Optimal Adaptive Control of Nonlinear Continuous-time Systems in Strict Feedback Form with Unknown Internal Dynamics*

H. Zargarzadeh¹, T. Dierks², and S. Jagannathan¹

Abstract— A novel approach is proposed for multi-input multi-output (MIMO) optimal adaptive control of nonlinear continuous-time systems in strict feedback form with uncertain internal dynamics. First, it is shown that the optimal adaptive tracking problem of strict feedback systems can be reduced to an optimal regulation problem of affine nonlinear continuous-time systems expressed as a function of tracking error by designing a properly chosen adaptive feedback control input. Then, an optimal adaptive feedback scheme is introduced for the affine system to estimate the solution of the Hamilton-Jacobi-Bellman (HJB) equation online which becomes the optimal feedback control input for the closed-loop system. A Lyapunov based approach is employed to show that the tracking error converges to zero as well as the cost function estimation and the internal dynamics estimation errors provided the system input is persistently exciting. Finally, numerical results are provided to verify the theoretical results.

I. INTRODUCTION

Stabilization of nonlinear systems is now an established field [1]-[4]. However, a controller should not only stabilize a nonlinear system but also minimize a prescribed performance index [6],[7],[8]. This problem becomes more challenging when the nonlinear system dynamics become partially uncertain [9]. The optimal control of partially uncertain nonlinear systems is not adequately addressed in the literature. The optimal control of nonlinear continuous or discrete-time systems is a challenging task, since usually the solution of the Hamilton-Jacobi-Bellman (HJB) equation (which does not have a closed-form solution) is required.

For the case of infinite horizon optimal control, linear systems require the algebraic Riccati equation (ARE) to be solved which is much more easier when compared to RE whereas the nonlinear systems require the solution of the HJB equation in real-time [10]. The system dynamics have to be known to solve RE, ARE and HJB equations while it becomes impractical when the system dynamics are uncertain. Recently, state dependent Riccati equation (SDRE) [14] is proposed to address optimal control in an iterative and numerical way by assuming the nonlinear system has a linear state dependent representation under certain conditions.

On the other hand, online adaptive approximation-based optimal controller designs referred to as adaptive critic designs (ACD) [11]-[13] by using online approximator (OLAs) have been introduced recently in the literature. In ACDs, dynamic programming (DP) or reinforcement learning (RL) is utilized. The ACD techniques [12] solve the optimal control forward-in-time by finding the solution to the HJB equation in an iterative manner via value (or cost) or policy (control input) iterations and without needing the system dynamics.

By contrast, in [6], a single online approximator-based ACD technique is introduced for continuous-time nonlinear system in affine form by using an initial admissible (stabilizing) control input. Lyapunov stability is included and policy and value iterations are not needed though full knowledge of the system dynamics is needed.

In the adaptive control literature, nonlinear systems in strict feedback form are represented in a variety of ways [5],[16]-[18] and their stability is studied using the standard backstepping scheme [1]-[5] without any optimality. In addition, in a few papers [4],[16]-[18] the control of such unknown strict feedback systems using adaptive neural network (NN)-based schemes is given. More recently, the inverse optimal control of strict feedback systems is introduced in [7] when the dynamics are assumed to be known. In the inverse optimal control method, first the control law is designed, and then the associated cost function is identified. In contrast, we aim to design a control law based on a given cost function.

Therefore, in our previous paper, the optimal control of nonlinear strict feedback continuous-time systems [15] is introduced without using policy iterations when the system dynamics are known. Continuing the work of [15], this paper tackles the optimal control of nonlinear continuous-time systems in strict feedback form with unknown internal dynamics. The optimal control law and cost function are approximated by using linear in the parameter (LIP) structures in a forward-in-time manner. Moreover, an initial admissible controller is not required and policy iterations are not utilized. It is demonstrated that by using a feedforward controller, the optimal tracking problem of the partially unknown strict feedback systems is equivalent to optimally stabilizing an affine system expressed in the tracking error form. Next, the optimal adaptive scheme is developed for affine nonlinear continuous-time systems without needing the internal dynamics and policy or value iterations. Lyapunov theorem is used to prove the overall stability.

The internal dynamics is also being estimated separately first through the use of a state estimator by using linear in the unknown parameters (LIP) assumption. Lyapunov theory is utilized to demonstrate the convergence of the optimal adaptive control scheme for the overall nonlinear system while explicitly considering the solution to the HJB equation (cost function) has a LIP form.

Next the tracking problem for nonlinear continuous-time systems in strict feedback form is introduced followed by the optimal control scheme development.

*Research is supported in part by NSF grants ECCS#0624644 and ECCS# 0901562.

[1] H. Zargarzadeh and S. Jagannathan are with Dept of Electrical and Computer Engineering Missouri University of Science and Technology Rolla, MO 65401, USA (corresponding author is H. Zargarzadeh; e-mail: hzrd47@mail.mst.edu).

[2] T. Dierks is with DRS Sustainment Systems, Inc., 201 Evans Lane, St Louis, MO 63121, USA.
II. THE TRACKING PROBLEM FOR STRICT FEEDBACK SYSTEMS

Consider the multi-input multi-output (MIMO) nonlinear continuous-time system described by
\[
\dot{x}_i = f_i(x_1, ..., x_N) + g_i(x_1, ..., x_N)u \quad \text{for } 1 \leq i \leq N-1 \text{ and } N \geq 2
\]
\[
y = x_1
\]
where each \(x \in \mathbb{R}^n\) denotes a state vector, \(u \in \mathbb{R}^m\) represents the input vector with \(f_i(x_1, ..., x_N) \in \mathbb{R}^n\), and \(g_i(x_1, ..., x_N) \in \mathbb{R}^{m \times n}\) being nonlinear smooth functions. It is assumed that systems (1)-(2) is reachable while its internal dynamics, \(f_i(x_1, ..., x_N)\), can be represented as LIP and it is given by
\[
f_i(x_1, ..., x_N) = \Psi(x)\Lambda_i,
\]
where \(\Psi(x) \in \mathbb{R}^{m \times n}\) and \(\Lambda_i \in \mathbb{R}^{n \times n}\) with \(\Lambda_i\) being unknown. Under the LIP assumption, the internal dynamics are estimated as \(\hat{f}_i(x_1, ..., x_N) = \Psi(x)\hat{\Lambda}_i\), where \(\hat{\Lambda}_i\) is the estimate of the target parameter vector \(\Lambda_i\), and the estimation error is then given by
\[
f_i(x_1, ..., x_N) - \hat{f}_i(x_1, ..., x_N) = \Psi(x)(\Lambda_i - \hat{\Lambda}_i) = \Psi(x)\varepsilon_i.
\]
The update law for \(\hat{\Lambda}_i\) will be designed based on Lyapunov stability analysis. Here for (1), the system state \(x_i\) is treated as the virtual control input. Nonetheless, the overall system (1)-(3) is being controlled through the control input \(u\). The following assumption is needed before we proceed.

Assumption 1. It is assumed that for the known functions \(g_i(x_1, ..., x_N)\), \(\|g_i(x_1, ..., x_N)\| \neq 0 \quad (1 \leq i \leq N)\) belongs to \(\Omega \in \mathbb{R}^n\), and \(\Omega\) is bounded satisfying \(g'_\text{min} \leq g_i(x_1, ..., x_N) \leq g'_\text{max}\), where \(\|\cdot\|\) is the Frobenius norm and \(g'_\text{min}\) and \(g'_\text{max}\) are positive constants.

Under the Assumption 1, the optimal control input for the nonlinear system (1)-(2) can be obtained [8] by using a backstepping approach. In other words, the objective of our scheme is to design an adaptive controller \(\alpha\) in order to have the output \(y\) to track a desired trajectory \(x_d\) in an optimal manner even when the internal dynamics, \(f_i(\cdot)\), are unknown.

To this end, by applying the backstepping approach [5], the system given by (1)-(2) tracks a predesigned trajectory \((x_{d1}, ..., x_{dN})\). Now, we follow the steps in the standard backstepping scheme to design the optimal adaptive scheme for strict-feedback systems. Consider \(J_i(E)\) as any positive definite radially unbounded function of \(E^T = [e_1^T \cdots e_N^T]\), (e.g. a quadratic function is selected in the Appendix), \(J_i(E)\) is the partial derivative of \(J(E)\) with respect to \(e_i\), and \(J_i(E)\) is the partial derivative of \(J_i(E)\) with respect to \(e_i\). Moreover, the tracking errors are defined as \(e_i = x_i - x_{di}\) with \(x_{di}\) being the desired value for \(x_i\) with \(1 \leq i \leq N\) that will be defined separately in the following.

To stabilize the tracking error, \(e_i = x_i - x_{di}\), the backstepping approach will use \(N\) steps [1] which are presented next.

**Step 1:** It is desired that \(x_i\) to follow a smooth desired trajectory \(x_{di}\). The system dynamics in (1) can be rewritten as
\[
\dot{x}_i = \dot{e}_i = -\dot{x}_{di} + f_i(x_1, ..., x_N) + g_i(x_1, ..., x_N)u,
\]
with the assumption that \(f_i(x_1) = \hat{f}_i(x_1)\) is the estimate of \(f_i(x_1)\), define the estimation error as \(\hat{f}_i(x_1) - \hat{f}_i(x_1) = \hat{f}_i(x_1)\). Then
\[
x_i - \dot{x}_{di} = \dot{e}_i = -\dot{x}_{di} + \hat{f}_i(x_1) + \hat{f}_i(x_1) + \hat{f}_i(x_1)
\]
\[
= \hat{f}_i(x_1) + \hat{f}_i(x_1) + \hat{f}_i(x_1),
\]
where virtual control input \(x_{d2}\) is chosen such that \(x_{d2} = \hat{f}_i(x_1)\) with \(x_{d2}\) being the optimal feedback control input and \(x_{d3}\) the feedforward virtual control input. The input \(x_{d2}\) is selected by solving
\[
-\dot{x}_{d2} + \hat{f}_i(x_1) + g_i(x_1)x_{d2} - g_i(x_1)J_{i
u}(E) = 0.
\]

Where, \(f_i(x_1) - f_i(x_1)\), \(\hat{f}_i(x_1) + \hat{f}_i(x_1)\) and \(\hat{f}_i(x_1)\) respectively. In the right hand side (RHS) of (6), the effect of the third and the last terms are cancelled during Lyapunov stability proof. This can be accomplished by choosing a proper desired trajectory (and the corresponding virtual controller) in the next step. Section III is devoted to present the existence of the optimal feedback control input \(x_{d3}\) and its design. Inevitably, \(e_i\) cannot be zero due to dynamics of the second system of (1) and the desired output \(x_i\) trajectory. Since the second to the \((N-1)\) steps are quite similar, we skip to the \(i^{th}\) step.

**Step 2:** In this step, we need an optimal controller for the system (1)-(3) such that \(e_i \to 0\). To this end, the system \(i\) in (1) can be rewritten as
\[
\dot{x}_i = \dot{e}_i = -\dot{x}_{di} + f_i(x_1, ..., x_N) + g_i(x_1, ..., x_N)x_{d(i+1)}
\]
\[
+ g_i(x_1, ..., x_N)(x_{di} - x_{d(i-1)})J_{i
nu}(E) - g_i(x_1, ..., x_N)J_{i
nu}(E) = 0,
\]
where \(x_{di}\) is chosen such that \(x_{d(i+1)} = x_{d(i+1)} + x_{d(i+1)}\), with the virtual control input \(x_{d(i+1)}^*\) satisfying (similar to the step in (7))
\[
-\dot{x}_{d(i+1)} + \hat{f}_i(x_1, ..., x_N) + g_i(x_1, ..., x_N)x_{d(i+1)}^* + \hat{f}_i(x_1, ..., x_N)
\]
\[
= \hat{f}_i(x_1, ..., x_N) - \hat{f}_i(x_1, ..., x_N)
\]
\[
= g_i(x_1, ..., x_N)J_{i
nu}(E) = 0.
\]

As mentioned in the previous step, there exists an optimal solution for the virtual input \(x_{d(i+1)}^*\) which will be designed in the next section. Moreover, the third term of (9) inevitably shows up due to the design procedure, while the fourth term is deliberately added due to stability considerations.

**Step N:** In this step, similar to the previous steps, the system input will be designed. To this end, the system (2) can be rewritten as
\[
\dot{x}_N - \dot{x}_{N(i)} = \dot{e}_n = -\dot{x}_{N(i)} + f_N(x_1, ..., x_N) + g_N(x_1, ..., x_N)u
\]
\[ \dot{\hat{x}} = \Psi(x)\hat{\Lambda} + g(x, \ldots, x) + \bar{Y}_i \hat{\xi} + \bar{T}_i (\hat{\xi}^T \bar{T}_i) \]

\[ \text{for } 1 \leq i \leq N - 1 \text{ and } N \geq 2 \]

where \( \bar{Y}_i, \bar{T}_i \) are chosen as positive definite constant design matrices. Define the state estimation error as \( \hat{x} = x - \hat{x} \). Now, by subtracting the dynamics (13)-(14) from (1)-(2) yields the state estimation error dynamics as

\[ \dot{\hat{x}} = \Psi(x)\hat{\Lambda} - \bar{Y}_i \hat{\xi} - \bar{T}_i (\hat{\xi}^T \bar{T}_i) \]

for \( 1 \leq i \leq N \) and \( N \geq 2 \)

Next, the following lemma is stated in order to convert the strict-feedback system into an affine system.

\textbf{Lemma 1.} Consider the tracking dynamics defined in (6), (9), and (11). Assume that the virtual and real control input vector \( U = [x_{id}, \ldots, x_{ud}, u] \) is designed such that \( U = U^a + U^r \) where \( U^a = [x_{id}, \ldots, x_{ud}, u^a] \) is the feedforward control input designed in (8), (10), (12) and \( U^r = [x_{id}, \ldots, x_{ud}, u^r] \) represent the feedback control input which optimally stabilizes the system

\[ \begin{bmatrix} \hat{e}_i \\ \vdots \\ \hat{e}_N \end{bmatrix} = \begin{bmatrix} f_i(e_i) \\ \vdots \\ f_N(e_N) \end{bmatrix} + \begin{bmatrix} \hat{g}_1(x_i) \\ \vdots \\ \hat{g}_N(x_i, \ldots, x_N) \end{bmatrix} U^r. \] (16)

In this case, optimal control of (1) and (2) is equivalent to the optimal controller design (16) with \( \hat{\Lambda} \) being updated using

\[ \hat{\Lambda} = -\left( \Psi^T(x) - \Psi^T(x_i) \right) J_{ic}(E) + \Psi^T(x) J_{ic}(E) + \Psi^T(x) \hat{x} \]

where \( \alpha_i > 0 \) is a design parameter. In the other words, by applying \( U = U^a + U^r \) to the system (1) and (2), the system dynamics (1) and (2) are transformed into the error system given by (16).

\textbf{Proof.} By choosing \( J(E, \hat{\Lambda}, \hat{\xi}) = (J_i(E) + tr(\hat{\Lambda}^T \hat{\Lambda}) + \hat{\xi}^T \hat{\Lambda} \) and \( \hat{\Lambda} = [A_1^T \ldots A_N^T] \) as the Lyapunov candidate, with \( J_i(E), J_{ic}(E), \) and \( J_{ic}(E) \) defined prior to the step 1 of the backstepping process. Taking the derivative and evaluating the system dynamics (6), (9), (11) along the desired trajectory we have

\[ \dot{J} = J_{ic}(E) E + tr(\hat{\Lambda}^T \hat{\Lambda}) \]

and \( \hat{\Lambda} = [A_1^T \ldots A_N^T] \) as the Lyapunov candidate, with \( J_i(E), J_{ic}(E), \) and \( J_{ic}(E) \) defined prior to the step 1 of the backstepping process. Taking the derivative and evaluating

\[ \text{III. OPTIMAL TRAJECTORY AND CONTROL INPUT DESIGN} \]

The objective of this section is to optimally force the tracking error \( E \) in (16) to converge to zero. It is desirable to design the optimal control input vector defined by \( [x_{id}, \ldots, x_{ud}, u] \) such that the tracking error \( (e_i, \ldots, e_N) \) is stable while minimizing the cost function

\[ J = J_{ic}(E) E + tr(\hat{\Lambda}^T \hat{\Lambda}) + \hat{\xi}^T \hat{\Lambda} \]

\[ + J_{ic}(E) \left( J_i(e_i, \ldots, e_i) + g(x_i, \ldots, x_i) \right) \]

\[ + J_{ic}(E) \left( J_i(e_i, \ldots, e_i) + g(x_i, \ldots, x_i) \right) u^r - \sum_{i=1}^{N} \hat{\xi}_i^T \hat{\xi}_i \]

\[ + \left( \sum_{i=1}^{N} J_i^T \left( J_i(e_i, \ldots, e_i) + g(x_i, \ldots, x_i) \right) \right) \]

\[ \left( \begin{array}{c} g_i(x_i) \\ \vdots \\ g_N(x_i, \ldots, x_N) \end{array} \right) \left( \begin{array}{c} 0 \\ \vdots \\ 0 \end{array} \right) \]

\[ U^* - \sum_{i=1}^{N} \hat{\xi}_i^T \hat{\xi}_i - \sum_{i=1}^{N} \hat{\xi}_i^T \hat{\xi}_i \]

(18)

From (18) one can easily recognize that if the optimal controller \( U^* \) stabilizes the affine system then the first term in the RHS of equation (17) becomes negative and therefore \( \dot{J} \) becomes negative semidefinite which implies that the closed-loop signals are bounded. This in turn guarantees the tracking error \( E \) and the state estimation error \( \hat{\xi} \) go to zero asymptotically by invoking Barbalat’s Lemma [4]. Moreover, convergence of \( \hat{\xi} \) implies that \( \Psi(x)\hat{\Lambda} \) will converge to zero based on (15). Therefore, if the input is persistently exciting, \( \Psi(x) \) will not be zero while \( \hat{\Lambda} \) will converge to zero over time which implies that \( \dot{J} \) will converge to \( J_i \) provided \( U^r \) optimally stabilizes the affine system.

\textbf{Remark 1:} In fact, this lemma shows that by properly selecting the feedforward term for strict feedback systems, the optimal tracking control problem reduces to optimally stabilizing the nonlinear continuous-time systems in affine form described by (16). However, the unknown parameters, \( \hat{\Lambda} \), requires a state estimator(or observer) (13)-(14) for estimating the parameters though the states are measurable.

\textbf{Remark 2:} The update law in (17) is chosen to provide the unknown parameters given \( U^r \). However, since the actual control input is not optimal initially but attains the optimality over time, an extra term is needed in (17) which is given in (40). This extra term vanishes when the actual control input approaches \( U^r \) thus making (17) valid.

The next step is to design \( U^r \) in equation (18) that stabilizes the system (16) in an optimal manner. Next section provides the necessary optimal stabilizing term \( U^r \) in (16) and (18) that makes \( J_i \) negative semi-definite.
\[ V = \int r(E(\tau), U(\tau))d\tau, \]  
where \( E = [e_1, \ldots, e_N] \), \( U = [x_1, \ldots, x_N] \), and \( [x_1, \ldots, x_N] = X \).

In (19), \( r(E, U) = Q(E) + U^T R U \), \( Q(E) \geq 0 \) is the positive semi-definite penalty on the states, and \( R > 0 \), \( M = mN \) is a positive definite matrix with \( m = mN \) since the size of \( X \) is \( m \) times that of \( x \).

Next, consider the optimal stabilization problem for an affine system (17) in the error domain \( X = F(E) + G(X)U \),

\[ \begin{bmatrix} \hat{J}_1^T(e_1) & \cdots & \hat{J}_N^T(e_1) \end{bmatrix} = \hat{F}(E) \text{ and } G(X) = \text{diag}[g(x_1), \ldots, g(x_N)] \].

It is desirable to design an optimal control input that forces \( E \) to converge to zero while the cost function (19) is minimized. The control input \( U \) needs to be designed such that the cost function (19) will be finite. We define the Hamiltonian for the cost function (11) with an associated admissible control input \( U \) as

\[ H(E, U) = r(E, U) + V^T(E)F(E) + G(X)U, \]  
where \( V(E) \) is the gradient of the \( V(E) \) with respect to \( E \).

In the sequel, we will use the same terminology for denoting gradient of functions i.e. for any function \( \Omega(\psi) \), \( \hat{\Omega}(\psi) \) means gradient of \( \Omega(\psi) \) with respect to \( \psi \). \( U^* \) that minimizes the cost function (19) also minimizes the Hamiltonian (21); therefore, the optimal control is found by using the stationarity condition \( \partial H(E, U)/\partial U = 0 \) and revealed to be [8]

\[ U^* = -R^{-1}G(X)^T V(E)/2. \]  
(22)

By substituting the optimal control (22) into the Hamiltonian (21), observing \( H(E, U) = 0 \) reveals the HJB equation and the necessary and sufficient condition for optimal control to be [8]

\[ Q(E) + V^T(E)F(E) - V^T(E)G(X)R^{-1}G(X)^T V(E)/4 = 0, \]  
(23)

with \( V^*(0) = 0 \). For linear systems, equation (23) yields the standard algebraic Riccati equation (ARE) [8]. Before proceeding, the following technical lemma is required.

**Lemma 2** [6]. Given the nonlinear system (20) with associated cost function (19) and optimal control (22), let \( J(E) \) be a continuously differentiable, radially unbounded Lyapunov candidate such that \( \hat{J}(E) = J(E) \), \( \hat{J}(E) = \hat{J}(E) + G(X)U \), \( J(E) \leq 0 \), where \( \hat{J}(E) \) is the radially unbounded partial derivative of \( J(E) \). Moreover, let \( \hat{Q}(E) \) be a positive definite matrix satisfying \( \hat{Q}(E) = 0 \) only if \( \| E \| = 0 \) and \( \hat{Q}(E) \leq \hat{Q}(E) \) for positive constants \( \alpha_{min}, \alpha_{max} \) and \( \alpha_{max} \). In addition, let \( \hat{Q}(E) \) satisfy \( \lim_{E \to 0} \hat{Q}(E) = \infty \) as well as

\[ V^T(E) \hat{Q}(E) J(E) = r(E, U^*) = Q(E) + U^T R U^*. \]  
(24)

Then, the following relation holds

\[ J(E) + G(X)U^* = -J(E) \hat{Q}(E) J(E) \]  
(25)

**Proof:** When the optimal control (22) is applied to the nonlinear system (20), the cost function (19) becomes a Lyapunov function rendering

\[ \dot{V}(E) = V(E) + Q(E) - U^T R U^* = -Q(E) + U^T R U^*. \]  
(26)

From (22), after manipulation and substitution of (24), equation (26) is rewritten as

\[ F(E) + G(X)U = -(V^T(E) + V^T(E) \hat{Q}(E)) J_{IE} = -\hat{Q}(E) J_{IE} \]  
(27)

Now, multiply both sides of (27) by \( J_{IE}^T \) yields the desired relationship in (25).

The generalized bound \( \delta(E) \) is taken as \( \delta(E) < \sqrt{K_{\theta} \| E \|} \) where \( K_{\theta} \) can be selected to satisfy general bounds and \( K' \) is a constant. Moving on, by using the LIP assumption, the cost function can be represented as

\[ V(E) = \theta^T \phi(E), \]  
(28)

i.e. it is linear in parameter (LIP) with respect to \( \theta \). The target parameter vector \( \theta \) is assumed to be bounded above according to \( \| \theta \| \leq \theta_{max} \) [3]. The gradient of the cost function (28) is written as

\[ \partial V(E)/\partial \theta = V(E) \theta = V^T(E) \theta. \]  
(29)

Now, using (29), the optimal control (14) and HJB equation (23) are rewritten as

\[ U^* = -R^{-1}G(E)^T V(E) \theta / 2 \]  
(30)

and

\[ H^*(E, \theta) = Q(E) + \theta \nabla \phi(E) F(E) - \theta \nabla \phi(E) D V^T(E) \theta / 4 = 0, \]  
(31)

where \( D = G(E) R^{-1} G(E)^T > 0 \) is bounded such that \( D_{min} \leq \| D \| \leq D_{max} \) for known constants \( D_{min} \) and \( D_{max} \). Moving on, the estimate of (11) is now written as

\[ \hat{V}(E) = \theta^T \phi(E) \]  
(32)

where \( \hat{\theta} \) is the estimate of the target parameter vector \( \theta \). Similarly, the estimate of the optimal control (14) is written in terms of \( \theta \) as

\[ \hat{U}^* = -R^{-1}G(E)^T V(E) \theta / 2 \]  
(33)

It is shown [6] that an initial stabilizing control is not required to implement the proposed scheme in contrast to [12] and [13], which require initial control policies to be stabilizing. Now, using (32), the approximate Hamiltonian can be written as

\[ H(E, \hat{\theta}) = Q(E) + \hat{\theta} \nabla \phi(E) F(E) - \hat{\theta} \nabla \phi(E) D V^T(E) \theta / 4. \]  
(34)

Observing the definition of cost function estimation (32) and the Hamiltonian function (34), it is evident that both become zero when \( \| E \| = 0 \). Thus, once the system states have converged to zero, the cost function estimation can no longer be updated. This can be viewed as a persistency of excitation (PE) requirement for the inputs to the cost function estimator [12], [13]. That is, the system states must be exciting long enough for the parameter update scheme to learn the optimal cost function.

Recalling the HJB equation in (23), the parameter estimate \( \hat{\theta} \) should be tuned to minimize \( H(E, \hat{\theta}) \). However, tuning to minimize \( H(E, \theta) \) alone does not ensure the stability of the nonlinear system (20) during the adaptation process. Therefore, the proposed tuning algorithm is
designed to minimize (34) while considering the stability of (20) and written as
\[
\hat{\dot{\theta}} = -\frac{\beta_2 \sigma}{(\sigma^2 + \sigma + 1)} \left( Q(E) + \hat{\theta}^T \nabla_{\theta} \phi(E) \hat{\theta} - \frac{1}{4} \hat{\theta}^T \nabla_{\theta} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{\theta} \right) \\
+ \beta_2 \Sigma(E, \hat{U}) \nabla_{\theta} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{J}_{1 \theta}(E)/2
\]
where \(\sigma = \nabla_{\phi} \phi(F(E) - \nabla_{\phi} \phi(E) \hat{\theta})/2\), \(\beta_2 > 0\) and \(\beta_2 > 0\) are design constants, \(J_{1 \theta}(E)\) is a Lyapunov function described in Lemma 2. The term \(\Sigma(E, \hat{U})\) is used to determine the stability condition of the closed system and it is defined as
\[
\Sigma(E, \hat{U}) = \begin{cases} 
0 & \text{if } J_{1 \theta}(E) = J_{1 \theta}(E) \\
- \frac{1}{2} G(E) R^{-1} G(E) \nabla_{\phi} \phi(E) \hat{\theta} \hat{\theta} < 0 & \text{and } \| \hat{X} \| \leq B_{\hat{X}} \\text{(36)} \\
1 & \text{otherwise}
\end{cases}
\]

The condition \(\| \hat{X} \| \leq B_{\hat{X}}\) assures that the state estimation error \(\hat{X}\) will be in a small enough bound \(B_{\hat{X}}\) before switching off the last term of the update law (35). Existence of \(B_{\hat{X}}\) is proven in the overall stability proof, therefore for calculating the value of \(B_{\hat{X}}\) the reader is referred to the proof of Theorem 1 in the appendix. Here, \(J_{1 \theta}(E)\) is a Lyapunov function whose described in the Lemma 2. The first term in (35) is the portion of the tuning law which seeks to minimize (34) and was derived using a normalized gradient descent scheme with the auxiliary HJB error defined as
\[
E_{HJB} = H(E, \hat{\theta})^2/2
\]

Meanwhile, the second term in the parameter tuning law (35) is included as it is required in the process of Lyapunov-based stability proof of the overall closed loop system. Moving on, we now form the dynamics of the cost function parameter estimation error \(\hat{\theta} = \dot{\theta} - \hat{\theta}\). Observing \(Q(E) = -\theta^T \nabla_{\phi} \phi(E) F(E) + \theta^T \nabla_{\phi} \phi(E) \nabla_{\phi}^2 \phi(E) \theta/4\) from (30), the approximate HJB equation (34) can be rewritten in terms of \(\hat{\theta}\) as
\[
\hat{H}(E, \hat{\theta}) = -\hat{\theta}^T \nabla_{\phi} \phi(E) \hat{\theta} - \hat{\theta}^T \nabla_{\phi} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{\theta}/2 \\
+ \hat{\theta}^T \nabla_{\phi} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{\theta}/4.
\]

Next, observing \(\hat{\dot{\theta}} = \hat{\theta} - \hat{\theta}\) and \(\hat{\sigma} = \nabla_{\phi} \phi(E) \hat{\theta} + \nabla_{\phi} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{\theta}/2\) where 
\(\hat{E} = F(E) + G(X) \hat{U}\), the error dynamics of (20) are written as
\[
\hat{\dot{\theta}} = \beta_1 \frac{1}{\rho^2} \left( \nabla_{\phi} \phi(E) \hat{E} - \hat{\theta}^T \nabla_{\phi} \phi(E) \hat{\theta} \right) \\
= \left( -\hat{\theta}^T \nabla_{\phi} \phi(E) \hat{E} - \hat{\theta}^T \nabla_{\phi} \phi(E) \hat{\theta} - \frac{1}{4} \hat{\theta}^T \nabla_{\phi} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{\theta} \right) \\
+ \beta_1 \Sigma(E, \hat{U}) \nabla_{\phi} \phi(E) \nabla_{\phi}^2 \phi(E) \hat{J}_{1 \theta}(E)/2
\]
where \(\rho = (\sigma^2 + \sigma + 1)\). As the next step a state estimator is also required and represented in Section II equations (13)-(14). As mentioned in the Remark 3, it is required to modify the update law (17) since \(\hat{U}\) is not available and the estimated optimal control law \(\hat{U}\) is used instead. This term vanishes once the actual control input converges to the optimal value. Therefore the update law chosen after incorporating the extra term is given by
\[
\hat{\dot{\theta}} = -\left( \Psi^T(x) - \Psi^T(x_d) J_{1 \theta}(E) + \Psi^T(x) J_{1 \theta}(E) + \Psi^T(X) \hat{X} \right) \\
+ (1 - \Sigma(E, \hat{U})) \Psi^T(x) J_{1 \theta}(E)
\]

Next the stability of the optimal adaptive scheme is given.

**Theorem 1**: (Stability of Optimal Adaptive Control Scheme with Partially Unknown Dynamics). Given the nonlinear system (6), (9), and (11) with the target HJB equation (23), and let the tuning law for the internal dynamics and the cost function estimation be given by (40) and (35) respectively. Then, when the design parameter is selected as \(\beta_1 / \beta_2 > 4096 \beta_1 / D_{min}^2\), \(\sigma_1(\bar{Y}) \sigma_1^2(\bar{Y}) > 5\beta_1 \| \nabla_{\phi} \phi(E) \|\| \theta \| \), and \(\sigma_2^2(\bar{Y}) \sigma_2^2(\bar{Y}) > 4096 \beta_1 \beta_2 / D_{min}^2\), the closed loop system is uniformly stable such that the tracking error \(E\), the internal dynamics parameter error \(\hat{\lambda}\), and cost function parameter error \(\hat{\theta}\) converge to zero if the input is PE which also implies that \(\hat{Y} \rightarrow V^*\) and \(\hat{U} \rightarrow U^*\).

**Proof**: The proof is omitted due to the page limit.

IV. SIMULATION RESULTS

Consider the following nonlinear system in the form of (1)-(2) described by
\[
\begin{bmatrix}
\dot{x}_1 \\
\dot{x}_2
\end{bmatrix} =
\begin{bmatrix}
x_2 \\
-x_1 (x_2 + \tan^{-1}(5x_3)) - \frac{5x_1^2}{2(1 + 5x_1^2)} + 4x_2
\end{bmatrix} +
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

Using the cost function (19) with \(Q(E) = E^T E\) and \(R = 1\), the regression vector for the cost function is selected as \(\phi(E) = [x_1, x_2, x_3, x_3, x_3, x_3, x_3, x_3, x_3, x_3, x_3, x_3, x_3, x_3]^T\) where, \(x_1 = x_1 - x_1^d\), \(x_2 = x_2 - x_2^d\), \(x_3 = x_3^d - x_3\), and \(z_2 = z_2 - z_2\). The regression function \(\phi(E)\) is chosen of polynomial entries though one of the entries is selected as \(z_3^d \tan^{-1}(5z_3^d)\) due to the system dynamics. Obviously, it can be replaced by a finite number of other polynomial entries. Moreover, it is obvious that finding \(K^*\) is difficult, and therefore, in order to make theorem 1 conditions hold, we choose \(\beta_1 = 0.007\), \(\beta_2 = 0.05\) for all the possible cases of \(K^*\). Moreover, \(Y = \bar{Y} = 50 \text{diag}(1,1)\) and the initial conditions of the system states are taken as \([x_1, x_2, x_3, z_1]^T = [10 -5 -2 2]^T\) while all the \(\lambda\) and \(\hat{\phi}\) parameter vectors are initialized to zero. That is, no initial stabilizing control was utilized. The regression matrix for the internal dynamics is chosen to be
\[
\Psi(x) = \begin{bmatrix}
x_1 \\
x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1 \\
x_1^d - 5x_1
\end{bmatrix}
\]

which renders the target internal dynamics parameters as
\[
\Lambda_1 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}^T
\]

\[
\Lambda_2 = \begin{bmatrix}
-2 & 0 & 0 & 0 & 1 & -4 & 0
\end{bmatrix}
\]

\[
\Lambda_3 = \begin{bmatrix}
-1 & 0 & 0 & 0 & 2 & 0 & -3
\end{bmatrix}
\]
Furthermore, it is desired that the output track 
\[ X_t = [\sin(t/50) \sin(t/40)]^T \] as the desired trajectory. 

Figure 1 illustrates the closed loop behavior while the states track the desired trajectory which is highly satisfactory. It is important to note that the system (41) is unstable and the system tends to diverge if suitable control input is not applied. However, the proposed optimal controller still keeps the system stable though the total energy applied is lower than a normal stabilizing controller. For the first 500 seconds of the simulation a probe noise is injected to the system dynamics to provide the PE condition that is required for the convergence of the cost function and the internal dynamics parameters to converge. Then, the noise is removed to show that the system tracks the desired trajectory when weights converge.

![Figure 1. Performance of the optimal adaptive controller.](image)

![Figure 2. The cost function parameter \( \Theta \) convergence.](image)

![Figure 3. Parameter convergence \( \hat{\Lambda} \).](image)

![Figure 4. The control input with \( U^* + U^\ast \).](image)

Convergence of the unknown parameters of the cost function is illustrated in the Figure 2. Figure 3 presents the convergence of the unknown parameters of internal dynamics. Finally, Figure 4 shows that the control input, which is applied to the system, is bounded.

V. CONCLUSION

This work proposes an adaptive optimal scheme for stabilizing nonlinear MIMO systems for strict feedback nonlinear continuous-time systems with unknown internal dynamics. An adaptive LIP assumption is proposed to solve the Hamilton Jacobi-Bellman equation forward-in-time while the other unknown dynamics are estimated by using a state estimator with adaptive elements. By applying Lyapunov analysis, it is demonstrated that the problem of online optimal tracking of strict feedback systems can be accomplished in two steps: a) finding a feedforward tracking controller; and b) designing an optimal feedback controller that stabilizes the tracking error dynamics of the affine system. Numerical results demonstrate the approach to an unstable plant whose optimal stabilization/tracking is more challenging.

REFERENCES