2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Self-organized Locally Linear Optimal Tracking Control for Unknown Nonlinear Systems

Yiming Chen, Wenjie Dong and Jay A. Farrell

Abstract— **This paper considers optimal full state feedback tracking control of unknown nonlinear systems. To deal with the unknown nonlinearities in the system, small learning regions are assigned online along the system trajectory in a manner dictated by a Lyapunov based self-organization method. In each of these regions, a local affine approximation is developed. A state observer-based approach adapts the approximator parameters. With the aid of this state observer, analytic optimal controllers are proposed by solving corresponding linear quadratic control problems in each learning region. To show the effectiveness of the proposed controllers, a numerical example is included.**

I. INTRODUCTION

Optimal control is an important branch of control theory research and has been applied extensively in practical applications. Originated from the seminal works of R. E. Bellman [4] and L. S. Pontryagin [30] during the late 1950s, the basic theory of optimal control includes Pontryagin's minimum principle, which leads to necessary conditions for the existence of optimal trajectories, as well as the concept of dynamic programming introduced by Bellman. Dynamic programming led to the celebrated Hamilton-Jacobi Bellman (HJB) partial differential equation, which provides a sufficient condition for optimality.

It is well known that for the Linear Quadratic Regulator (LQR) problem [1], the corresponding HJB equation becomes a Riccati differential equation, which can be solved very efficiently. For a plant with unknown parameters, a linear quadratic (LQ) controller can be derived by Adaptive Linear Quadratic Control (ALQC), where a parameter adaptation is used [19]. Based on the *Certainty Equivalence Principle*, by replacing unknown parameters with estimated ones in the original Riccati equation, a control law can be derived by solving this new linear quadratic control problem at each time t.

However, as discussed in [13, 14], when it comes to general nonlinear systems, or even the systems with uncertain nonlinearities, it is extremely difficult to obtain the optimal controllers. In [13, 14], a locally weighted learning observer (LWLO) is used to estimate the unknown nonlinear system and then the original problem is transformed into a point-wise min-norm form. Many forms of on-line approximation based control have been suggested for unknown nonlinear systems [3, 7, 8, 10–12, 15–18, 20, 22–29, 31–35]. For the on-line approximation, the distribution of the learning regions can be preassigned, or be allocated automatically during the operation, i.e., be self-organized [16, 35]. A survey of on-line approximation based control and self-organization can be found in [11].

In this paper, we consider the optimal tracking control of uncertain nonlinear systems. A self-organized locally linear optimal tracking control method is proposed, without solving the Hamilton-Jacobi-Bellman equation explicitly. While the learning regions are preassigned in the methods proposed by [13, 14], the main contribution of this paper is that self-organizing on-line approximation is used to solve the optimal tracking control problem for unknown nonlinear systems. Following the idea proposed in [11, 16, 35], the structure self-organization is derived within the Lyapunov context which provides a theoretical basis for a *performance-based* self-organization approach. In this approach, along the system operation trajectory small learning regions are only added if necessary to achieve the performance objective. In each of these small regions, a local affine approximation is used for the unknown nonlinearity. Thus, this control problem is transformed into solving linear system optimal control problems in each local region. To show effectiveness of the proposed optimal controller, a numeric example is presented.

II. PROBLEM STATEMENT

Consider the system

$$\dot{x}_i = x_{i+1}, \qquad 1 \le i \le n-1 \tag{1}$$

$$\dot{x}_n = f(x) + u, \tag{2}$$

where $x(t) = [x_1(t), \ldots, x_n(t)]^\top : \mathbb{R}^+ \mapsto \mathbb{R}^n$ is the system state vector which is assumed to be measured and available, and $u \in \mathbb{R}$ is the control signal. The function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is unknown, nonlinear and assumed Lipschitz continuous with respect to x. Eqns. (1-2) can be written as

$$\dot{x} = Ax + B\big(u + f(x)\big), \tag{3}$$

where (A, B) is in controllable canonical form [9].

There is a desired bounded trajectory $x^d(t) = [x_1^d(t), x_2^d(t), \ldots, x_n^d(t)]^\top : \mathbb{R}^+ \mapsto \mathbb{R}^n$ which satisfies

$$\dot{x}_1^d = x_2^d, \quad \dot{x}_2^d = x_3^d, \quad \cdots, \quad \dot{x}_{n-1}^d = x_n^d. \tag{4}$$

Furthermore, $x^d(t)$ and the derivative $\dot{x}_n^d(t)$ are always available during the control process.

Yiming Chen and Jay A. Farrell are with Department of Electrical Engineering, University of California, Riverside, CA 92521, USA (e-mail: yichen, farrell@ee.ucr.edu).

Wenjie Dong is with the Department of Electrical Engineering, University of Texas-Pan American, Edinburg, TX 78539 USA (e-mail: dongw@utpa.edu).

Acknowledgements: This material is based upon work supported by the National Science Foundation under Grant No. ECCS-0701791. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

978-1-61284-799-3/11/$26.00 ©2011 IEEE

Control Objective: The control objective is to design an optimal controller to ensure that the following cost function

$$J(u)=\frac{1}{2}\int_{0}^{\infty}\big[e^{\top}Qe+\frac{u^{2}}{R}\big]dt \tag{5}$$

achieves its minimum. where

$$\begin{aligned}&e(t)=x(t)-x^{d}(t):\mathbb{R}^{+}\mapsto\mathbb{R}^{n}, \\ &Q\geq 0 \text{ is an } n\times n \text{ matrix, and } R\in\mathbb{R}^{+}.\end{aligned} \tag{6}$$

The optimal control problem can be stated as: *minimize the function $J(u)$ in eqn. (5) by choosing the optimal control policy u^*, subject to the constraint defined by eqn. (3).*

Equivalently, a standard dynamic programming argument could reduce this optimization problem to finding the value function J^* solving the Hamilton-Jacobi-Bellman (HJB) partial differential equation [6],

$$J_x^{*}f_e-\Big(J_x^{*}BB^{\top}J_x^{*\top}\Big)+e^{\top}Qe=0 \tag{7}$$

where J_x^* denotes $\frac{\partial J^*}{\partial x}$, $f_e=[x_2,x_3,\cdots,x_n,f(x)]^{\top}$. If there exists a continuously differentiable positive definite solution to eqn. (7), then the optimal controller is

$$u^{*}=-RB^{\top}J_x^{*\top}.$$

However, it is extremely difficult to solve the HJB equation for a nonlinear optimal control problem, even when the nonlinearity $f(x)$ is known. In this paper, we proposed a Locally Linear Optimal Control (LLOC) approach to derive an approximately optimal controller for the unknown nonlinear system, without solving the Hamilton-Jacobi-Bellman equation explicitly.

III. LLOC DESIGN

The main concept of the LLOC method is straightforward. By detecting the effect of the nonlinearity, small learning regions $\mathcal{S}_k\subset\mathbb{R}^n$ are assigned on-line along the system operation trajectory where they are necessary to satisfy the performance requirement. In each of these small zones, an affine approximation is used to approximate the unknown function through parameter adaption. Thus, on each $\mathcal{S}_k$, we solve the corresponding linear quadratic control problem, to get an optimal control law.

First, herein we introduce some useful notation:

$$\begin{aligned}\mathcal{D}\subset\mathbb{R}^{n} &= \text{the system compact operational region;}\\ \mathcal{N}(t) &= \text{the total number of regions } S_k \text{ at time } t;\\ \mathcal{A}^{\mathcal{N}(t)}\subset\mathbb{R}^{n} &= \bigcup_{1\leq k\leq\mathcal{N}(t)}\mathcal{S}_k,\ \text{the learning region;}\\ \mathcal{D}-\mathcal{A}^{\mathcal{N}(t)} &= \text{the operational region without learning.}\end{aligned}$$

From eqn. (3) and (6), it follows that

$$\dot{e}=Ae+B\big(u+h(x,\dot{x}_n^d)\big), \tag{8}$$

where $h(x,\dot{x}_n^d)=f(x)-\dot{x}_n^d$ is the unknown function with the nonlinearity $f(x)$.

Let us first consider an ideal situation that the unknown function $h(x,\dot{x}_n^d)$ equals zero (i.e. there is no nonlinearity). Then the system (8) can be treated as a Linear Time-Invariant (LTI) system and consequently this control problem becomes a Linear Quadratic Regulation (LQR) problem, of which the solution is well known [2,5,21]. Since Linear Quadratic Regulator is very robust [36], the following controller is derived.

A. Analysis for $|h(x,\dot{x}_n^d)|\leq\epsilon^*$

If $|h(x,\dot{x}_n^d)|\leq\epsilon^*$, where $\epsilon^*>0$ is a criterion selected by designer, then the system (8) can be treated as a linear time-invariant (LTI) system with a small disturbance $d(t)=h\big(x(t),\dot{x}_n^d(t)\big)$ having magnitude less than ϵ^*. Consider the cost function defined in eqn. (5). Since (A,B) is controllable, a control policy can be derived by solving the following optimal control problem

Problem 1:

$$\begin{aligned}\min_{u}\quad &\frac{1}{2}\int_{t_0}^{\infty}\big[e(t)^{\top}Qe(t)+\frac{u^2(t)}{R}\big]dt,\\ \text{s.t.}\quad &\dot{e}(t)=Ae(t)+B\big(u(t)+d(t)\big),\ \text{and } |d(t)|\leq\epsilon^*.\end{aligned}$$

This LQR problem has the solution [2]

$$u_0^*(t)=-RB^{\top}Pe(t). \tag{9}$$

where P is a (constant) $n\times n$ positive definite matrix which is the solution of the *algebraic Riccati equation*,

$$PA+A^{\top}P+Q-RPBB^{\top}P=0.$$

In this case, the optimal trajectory is the solution of the linear time-invariant homogeneous system

$$\dot{e}(t)=Ge(t)+Bd(t), \tag{10}$$

where $G=A-BRB^{\top}P$ and $e(t_0)=x(t_0)-x^d(t_0)$. Regarding the stability of the closed-loop system (10), we have the following lemma.

Lemma 1: If $G=A-BRB^{\top}P$, then the eigenvalues of G will have negative real parts; therefore the stability of the corresponding closed-loop system is guaranteed.

Proof: This result is proved as *Lemma* 9-8 in [2]. ■

Remark 1: For $|h(x,\dot{x}_n^d)|\leq\epsilon^*$, with $d(t)=h(x(t),\dot{x}_n^d(t))$ treated as a small disturbance, the closed-loop system (10) is BIBO stable. △

B. Structure Adaptation

In this part, we consider when and how to assign new learning regions $\mathcal{S}_k$ along the system trajectory.

The controller is initiated with the assumption that $|h(x,\dot{x}_n^d)|\leq\epsilon^*$ and $\mathcal{N}(0)=0$; therefore, the set $\mathcal{A}^0$ is initially empty.

When $x\in\mathcal{D}-\mathcal{A}^{\mathcal{N}(t)}$, the control law (9) is used. Then, from eqns. (3) and (9), it follows that

$$\dot{x}=Gx+B\big(RB^{\top}Px^d+f(x)\big). \tag{11}$$

A state estimator is defined as

$$\dot{\hat{x}}=G\hat{x}+B\big(RB^{\top}Px^d+\dot{x}_n^d\big). \tag{12}$$

From eqn. (11) and (12), it follows that

$$\dot{\tilde{x}}=G\tilde{x}-Bh(x,\dot{x}_n^d), \tag{13}$$

where $\tilde{x} = \hat{x} - x$.

Consider the Lyapunov function $V_0(t) = \tilde{x}^\top N \tilde{x}$, where $N = N^\top > 0$ is the solution of the algebraic Lyapunov equation $G^\top N + NG = -M$ and $M > 0$ is selected by the designer. The time derivative of $V_0(t)$ along solutions of (13) is

$$\dot{V}_0 = -\tilde{x}^\top M \tilde{x} - 2\tilde{x}^\top N B h(x, \dot{x}_n^d). \quad (14)$$

For all x such that $|h(x, \dot{x}_n^d)| \leq \epsilon^*$, when $\tilde{x}^\top(t) M \tilde{x}(t) > 2|B^\top N \tilde{x}(t)|\epsilon^*$, it will be ensured that $\dot{V}_0 \leq 0$.

If V_0 increases while $\tilde{x}^\top(t) M \tilde{x}(t) > |2B^\top N \tilde{x}(t)|\epsilon^*$, then it must be true that $|h(x, \dot{x}_n^d)| > \epsilon^*$. Therefore, the Lyapunov function V_0 provides a mechanism to detect those locations along the trajectory where $|h(x, \dot{x}_n^d)| > \epsilon^*$.

We define the following criteria for adding a new local learning region $\mathcal{S}_k$ to the approximation structure. A local approximator $\hat{f}_k$ is added and $\mathcal{N}(t)$ is increased by one:

1) if the present operating point $x(t)$ is not in any of the existing local learning regions; and
2) $\dot{V}_0(t) \geq 0$ while $\tilde{x}^\top(t) M \tilde{x}(t) > |2B^\top N \tilde{x}(t)|\epsilon^*$.

With the above criteria, $\mathcal{N}(t)$ is non-decreasing. The distribution of regions $\mathcal{S}_k$ and the region $\mathcal{A}^{\mathcal{N}(t)}$ change as $N(t)$ increases.

For $k \geq 1$, we denote the time at which the k-th local learning region is added as t_k (i.e., $\mathcal{N}(t_k) = k$ and $\lim_{\epsilon \to 0} \mathcal{N}(t_k - \epsilon) = k - 1$). With this notation, $\mathcal{N}(t)$ is constant with value k for $t \in [t_k, t_{k+1})$. It is possible that for some k, the approximator has sufficient approximation capability, in which case $t_{k+1} = \infty$. The center location of the new local approximator is denoted as c_k. At $t = t_k$, when the k-th node is added, it is the case that $x(t) \notin \mathcal{S}_i$ for $i = 1, ..., \mathcal{N}(t_k) - 1$. The center location $c_k \in \mathcal{D}$ should be selected such that $x(t_k) \in \mathcal{S}_k$, $c_k \notin \mathcal{S}_i$ and $c_i \notin \mathcal{S}_k$, for $1 \leq i < \mathcal{N}(t_k)$. Furthermore, the size of $\mathcal{S}_k$ can be specified by the radius $r = [r_1, ..., r_n] \in \mathbb{R}^n$ such that

$$|x_i - c_{ki}| \leq r_i, \ i = 1, \ldots, n, \text{ for all } x \in \mathcal{S}_k.$$

Remark 2: Usually, the Lyapunov function is used only for stability analysis. Herein, V_0 is computed and used online for structure self-organizing. In practical applications of the method proposed here, the value of the Lyapunov function V_0 would be computed at every sample time, when $x \in \mathcal{D} - \mathcal{A}^{\mathcal{N}(t)}$. If it is increasing in consecutive time instants, then it implies that $\dot{V}_0(t) \geq 0$. △

C. For $x(t) \in \mathcal{A}^{\mathcal{N}(t)}$

When $x(t) \in \mathcal{A}^{N(t)}$, it is the case that $x(t) \in \mathcal{S}_k$'s, for at least one $k \in [1, N(t)]$. Self-organized function approximation is proposed to deal with the effect of the uncertainty. Based on the assumption about the Lipschitz continuity of $f(x)$ and definition of $h(x, \dot{x}_n^d)$, on each of the learning regions, there exists a local affine approximation

$$h_k(x, \dot{x}_n^d) = \theta_k^\top \phi_k(x) - \dot{x}_n^d, \quad (15)$$

such that $h(x, \dot{x}_n^d) = h_k(x, \dot{x}_n^d) + \delta_{h_k}(x, \dot{x}_n^d)$. Herein we assume that the radius of S_k is small enough such that the *inherent appoximation error* $|\delta_{h_k}(x, \dot{x}_n^d)| \leq \epsilon^*$ for all $x \in \mathcal{S}_k$ and $\dot{x}_n^d$.

In eqn. (15), $\theta_k = [\theta_{k_1}, \ \theta_{k_2}, \cdots, \theta_{k_{n+1}}]^\top \in \mathbb{R}^{n+1}$ are the weights and $\phi_k(x) = [x - c_k, \ 1]^\top \in \mathbb{R}^{n+1}$ is the basis for the function approximation, where $c_k \in \mathbb{R}^n$ is the center of the region $\mathcal{S}_k$.

To realize locally weighted learning, we introduce a continuous, non-negative and locally supported weighting function $\bar{\omega}_k(x)$, which is specified in [11, 16, 35], for the k-th local approximator, with $k \in 1, \ldots, N(t)$.

Then, we have the *ideal* locally weighted approximation

$$h(x, \dot{x}_n^d) = \sum_k \bar{\omega}_k h_k(x, \dot{x}_n^d) + \delta_h(x, \dot{x}_n^d),$$

where $|\delta_h(x, \dot{x}_n^d)| \leq \epsilon^*$ for all $x \in \mathcal{A}^{\mathcal{N}(t)}$. Thus, in $\mathcal{A}^{\mathcal{N}(t)}$, the system dynamics becomes

$$\dot{x} = Ax + B\big(u + \sum_k \bar{\omega}_k \theta_k^\top \phi_k + \delta_h\big). \quad (16)$$

To approximate $h(x, \dot{x}_n^d)$ in $\mathcal{A}^{\mathcal{N}(t)}$, we introduce the following state estimator,

$$\dot{\hat{x}} = A\hat{x} + B\big(u + \sum_k \bar{\omega}_k \hat{\theta}_k^\top \phi_k - v\big), \quad (17)$$

where $\hat{\theta}_k$ is the estimated value of θ_k and v is the estimation error stabilizing signal defined later. From eqn. (16) and (17), we have

$$\dot{\tilde{x}} = A\tilde{x} + B\big(\sum_k \bar{\omega}_k \tilde{\theta}_k^\top \phi_k - \delta_h - v\big), \quad (18)$$

where $\tilde{x} = \hat{x} - x$ and $\tilde{\theta}_k = \hat{\theta}_k - \theta_k$.

Let $z(t) = L^\top \tilde{x}(t)$, where

$$\begin{aligned} L &= [l_1, l_2, \ldots, l_{n-1}, l_n]^\top \\ &= [\lambda^{n-1}, C_{n-1}^1 \lambda^{n-2}, \ldots, C_{n-1}^{n-2}\lambda, 1]^\top, \end{aligned} \quad (19)$$

λ is a positive constant, and $C_n^m = \frac{n!}{m!(n-m)!}$ for $1 \leq m \leq n-2$. This definition ensures the stability analysis in the proof of *Lemma* 3.

Lemma 2: ([35]) If $\lim_{t\to\infty} |z(t)| \leq \mu_e$, then $\lim_{t\to\infty} |\tilde{x}_i| \leq \frac{2^{i-1}}{\lambda^{n-i}} \mu_e$ for $1 \leq i \leq n$, where μ_e is a positive constant. Furthermore, if $\lim_{t\to\infty} z(t) = 0$, then $\lim_{t\to\infty} \tilde{x}_i = 0$ for $1 \leq i \leq n$.

Note that $\hat{x}$ asymptotically converging to x is equivalent to that $\tilde{x}$ converging to zero. By Lemma 2, $\tilde{x}$ will converge to zero if $z(t)$ converges to zero; therefore, it is sufficient to choose v and $\hat{f}_k$ to ensure that $z(t)$ converges to zero.

With the notation above, the adaptive law is defined as

$$\dot{\hat{\theta}}_k = \begin{cases} -z\Gamma\phi_k(x), & \text{if } x \in \mathcal{S}_k \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

where Γ is a symmetric positive definite adaptive gain matrix. This definition ensures the stability analysis in the proof of *Lemma* 3.

The stabilizing signal v is chosen as [13, 14]

$$v = l_1\tilde{x}_2 + ... + l_{n-1}\tilde{x}_n + Fz + \epsilon^* sign(z), \quad (21)$$

where L is defined in eqn. (19) and F is a positive constant. Then we have the following lemma.

Remark 3: For implementation, it is impractical to apply the stabilizing signal (21) with the $sign(\cdot)$ term. Instead, in application, that term can be replaced by a $sat(\cdot)$ term or just neglected by the designer. The analysis for these alternative solutions is trivial, but for this paper, the space is too limited to show it. △

Lemma 3: For system (17) with the adaptive law (20) and the stabilizing signal (21), we have that

$$\lim_{t\to\infty}(\hat{x}-x)=0.$$

Proof: This lemma can be proved by following the same logic in the proof of $Lemma$ 2 of [14]. ■

With the aid of *Lemma* 3, optimal controllers can be derived for the system dynamics (17) to solve the corresponding optimal tracking problems. Since $\hat{x}$ converges asymptotically to x, x will converge to zero when $\hat{x}$ does.

For convenience of later analysis, we introduce some useful notation here. With $\hat{\theta}_k$, it follows that

$$\begin{aligned}\hat{h}(x,\dot{x}_n^d) &= \sum_k \bar{\omega}_k \hat{h}_k(x,\dot{x}_n^d) = \sum_k \bar{\omega}_k \hat{\theta}_k^\top \phi_k(x) - \dot{x}_n^d \\ &= \Theta^\top x + \sum_k \bar{\omega}_k \hat{\theta}_k^\top \begin{bmatrix} -c_k \\ 1 \end{bmatrix} - \dot{x}_n^d \end{aligned} \tag{22}$$

where

$$\begin{aligned}\Theta &= \sum_k \bar{\omega}_k \bar{\theta}_k = [\Theta_1,\ \cdots,\Theta_n], \\ \bar{\theta}_k &= [\hat{\theta}_{k1},\ \hat{\theta}_{k2},\cdots,\hat{\theta}_{kn}]^\top,\ \text{and} \\ \hat{\theta}_k &= [\hat{\theta}_{k1},\ \hat{\theta}_{k2},\cdots,\hat{\theta}_{kn},\hat{\theta}_{k(n+1)}]^\top.\end{aligned}$$

From eqns. (17) and (22), with $\hat{e}(t)=\hat{x}-x^d$, it follows that

$$\begin{aligned}\dot{\hat{e}} &= A\hat{e} + B\Big(u + \Theta^\top x + \sum_k \bar{\omega}_k \hat{\theta}_k^\top \begin{bmatrix} -c_k \\ 1 \end{bmatrix} - \dot{x}_n^d - v\Big) \\ &= \big(A+B\Theta^\top\big)\hat{e} + B\Big(u + \Theta^\top x^d - \Theta^\top \tilde{x} \\ &\qquad + \sum_k \bar{\omega}_k \hat{\theta}_k^\top \begin{bmatrix} -c_k \\ 1 \end{bmatrix} - \dot{x}_n^d - v\Big) \end{aligned} \tag{23}$$

Based on eqn. (23), a control policy can be derived by solving the following optimal control problem,

Problem 2:

$$\begin{aligned}\min_u \quad & \frac{1}{2}\int_{\bar{t}}^{\infty} \big[\hat{e}(t)^\top Q \hat{e}(t) + \frac{u^2(t)}{R}\big]dt, \\ \text{s.t.} \quad & \dot{\hat{e}}(t) = \bar{A}(t)\hat{e}(t) + B\big(u(t)+\zeta(t)\big).\end{aligned} \tag{24}$$

where

$$\begin{aligned}\hat{e} &= \hat{x}-x^d, \\ \bar{t} & \text{ is the current sample time}, \\ \bar{A} &= A + B\Theta^\top, \\ \zeta &= \Theta^\top x^d - \Theta^\top \tilde{x} + \sum_k \bar{\omega}_k \hat{\theta}_k^\top \begin{bmatrix} -c_k \\ 1 \end{bmatrix} - \dot{x}_n^d - v.\end{aligned} \tag{25}$$

To solve the optimal regulation problem above, we first consider the controllability of the system. From eqn. (25), it follows that

$$\bar{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \Theta_1 & \Theta_2 & \Theta_3 & \dots & \Theta_n \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

which is in controllable canonical form at every time t. From $Theorem$ 6.12 in [9] for time-varying systems, the controllability is always ensured.

To solve this problem, based on the ALQC idea and the Certainty Equivalence Principle [19], we have the following remark on solving *Problem* 2.

Remark 4: At every time $\bar{t}$, $\hat{\theta}_k$ is updated by the adaptive law (20) before solving Problem 2. Then, the parameters $\bar{A}$ and ζ are both considered as constant when solving Problem 2. Because this is the best knowledge of the local affine approximation to the unknown nonlinearity $f(x)$ that the controller could have at every time t. △

When the parameters $\bar{A}$ and ζ are all treated as constants, a control policy can be derived for this infinite horizon problem by following the same logic provided in Section 9-10, Chapter 9 of [2] about the solution of optimal tracking control of linear systems.

Through minimizing the *Hamiltonian* of Problem 2

$$H = \frac{1}{2}\hat{e}^\top Q\hat{e} + \frac{u^2}{2R} + \lambda^\top \bar{A}\hat{e} + \lambda^\top B(u+\zeta),$$

we have that the control law for the learning region is

$$\bar{u} = RB^\top\Big((\bar{G}^\top)^{-1}\bar{P}B\zeta - \bar{P}\hat{e}\Big), \tag{26}$$

where $\bar{G} = \bar{A} - BRB^\top\bar{P}$ and $\bar{P}$ is the positive definite solution of the algebraic Riccati equation

$$\bar{P}\bar{A} + \bar{A}^\top\bar{P} + Q - R\bar{P}BB^\top\bar{P} = 0. \tag{27}$$

Thus, the closed-loop system is

$$\dot{\hat{e}} = \bar{G}\hat{e} + B\Big(RB^\top(\bar{G}^\top)^{-1}\bar{P}B\zeta + \zeta\Big). \tag{28}$$

From *Lemma* 1, it follows that system (28) is stable and then $(\bar{G}^\top)^{-1}$ always exists.

D. Analysis

Based on the solution of Problem 1 and 2, the control policy for system (3) is concluded as,

$$u(t) = \begin{cases} -RB^\top Pe(t), & x\in\mathcal{D}-\mathcal{A}^{\mathcal{N}(t)} \\ \bar{u}, & x\in\mathcal{A}^{\mathcal{N}(t)}. \end{cases} \tag{29}$$

with the state observer

$$\dot{\hat{x}}(t) = \begin{cases} G\hat{x} + B\big(RB^\top Px^d + \dot{x}_n^d\big), & x\in\mathcal{D}-\mathcal{A}^{\mathcal{N}(t)} \\ A\hat{x} + B\big(u + \sum_k \bar{\omega}_k\hat{\theta}_k^\top\phi_k - v\big). & x\in\mathcal{A}^{\mathcal{N}(t)}, \end{cases}$$

In eqn. (29), $\bar{u}$ is defined by eqn. (26).

With all the analysis above, we have the following result.

Theorem 1: For system (3) with the state estimator (17) and the adaptive law (20), the optimal control (29) solves Problem 1 and Problem 2, and makes x track x^d while the number of allocated learning regions $N(t)$ is finite $\forall t \geq 0$.

Proof: For $x \in \mathcal{D} - \mathcal{A}^{\mathcal{N}(t)}$, with *Lemma* 3 and the analysis in Subsection III-A, the control law (29) solves Problem 1 and makes x track x^d.

For $x \in \mathcal{S}_k$, from the analysis in Subsection III-C, the control law (29) solves Problem 2, i.e. makes $\hat{x}$ track x^d. Then with *Lemma* 3, it is proved that x also tracks x^d.

The proof of that $N(t)$ is finite can follow the same idea of the proof of $Theorem$ 1 in [11]. ■

IV. Numeric Example

In this section, a numerical simulation is presented to illustrate the control method proposed in the previous sections. Consider the following nonlinear system

$$\dot{x} = Ax + B\big(u + 0.5|x_1|\big),$$

where

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We used two desired trajectories : (1) x_1^d is a square wave, with magnitude 1.5 and a 14-second period while $x_2^d = 0$; (2) $x_1^d \in [-2, 2]$ is a uniform random number while $x_2^d = 0$. The control objective is to design a self-organized locally linear optimal controller proposed in this paper to realize the optimal tracking, i.e., the tracking error $e(t)$ converges to zero asymptotically while minimizing the cost function

$$J(u) = \frac{1}{2}\int_0^\infty \left[e^\top Qe + \frac{u^2}{R}\right]dt,$$

where we choose

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 10.$$

For the function approximation, we use the learning regions $\mathcal{S}_k$ without any overlap between them. The weighting function is defined as

$$\bar{\omega}_k(x) = \begin{cases} 1, & x \in \mathcal{S}_k \\ 0, & \text{otherwise.} \end{cases}$$

And we specify that

$$\begin{aligned} \text{the adaptation gain } \Gamma &= 10 \times I_{3\times 3}, \\ \text{the approximation accuracy } \epsilon^* &= 0.05, \\ \text{the radius of local learning regions } r &= 0.1. \end{aligned}$$

The function approximation is initialized with $N(0) = 0$ and $\hat{\theta}_k = 0$. In this problem the unknown nonlinear function is $h(x, \dot{x}_n^d) = 0.5|x_1(t)|$. Therefore, the self-organized approximator should not allocate centers in the region $-0.1 \leq x_1 \leq 0.1$, since the prespecified accuracy requirement $\epsilon^* = 0.05$ is satisfied in this region, even without any function approximation.

The simulation results are shown in Fig. 1-3. In Fig. 1, it can be observed that the curve in the second period is

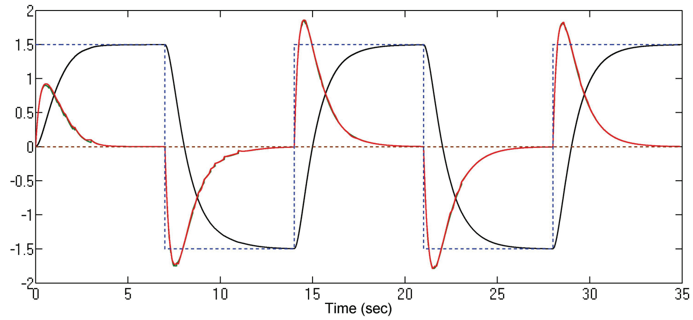

Fig. 1. Performance of Self-organized Locally Linear Optimal Tracking Controller for $t \in [0,\ 35]s$ when $x_1^d(t)$ is the square wave: $x_1^d(t)$ (blue, dashed), $x_2^d(t)$ (brown, dashed), $x_1(t)$ (black, solid), $x_2(t)$ (red, solid), $\hat{x}_2(t)$ (green, solid). Initial state is (0,0).

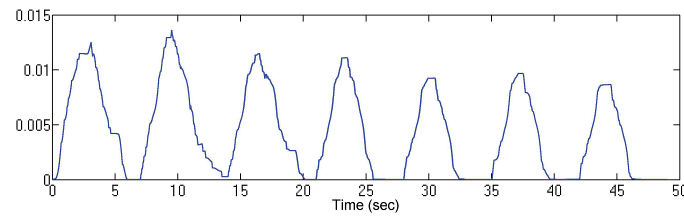

Fig. 2. Performance of Self-organized Locally Linear Optimal Tracking Controller for $t \in [0,\ 49]s$ when $x_1^d(t)$ is the square wave: $\|\tilde{x}(t)\|$ (blue, solid), which is filtered by a moving average filter.

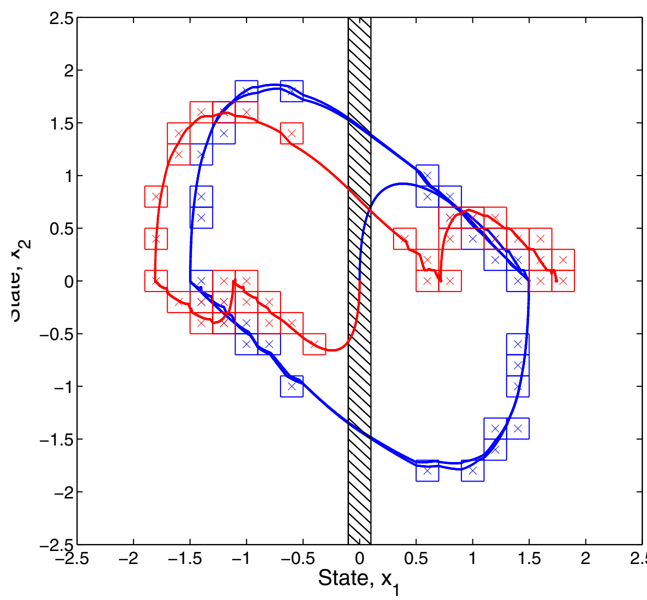

Fig. 3. Phase plane plot of x_1 versus x_2 for $t \in [0,\ 35]s$: blue for that $x_1^d(t)$ is the square wave while red for the uniform random number. The $\times' s$ indicate the assigned center locations and the small square around each center location represents the associated region of support. Initial states are (0,0).

smoother than that in the first one, since the accuracy of local affine approximations is improved through learning. Fig. 2 presents the time history of estimation error $\|\tilde{x}\|$ for $t \in [0, 49]s$ when the square wave is desired. To analyze the result more efficiently, $\|\tilde{x}\|$ is filtered by a moving average filter. It can be observed that after steps occur at every half-period, during the first several seconds $\tilde{x}$ keeps increasing until touches each peak. On one hand, it is because that at the beginning the trajectory crosses learning regions very fast and then in each S_k the controller has limited time to improve the function approximation. On the other hand, when the trajectory comes around the zone where $|f(x)| < \epsilon^*$, the criteria for structure adaptation are not satisfied until $\|\tilde{x}\|$ becomes large enough and S_k is activated. Then, with the function approximation in the learning region, $\tilde{x}$ decreases rapidly and converges to zero. Compare the curves in three periods $[7, 21)s$, $[21, 35)s$ and $[35, 49)s$, it is shown that the heights of the peaks in same interval of each period become smaller and smaller, which means that the accuracy of function approximation is improved through learning. By Fig. 3, it is demonstrated that small local learning regions were assigned along the trajectory. The controller does not allocate any centers between $x_1 = -0.1$ and $x_1 = 0.1$, where $|h(x, \dot{x}_n^d)| < \epsilon^*$. This zone was unknown to the controller before the control process.

V. CONCLUSION AND FUTURE WORK

This paper considers optimal control of uncertain nonlinear systems using a self-organizing locally linear method. Small learning regions are self-organized online along the system trajectory. With locally weighted learning, a state estimator is introduced to estimate the parameters for local approximators. Optimal controllers are proposed for the estimated piecewise linear models and the stability is guaranteed. Numerical simulations show the efficiency of this approach. Future work may extend this approach to more general affine models like $\dot{x}_n = f(x) + g(x)u$. In this paper, only the situation where $g(x) = 1$ has been considered.

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[2] M. Athans and P. L. Falb, *Optimal Control: An Introduction to the Theory and its Applications*. New York: McGraw-Hill, 1966.

[3] C. G. Atkeson, A. W. Moore, and S. Schaal, “Locally weighted learning for control,” *Artifical Intelligence Review*, vol. 11, pp. 75–113, 1996.

[4] R. Bellman, “The theory of dynamic programming,” *Proc. of Nat. Acad. Sci. USA*, vol. 38, 1952.

[5] A. E. Bryson and Y. C. Ho, *Applied Optimal Control: Optimization, Estimation and Control*. Waltham, MA: Ginn Company, 1969.

[6] ——, *Applied Optimal Control: Optimization, Estimation and Control*. Hemisphere Publishing Corporation, 1975.

[7] A. J. Calise, N. Hovakimyan, and M. Idan, “Adaptive output feedback control of nonlinear systems using neural networks,” *Automatica*, vol. 37, no. 1, pp. 1201–1211, 2001.

[8] M. Cannon and J.-J. Slotine, “Space-frequency localized basis function networks for nonlinear system estimation and control,” *Neurocomputing*, vol. 9, pp. 293–342, 1995.

[9] C.-T. Chen, *Linear System Theory and Design*, 3rd ed., 1999.

[10] F.-C. Chen and H. K. Khalil, “Adaptive control of a class of nonlinear discrete-time systems using neural networks,” *IEEE Trans. on Automatic Control*, vol. 40, no. 5, pp. 791–801, 1995.

[11] Y. Chen and J. A. Farrell, “Self-organizing approximation based control with $\mathcal{L}_1$ transient performance guarantees,” *Proc. of American Control Conference*, 2011.

[12] J. Y. Choi and J. A. Farrell, “Nonlinear adaptive control using networks of piecewise linear approximators,” *IEEE Trans. on Neural Networks*, vol. 11, no. 2, pp. 390–401, 2000.

[13] W. Dong and J. A. Farrell, “Optimal regulation of unknown nonlinear systems based on locally weighted learning,” in *Intelligent Control, 2008. ISIC 2008. IEEE International Symposium on*, pp. 1079–1084.

[14] ——, “Optimal tracking control for unknown nonlinear systems based on locally weighted learning,” in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pp. 2148–2153.

[15] J. A. Farrell and M. M. Polycarpou, *Adaptive Approximation Based Control: Unifying Neural, Fuzzy, and Traditional Adaptive Approximation Approaches*. NJ: John Wiley & Sons, Hoboken, 2006.

[16] J. A. Farrell and Y. Zhao, “Self-organizing approximation based control,” *Proc. of American Control Conference*, pp. 3378–3384, 2006.

[17] S. S. Ge and J. Zhang, “Neural-network control of nonaffine nonlinear system with zero dynamics by state and output feedback,” *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 900–918, 2003.

[18] J. Hu and G. Pratt, “Self-organizating CMAC neural networks and adaptive dynamic control,” 1999, pp. 259–265.

[19] P. Ioannou and J. Sun, *Robust Adaptive Control*. Prentice Hall, 1996.

[20] J.-P. Jiang and L. Praly, “Design of robust adaptive controllers for nonlinear systems with dynamic uncertainties,” *Automatica*, vol. 34, no. 7, pp. 825–840, 1998.

[21] D. E. Kirk, *Optimal Control Theory: An Introduction*. Prentice-Hall, 1970.

[22] E. Lavretsky and N. Hovakimyann, “Adaptive dynamic inversion for nonaffine-in-control systems via time-scale separation: part ii,” *Proceedings of America Control Conference*, pp. 3548–553, 2005.

[23] F. L. Lewis, K. Liu, and Y. A, “Neual net robot control with guaranteed tracking performance,” *IEEE Trans. on Neural Networks*, vol. 6, no. 3, pp. 703–715, 1995.

[24] F. L. Lewis, A. Yesildirek, and K. Liu, “Multilayer neural-net robot controller with guaranteed tracking performance,” *IEEE Trans. on Neural Networks*, vol. 7, no. 2, pp. 388–399, 1996.

[25] J. Lian, Y. Lee, S. D. Sudhoff, and S. H. Zak, “Variable structure neural network based direct adaptive robust control of uncertain systems,” in *Proceedings of the American Control Conference*, 2008, pp. 3402–3407.

[26] C.-M. Lin and T.-Y. Chen, “Self-organizing cmac control for a class of MIMO uncertain nonlinear systems,” *Neural Networks, IEEE Transactions on*, vol. 20, no. 9, pp. 1377 –1384, 2009.

[27] J. Nakanishi, J. A. Farrell, and S. Schaal, “Composite adaptive control with locally weighted statistical learning,” *Neural Networks*, vol. 18, no. 1, pp. 71–90, 2005.

[28] J. Nakanishi, J. A. Farrell, and S. Schall, “A locally weighted learning composite adaptive controller with structure adaptation,” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 882–889, 2002.

[29] J.-H. Park, S.-H. Huh, S.-H. Kim, S.-J. Seo, and G.-T. Park, “Direct adaptive controller for nonaffine nonlinear systems using self-structuring neural networks,” *IEEE Trans. on Neural Networks*, vol. 16, no. 2, pp. 414–422, 2005.

[30] L. S. Pontryagin, “Optimal control processes,” *Uspehi Mat. Nauk (in Russian),*, vol. 14, pp. 3–20, 1959.

[31] R. M. Sanner and J. J. Slotine, “Gaussian networks for direct adaptive control,” *IEEE Trans. on Neural Networks*, vol. 3, pp. 837–863, 1992.

[32] S. Schaal and C. G. Atkeson, “Constructive incremental learning from only local information,” *Neural Computation*, vol. 10, no. 8, pp. 2047–2084, 1998.

[33] D. Vrabie and F. Lewis, “Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems,” *Neural Networks*, vol. 22, no. 3, pp. 237–246, 2009.

[34] Y. Zhao and J. A. Farrell, “Performance-based self-organizing approximation for scalar state estimation and control,” *Proc. of American Control Conference*, 2007.

[35] ——, “Self-organizing approximation-based control for higher order systems,” *IEEE Transactions on Neural Networks*, vol. 18, no. 4, pp. 1220–1231, 2007.

[36] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Prentice Hall, 1998.