q}(x))J_x - \nu\tilde{x}^T K\tilde{x} \\ & + \kappa\tilde{W}^T\nabla_x\phi(x)g(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\hat{W} + V_x^T(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\tilde{W}/2 \\ & + J_x^T(x)R^{-1}g(x)^T\nabla_x\varepsilon(x)/2 + V_x^T(x)R^{-1}g(x)^T\nabla_x\varepsilon(x)/2 \\ & - \nu\tilde{x}^T g(x)(Q - (1/4)I)g(x)^T\nabla_x\varepsilon(x))\end{aligned} \tag{26}$$

Next, adding and subtracting $\kappa\tilde{W}^T\nabla_x\phi(x)g(x)R^{-1}g(x)^T V_x(x)$ and using (15) allows (26) to be rewritten as

$$\begin{aligned}\dot{L}(x,\tilde{x},\tilde{W}) = {} & -V_x^T(x)(Q + g(x)R^{-1}g(x)^T/4)V_x(x) - J_x^T(qI + \overline{Q}(x))J_x \\ & -\nu\tilde{x}^T K\tilde{x} - \kappa\tilde{W}^T\nabla_x\phi(x)g(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\tilde{W} \\ & +(\kappa + 1/2)\tilde{W}^T\nabla_x\phi(x)g(x)R^{-1}g(x)^T V_x(x) \\ & +J_x^T(x)R^{-1}g(x)^T\nabla_x\varepsilon(x)/2 + V_x^T(x)R^{-1}g(x)^T\nabla_x\varepsilon(x)/2 \\ & -\kappa\tilde{W}^T\nabla_x\phi(x)g(x)R^{-1}g(x)^T\nabla_x\varepsilon(x) \\ & -\nu\tilde{x}^T(Q - (1/4)I)g(x)^T\nabla_x\varepsilon(x))\end{aligned}$$

After completing the squares, $\dot{L}(x,\tilde{x},\tilde{W})$ is upper bounded according to

$$\begin{aligned}\dot{L}(x,\tilde{x},\tilde{W}) \le {} & -J_x^T(qI + \overline{Q}(x) - g(x)R^{-1}g(x)^T)J_x/2 \\ & -V_x^T(x)\left(Q + \left(1/8 - 3(\kappa + 1/2)^2/4\kappa\right)g(x)R^{-1}g(x)^T\right)V_x(x) \\ & -\nu\tilde{x}^T(K - (Q - g(x)R^{-1}g(x)^T/4)^2)\tilde{x} \\ & -\frac{\kappa}{3}\tilde{W}^T\nabla_x\phi(x)g(x)R^{-1}g(x)^T\nabla_x^T\phi(x)\tilde{W} \\ & +\left(\lambda_{\max}(R^{-1})\left(5/8 + 3\kappa g_M^2/4\right) + \nu/4\right)\varepsilon_M'^2\end{aligned}$$

where $\lambda_{\max}(\bullet)$ is the maximum singular value operator. Further, $\dot{L}(x,,\tilde{x}\tilde{W}) \le 0$ provided $\overline{Q}(x) > g(x)R^{-1}g(x)^T/2$, $\lambda_{\min}(Q) > 3(\kappa + 1/2)^2 g_M^2 \lambda_{\max}(R^{-1})/(4\kappa) - 1/8 g_m^2 \lambda_{\min}(R^{-1})$, where $\lambda_{\min}(\bullet)$ is the minimum singular value operator, $K_{\min} > 2\lambda_{\max}(Q)^2 + g_M^4 \lambda_{\max}(R^{-1})/8$ and the following inequalities hold

$$\begin{aligned} & \| V_x(x) \|^2 > \frac{\varepsilon_M'^2(\lambda_{\max}(R^{-1})\left(5/2 + 3\kappa g_M^2\right) + \nu)}{4(\lambda_{\min}(Q) + 1/8 g_m^2 \lambda_{\min}(R^{-1}) - \mu_V)} \quad \text{or} \\ & \| J_x \|^2 > \frac{\varepsilon_M'^2(\lambda_{\max}(R^{-1})\left(5/2 + 3\kappa g_M^2\right) + \nu)}{4q} \quad \text{or} \\ & \| \tilde{x} \|^2 > \frac{8\varepsilon_M'^2 \lambda_{\max}(R^{-1})\left(5/2 + 3\kappa g_M^2\right) + \nu)}{4\nu(K_{\min} - \mu_{\tilde{x}})} \quad \text{or} \\ & \| \tilde{W}^T\nabla_x\phi(x) \|^2 > \frac{3\varepsilon_M'^2 \lambda_{\max}(R^{-1})\left(5/2 + 3\kappa g_M^2\right) + \nu)}{4\kappa g_m^2 \lambda_{\min}(R^{-1})}\end{aligned} \tag{27}$$

where $K_{\min}$ is the minimum singular value of K,

$\mu_V = 3(\kappa + 1/2)^2 g_M^2 \lambda_{\max}(R^{-1})/(4\kappa)$ and $\mu_{\tilde{x}} = 2\lambda_{\max}(Q)^2 + g_M^4 \lambda_{\max}(R^{-1})$. According to standard Lyapunov extensions [10], the inequalities in (27) guarantee that $\dot{L}(x, \tilde{x}, \tilde{W})$ is less than zero outside of a compact set. Thus, $\| J_x(x) \|$, $\| V_x(x) \|$, $\| \tilde{x} \|$ and $\| \tilde{W}^T \nabla_x \phi(x) \|$ remain bounded, and recalling the properties of $\left\| J_x(x) \right\|$ set in *Lemma 1*, the boundedness of $\left\| J_x(x) \right\|$ implies the boundedness of the system states, $\left\| x \right\|$. Further, the boundedness of the system states ensures the basis function and its gradient are also bounded such that $\| \phi(x) \| \le \phi_M$ and $\| \nabla_x \phi(x) \| \le \phi'_M$ for positive constants ϕ_M and ϕ'_M. As a result, the boundedness of $\| \tilde{W}^T \nabla_x \phi(x) \|$ and $\| \nabla_x \phi(x) \|$ also ensures the boundedness of $\| \tilde{W} \|$ such that $\| \tilde{W} \| \le B_{\tilde{W}}$ for a constant $B_{\tilde{W}}$ when $\| x \| > 0$. Note that $\dot{\tilde{W}} = 0$ when $\| x \| = 0$; thus, $\| \tilde{W} \|$ is bounded for this case as well. To complete the proof, observe

$$| \hat{V}(x) - V(x) | = | -\tilde{W}^T \phi(x) - \varepsilon(x) | \le B_{\tilde{W}} \phi_M + \varepsilon_M \equiv \varepsilon_V$$

and

$$| \hat{u}(x) - u^*(x) | = | -\nabla_x^T \phi(x) \tilde{W} - \nabla_x \varepsilon(x) | \le B_{\tilde{W}} \phi'_M + \varepsilon'_M \equiv \varepsilon_u \blacksquare$$

Remark 4: To combat parameter drift which may occur when the PE condition is not satisfied, standard NN weight tuning algorithms are often augmented with extra terms such as the e-modification or σ-modification [10]. In such works, the system states and parameter estimation errors are shown to be *UUB* with fundamental lower bounds dependent upon the constant target weight matrix. Consequently, even if NN reconstruction errors are negligible, asymptotic convergence of the system states cannot be shown. In contrast, the novel augmented tuning term introduced in (24) of this work does not suffer from this deficiency since the lower bound of the systems states is a function of the NN reconstruction error alone. Therefore, asymptotic convergence can be shown in this work when the NN reconstruction errors are negligible.

Remark 5: It is observed that convergence to the optimal control is achieved independently of an initial stabilizing control. This is accomplished through the first term in the tuning law (24) which tunes $\hat{W}$ to become a stabilizing control. Thus, an initial stabilizing control does not have to be calculated which can be difficult when $f(x)$ is unknown. A similar approach to relax the requirement for an initial stabilizing control was used in [8] where an indicator function was used to monitor the stability of the system while the Hamiltonian (3) was being learned. In contrast to [8], an indicator function is not used in this work since the Hamiltonian is learned in a fundamentally different manner and without knowledge of the internal dynamics.

IV. Simulation Results

To demonstrate the effectiveness of the single NN-based regulator design of this work, the optimal regulation problem is solved for a nonlinear system. To implement the online SOLA-based designs, a linear in the parameter (LIP) NN is utilized as the OLA. In addition, the Lyapunov candidate from *Lemma 1* was taken as $J(x) = x^T x / 2$ so that $J_x(x) = x$ in (24).

Consider the nonlinear system in the form of (1) with $x = [x_1 \; x_2]^T$ and

$$\dot{x} = \begin{bmatrix} -2x_1 x_2^2 - \pi x_1 \\ 5\left(x_2 x_1^2 + x_2\right)/2 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} u(x)$$

Using the cost function (2) with $Q(x)$ given by (6) with $Q = I$ and $R = 1$, the optimal cost function is given by $V^*(x) = W_{c4}^* x_1^2 + W_{c5}^* x_2^2 + W_{c6}^* x_2^2 x_1^2$ with $W_{c4}^* = \pi / 2$ and $W_{c5}^* = W_{c6}^* = 1$. Converse optimal control techniques [11] were used to determine these parameters. The basis vector for the SOLA-based scheme implementation was selected as $\phi(x) = [x_1 \;\; x_2 \;\; x_1 x_2 \;\; x_1^2 \;\; x_2^2 \;\; x_1^2 x_2^2 \;\; x_1^3]^T$. The tuning parameters were selected as $\alpha = 0.01$, $\kappa = 0.01, \nu = 200$, and $K = 10$, and the initial conditions were taken as $x(0) = \hat{x}(0) = [1 \; -1]^T$ while all NN weights were initialized to zero. That is, no initial stabilizing control was utilized for implementation of this online design.

Fig. 1 depicts the evolution of the NN weights during the online learning. Starting from zero, the NN weights are tuned to learn the internal system dynamics, and the final values were $\hat{W}_{c4} = 1.5709$, $\hat{W}_{c5} = 0.9956$, and $\hat{W}_{c6} = 1.0050$ with $[\hat{W}_{c1} \;\; \hat{W}_{c2} \; \hat{W}_{c3} \; \hat{W}_{c7}] = [-0.0001 \quad 0.0015 \quad -0.0049 \quad 0.0001]$. These results confirm that the single NN design converges to the actual optimal cost function with small bounded error as the theoretical results suggested. The internal system dynamics estimation error is shown in Fig. 2 where it is observed that convergence to the actual internal system dynamics occurs as the NN weights converge to their target values. In addition, from the results of *Lemma 2,* the convergence of the estimated internal dynamics to their target values guarantee that the Hamiltonian has been learned and minimized.

The evolution of the system states during the online learning is shown in Fig. 3 where probing noise was added to ensure the PE condition is satisfied [5], [8]. After 90 seconds, the PE condition was no longer required since the NN weights converge to constant values and hence it was removed. Even though no initial stabilizing control was used, the system states remained bounded as *Theorem 2* predicted. Not shown, the state estimation error, $\tilde{x}$, is also observed to converge to the origin as the theoretic conjectures ensured.

As a comparison, the single OLA-based algorithm in [8] which requires $f(x)$ to be known was implemented, and the rate of convergence of the NN weights is compared to the convergence rate observed in this work. Fig. 4 shows the results of implementing the algorithm in [8], and it is observed that $\hat{W}_{c4}$ and $\hat{W}_{c5}$ quickly converge to their target

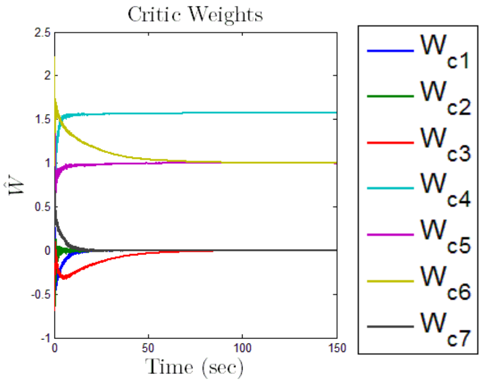

Fig. 1: Convergence of the NN Weights.

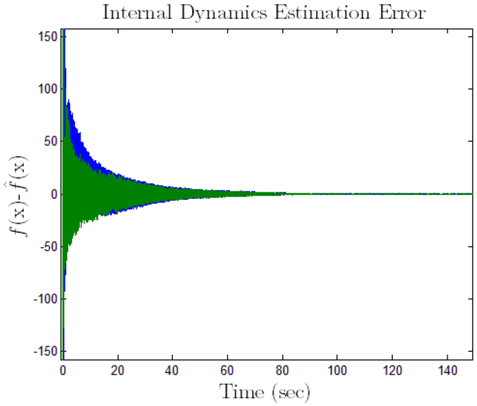

Fig. 2: Convergence of the Hamiltonian and Estimated Internal Dynamics.

values in both cases. On the other hand, $\hat{W}_{c6}$ is observed to converge to its target value after roughly 90 seconds in the proposed algorithm in this work whereas full convergence of $\hat{W}_{c6}$ in [8] is observed to occur after roughly 75 seconds. Thus, the proposed algorithm which does not require $f(x)$ to be known converges at nearly the same rate observed for the algorithm proposed in [8] which requires knowledge of the internal dynamics.

V. Conclusion

In this work, the optimal regulation of nonlinear systems with unknown internal dynamics was undertaken through the definition of a specific cost function. As a result, the proposed optimal regulator scheme simultaneously learned the optimal control input, cost function and internal dynamics using a single NN. The proposed adaptive controller did not require an initial stabilizing control since the NN weight tuning law tuned itself in order to ensure system stability in addition to tuning to learn the optimal control input. Novel NN weight tuning laws were derived, and Lyapunov methods demonstrated the closed loop stability of the approach. Simulation results confirmed the theoretical results, and through a simulation example, the proposed algorithm was shown to converge to the target parameter values at nearly the same rate observed for the SOLA algorithm from the literature which required knowledge of the internal dynamics.

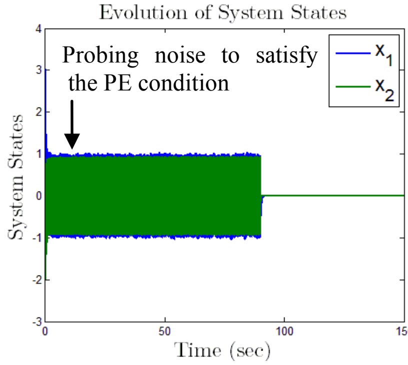

Fig. 3: Evolution of the System States with the PE Condition Satisfied.

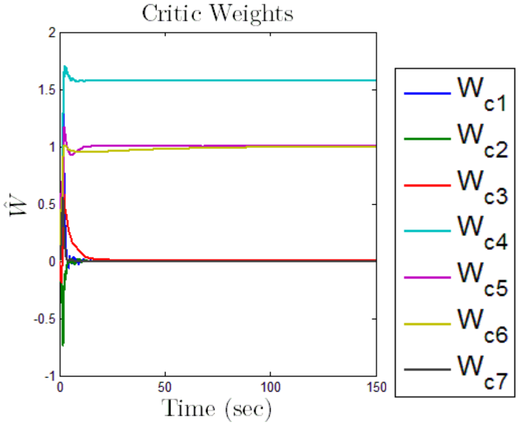

Fig. 4: NN weights for the SOLA Algorithm in [8] with *f*(*x*) known.

References

[1] F. L. Lewis and V. L. Syrmos, *Optimal Control* (2nd ed), Wiley: Hoboken, NJ, 1995.

[2] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Belmonth, MA: Athena Scientific, 2000.

[3] Z. Chen and S. Jagannathan, "Generalized Hamilton-Jacobi-Bellman formulation based neural network control of affine nonlinear discrete-time systems, "*IEEE Trans. Neural Networks,* vol. 10, pp. 90-106, 2008.

[4] J. Si, A. G. Barto, W. B. Powell, and D. Wunsch, *Handbook of Learning and Approx. Dynamics Prog.* Wiley: IEEE Press, 2004.

[5] K. G. Vamvoudakis and F. L. Lewis, "Multi-Player non zero sum Games: Online Adaptive Learning Solution of Coupled Hamilton-Jacobi Equations," *Automatica*, vol. 47, no. 8, pp. 1556-1569, 2011.

[6] T. Dierks and S. Jagannathan, "Optimal tracking control of affine nonlinear discrete-time systems with unknown internal dynamics," *Proc. of the IEEE Conf. on Dec. and Control,* pp. 6750-6755, 2009.

[7] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear discrete-time systems," *in Proc. of the Mediterranean Conference on Control and Automation*, pp. 1390 – 1395, 2009.

[8] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," *Proc. of the American Control Conference* 2010, pp. 1568-1573.

[9] D. Vrabie and F. L. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks*, no. 22, pp. 237-246, 2009.

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

[11] T. Doyle, J. Primbs, B. Shapiro, and V. Nevistic, "Nonlinear games: examples and counter examples," *Proc. Conf. on Decision. Contr.*, pp. 3915-3920, 1996.

[12] K. Hornik, M. Stinchcombe, and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks." *Neural Networks*, vol. 3, pp. 551-560, 1990.

[13] S. K. Narendra, A. M. Annaswamy, *Stable Adaptive Systems*, Prentice-Hall: Englewood Cliffs, NJ, 1989.