Optimal Tracking Control for Uncertain Nonlinear Systems with Prescribed Performance via Critic-Only ADP

Hongyang Dong, Xiaowei Zhao, and Biao Luo

Abstract—**This paper addresses the tracking control problem for a class of nonlinear systems described by Euler-Lagrange equations with uncertain system parameters. The proposed control scheme is capable of guaranteeing prescribed performance from two aspects: 1) A special parameter estimator with prescribed performance properties is embedded in the control scheme. The estimator not only ensures the exponential convergence of the estimation errors under relaxed excitation conditions but also can restrict all estimates to pre-determined bounds during the whole estimation process; 2) The proposed controller can strictly guarantee the user-defined performance specifications on tracking errors, including convergence rate, maximum overshoot, and residual set. More importantly, it has the optimizing ability for the trade-off between performance and control cost. A state transformation method is employed to transform the constrained optimal tracking control problem to an unconstrained stationary optimal problem. Then a critic-only adaptive dynamic programming algorithm is designed to approximate the solution of the Hamilton-Jacobi-Bellman equation and the corresponding optimal control policy. Uniformly ultimately bounded stability is guaranteed via Lyapunov-based stability analysis. Finally, numerical simulation results demonstrate the effectiveness of the proposed control scheme.**

Index Terms—**Prescribed Performance; Adaptive Dynamic Programming; Optimal Tracking Control; Parameter Estimation; Reinforcement Learning.**

I. Introduction

Tracking control problems of nonlinear systems with uncertainties have aroused extensive attention and been widely investigated by the adaptive control community. Traditional certainty-equivalence (CE) adaptive control methods [1] can guarantee the convergence of tracking errors in the presence of uncertainties. However their closed-loop transient performance (e.g., convergence rate and overshoot) is difficult to analyze. Besides, due to the inherent limitations arising from the CE design structure, CE adaptive controllers may result in significant performance degradation [2], [3] when compared with the corresponding deterministic-case controllers. Some non-CE adaptive control schemes, such as the immersion-and-invariance method [4], [5] and its extensions [6], [7], can partially address these limitations and achieve improved closed-loop performance. But it is still a challenging problem to quantitatively analyze or characterize the performance of these non-CE adaptive controllers. We note that specifying the closed-loop performance is very important for many practical control systems, for example a large overshoot may cause severe damage to the structure of many mechanical systems. Nevertheless, it is quite challenging to achieve user-defined performance specifications for nonlinear systems, particularly in the presence of uncertain parameters. To address this issue, a novel prescribed-performance control (PPC) method was recently proposed by Bechlioulis and Rovithakis [8], [9], [10]. State in a nutshell, PPC ensures that the specific requirements on transient performance (mainly including convergence rate and overshoot) and residual set can be quantitatively characterized by employing judiciously designed prescribed performance functions (PPF). Given the merits of PPC, it has been applied to various classes of nonlinear systems [11], [12], [13], [14], [15], [16], [17]. A fault-tolerant adaptive attitude controller with prescribed performance was proposed in [14] for spacecraft under parameter uncertainties and actuator faults. A neural network (NN)-based PPC method for the tracking control problem of robot manipulators was designed in [15]. An adaptive controller for a switched nonlinear system with multiple prescribed performance bounds was proposed in [16].

All these aforementioned results lack optimizing abilities to make the trade-off between performance and control cost, while this issue is of great importance for many systems (for example, motion control costs of spacecraft, usually in terms of fuel and electricity, are the most essential resources of on-orbit missions). To the best knowledge of the authors, it still remains an open problem to design an optimal controller with prescribed performance for uncertain nonlinear systems, especially for tracking control cases. On the one hand, optimal control problems usually require to solve the Hamilton-Jacobi-Bellman (HJB) equations, which is intractable for nonlinear systems. Although iterative learning algorithms [18], [19], [20], [21] can approximate the solutions of HJB equations, how to guarantee the prescribed performance specifications during the whole control process needs to be investigated. On the other hand, the existence of parameter uncertainties also increases the complexity of this problem. How to accurately analyze the performance of parameter estimator/identifier is also a challenging task.

To solve the optimal control problems of nonlinear systems, recently the reinforcement learning (RL)-based control technique, which is commonly referred to as adaptive dynamic

This work was funded by the UK Engineering and Physical Sciences Research Council (grant number: EP/S001905/1).

H. Dong and X. Zhao (Corresponding Author) are with the School of Engineering, University of Warwick, Coventry, CV4 7AL, UK. Emails: {hongyang.dong, xiaowei.zhao}@warwick.ac.uk.

B. Luo is with the School of Automation, Central South University, Changsha 410083, China. Email: biao.luo@hotmail.com.

programming (ADP) [22], [23], [24], [25], [26], [27], [28], has attracted extensive research interest. The fundamental principle of ADP is to improve actions by properly evaluating feedback from the environment, forming the actor-critic architectures [20], [29], [30], [31], [32]. This innovative method has been utilized to approximate the solutions of HJB equations and subsequently achieve near-optimal control. In the presence of uncertainties, estimators (or identifiers, when neural networks are employed to approximate system dynamics) can be further designed to estimate unknown parameters and then embedded into the actor-critic architectures. It should be emphasized that the performance and accuracy of the parameter estimator are essential for the whole control scheme design. A real-time estimator with poor performance and accuracy can lead to the instability of the closed-loop system. Besides, for many practical systems, some prior information of uncertain parameters are often available. For example, it is common to know the lower/upper bounds of the mass of a rigid body. These kinds of information, to some extent, can be regarded as estimation constraints. Violation of such constraints may cause that the parameter update process happens outside the feasible region, which makes no sense and leads to poor transient performance. However, the constraint handling abilities of ADP methods are still immature [33], and relevant studies are still very limited as mentioned in [13], [34]. Most of existing ADP controllers neither consider the estimation bounds of unknown parameters nor can meet the prescribed performance specifications (which can also be regarded as constraints) of system states.

Motivated by these facts, in the present paper, a novel tracking controller with both prescribed-performance and optimizing abilities is proposed for nonlinear systems described by the Euler-Lagrange (EL) equation under parameter uncertainties. We mention that the EL equation can represent a large class of nonlinear systems, such as robot manipulators [15], [35], [36] and rigid-body attitude dynamics [37]. A continuous PPF is employed to specify the transient performance and residual set of the coordinate tracking error. Based on the PPF, an augmented state with a transformation law is designed to transform the constrained control problem into an unconstrained one. Then by employing a virtual reference control signal, we further turn the optimal tracking control problem of the augmented system into a stationary optimal control problem. A novel real-time estimator is designed to identify unknown parameters, and then a critic-only structure is designed to approximate the optimal cost function and control policy. Uniformly ultimately bounded (UUB) stability of the proposed control scheme is guaranteed via Lyapunov-based stability analysis. The main contribution of this paper includes:

1) We enable PPC to has essential optimizing abilities. By utilizing the ADP technique, the proposed controller can make a trade-off between performance and cost while strictly guaranteeing performance specifications. From another point of view, we also show how to handle performance constraints of states in the ADP-based control architecture.

2) A novel constrained estimator is proposed to deal with system uncertainties. By employing a special projection law, the estimator can restrict all real-time estimates to a feasible region. Moreover, it also guarantees the exponential convergence of estimation errors with a user-defined convergence rate, subject to the satisfaction of finite excitation (FE) conditions [38], [39], [40], [41]. Thus the performance of the estimator is also prescribed, to some extent.

3) Motivated by the concurrent learning (CL) technique [38], [39], [40] and its extensions [30], [31], [32], real-time data and past measurements are concurrently introduced into the update laws of both the estimator and the critic NN. This guarantees the UUB of the closed-loop system under FE conditions, which is a significant relaxation when compared with conventional persistent excitation (PE) conditions. This design also allows us to employ a critic-only control structure, simplifying the commonly-used actor-critic ADP scheme. Moreover, in the parameter estimator, we circumvent the requirement of immeasurable derivatives in the CL technique [38], [39], [40].

The remainder of this paper is organized as follows. In Sec. II, mathematical preliminaries are introduced, and the optimal tracking control problem with prescribed performance specifications is formalized. Then an estimator-based critic-only control scheme is designed in Sec. III with Lyapunov-based stability analysis. Simulation results are demonstrated in Sec. IV to show the features and effectiveness of the proposed method. Finally, we conclude the paper in Sec. V.

II. Preliminaries and Problem Formulation

A. Preliminaries

Throughout the paper, the time domain of all functions is $\mathbb{R}_{\geq 0}$. The notation $\|\cdot\|$ denotes the Euclidean norm of vectors and the induced norm of matrices. We denote $\nabla_x(\cdot) = (\partial(\cdot)/\partial x)^{\mathrm{T}}$, where $(\cdot)^{\mathrm{T}}$ is the transpose of the corresponding vector/matrix, and we set $\nabla_x^{\mathrm{T}}(\cdot) = (\nabla_x(\cdot))^{\mathrm{T}}$. Besides, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are employed to represent the minimum and maximum eigenvalues of the corresponding matrix, respectively.

The definitions of finite excitation and persistent excitation conditions are presenting as follows.

Definition 1 (Finite Excitation, FE) [40]: A bounded signal $s(\cdot) : \mathbb{R} \to \mathbb{R}^{n\times m}$ is said to be finite exciting over an interval $[t, t+T]$, where $t \geq 0$ is a fixed time index, if there exist finite constants $T > 0$ and $c > 0$ such that

$$\int_t^{t+T} s^{\mathrm{T}}(\tau)s(\tau)\mathrm{d}\tau \geq cI_{m\times m} \tag{1}$$

where $I_{m\times m}$ is the m dimensional identity matrix.

Definition 2 (Persistent Excitation, PE) [42]: A bounded signal $s(\cdot) : \mathbb{R} \to \mathbb{R}^{n\times m}$ is said to be persistently exciting if there exist finite positive constants c and T such that for arbitrary $t \geq 0$, one has

$$\int_t^{t+T} s^{\mathrm{T}}(\tau)s(\tau)\mathrm{d}\tau \geq cI_{m\times m} \tag{2}$$

The contrast between FE and PE conditions is clearly indicated by their definitions. The former requires the signal to be excited just over a specific finite time interval, whereas, qualitatively speaking, PE implies the satisfaction of FE throughout the whole timeline.

B. System Model

In this paper, a class of nonlinear system described by the following Euler-Lagrange equation is considered:

$$\dot{x}_1 = x_2 \tag{3}$$

$$M(x_1)\dot{x}_2 + C(x_1, x_2)x_2 + g(x_1, x_2) = u \tag{4}$$

where $x_1, x_2 \in \mathbb{R}^n$ respectively denote the generalized coordinates and velocities, $M(x_1) \in \mathbb{R}^{n\times n}$ is the generalized mass matrix, $C(x_1, x_2) \in \mathbb{R}^{n\times n}$ denotes the Coriolis matrix, $u \in \mathbb{R}^n$ represents the control input to be designed, and $g(x_1, x_2) \in \mathbb{R}^n$ denotes the gravity or friction-related vector. Note that $M(x_1)$, $C(x_1, x_2)$ and $g(x_1, x_2)$ all contain uncertain parameters, and they satisfy a parameter-affine representation as follows

$$M(x_1)\dot{x}_2 + C(x_1, x_2)x_2 + g(x_1, x_2) = H(x_1, x_2, \dot{x}_2)\theta \tag{5}$$

where H is a regressor matrix, and $\theta \in \mathbb{R}^m$ is called the parameter vector which contains all the unknown parameters of the system. Please note that though θ is unknown, the regressor matrix H is available for controller design. A simple way to get the expression of H is to take Jacobian of the left-hand side of (5) with respect to θ. Other commonly-used properties of the system in (3) and (4) include: 1) $M(x_1)$ is positive definite for all $x_1 \in \mathbb{R}^n$, and there exist positive constants m_m and m_M such that $m_m \leq \|M(x_1)\| \leq m_M$. 2) $M(x_1) - 2C(x_1, x_2)$ is an anti-symmetric matrix. Hereinafter, for ease of notation, the arguments of M, C and g are omitted when there is no ambiguity.

The system state $x \triangleq [x_1^{\mathrm{T}}, x_2^{\mathrm{T}}]^{\mathrm{T}}$ is required to track a reference signal $x_r \triangleq [x_{r1}^{\mathrm{T}}, x_{r2}^{\mathrm{T}}]^{\mathrm{T}}$ with $\dot{x}_{r1} = x_{r2}$ and $x_{r1}, x_{r2}, \dot{x}_{r2} \in \mathcal{L}_\infty$. Since x_{r1} and x_{r2} are user-defined, we assume that $\dot{x}_{r2} = f_r(x_r)$ as [21], [32], where $f_r : \mathbb{R}^{2n} \to \mathbb{R}^n$. Defining error states to be: $z_1 \triangleq x_1 - x_{r1}$ and $z_2 \triangleq x_2 - x_{r2}$, one can get the following error model:

$$\dot{z}_1 = z_2 \tag{6}$$

$$\begin{aligned}\dot{z}_2 =& - f_r(x_r) + M^{-1}(x_1)[-C(x_1, x_2)x_2 - g(x_1, x_2) + u]\\ =& M^{-1}(x_1)[-H(x_1, x_2, f_r(x_r))\theta + u]\end{aligned} \tag{7}$$

A technical challenge of designing ADP-based tracking controllers for continuous nonlinear systems is the non-autonomous nature associated with the trajectory tracking problems. And directly employing the original control input u into the cost index will render the cost index ill-defined (since u can be persistently exciting in tracking control cases). Following the strategy given in [42] and [43], a reference control signal is designed in our paper to solve this problem. A fundamental requirement for the design of u_r is that $u - u_r$ should converge to zero when all tracking errors converge to zero. So that $\mu \triangleq u - u_r$ can be employed into the cost index. The reference control signal employed here is

$$\begin{aligned}u_r(x_r) \triangleq& M(x_{r1})f_r(x_r) + C(x_{r1}, x_{r2})x_{r2} + g(x_{r1}, x_{r2})\\ =& H(x_{r1}, x_{r2}, f_r(x_r))\theta\end{aligned} \tag{8}$$

This renders

$$\dot{z}_1 = z_2 \tag{9}$$

$$\dot{z}_2 = M^{-1}(x_1)[Y(z_1, z_2, x_{r1}, x_{r2})\theta + \mu] \tag{10}$$

where $Y(z_1, z_2, x_{r1}, x_{r2}) \triangleq H(x_{r1}, x_{r2}, f_r(x_r)) - H(x_1, x_2, f_r(x_r))$ is employed for ease of notation. One can see that $\mu \to 0_n$ when $x_1 \to x_{r1}$, $x_2 \to x_{r2}$, and $\dot{x}_2 \to \dot{x}_{r2}$, satisfying the requirement as discussed before.

Based on the error model in (9) and (10), the control objective is to guarantee the convergence of tracking errors (z_1 and z_2) by designing the augmented control input μ under the uncertain parameter vector θ. Besides, the generalized coordinate tracking error z_1 is required to satisfy the performance specifications as discussed in the following subsection.

C. Prescribed Performance Specifications and Error Transformation

Following the design philosophy of PPC in [8], [9], [10], a continuous PPF is employed in this paper to restrict z_1:

$$\rho_i(t) \triangleq (\rho_{i0} - \rho_{i\infty})\mathrm{e}^{-l_i t} + \rho_{i\infty}, \; i = 1, 2, ...n \tag{11}$$

where ρ_{i0}, $\rho_{i\infty}$ and l_i are positive constants, with $\rho_{i0} > \rho_{i\infty}$ and $\rho_{i0} > z_{1i}(0)$, and here z_{1i} denotes the i^{th} entry of z_1, $i = 1, 2, ...n$.

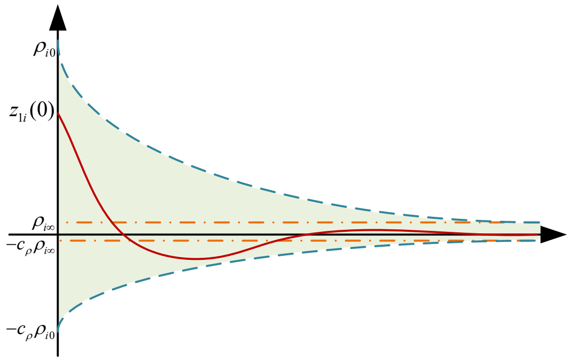

Figure 1: Prescribed performance illustration.

The coordinate tracking error z_1 is required to satisfy:

$$-[c_\rho + l_{ei}(1 - c_\rho)]\rho_i(t) < z_{1i}(t) < [1 - l_{ei}(1 - c_\rho)]\rho_i(t) \tag{12}$$

for all $t \geq 0$ and $i = 1, 2, ..., n$, where $0 < c_\rho \leq 1$ is a positive constant which is employed to restrict the maximum overshoot, and

$$l_{ei} = \begin{cases} 0, & \text{if } z_{1i}(0) \geq 0 \\ 1, & \text{if } z_{1i}(0) < 0 \end{cases} \tag{13}$$

is employed to adjust the performance requirements based on the initial condition of z_{1i}. Eq. (12) actually defines an admissible domain for z_1, we denote it by $\mathfrak{D}_z$. For the case $z_{1i}(0) \geq 0$ and $z_{1i}(0) \in \mathfrak{D}_z$, the performance specifications and the admissible domain $\mathfrak{D}_z$ are illustrated in Fig. 1.

As a brief summary, $\rho_i(t)$ quantifies the overshoot, steady-state error, and convergence rate of the coordinate tracking error. For each entry z_{1i} of z_1, $i = 1, 2, ..., n$, the maximum overshoot of z_{1i} is less than $|c_\rho \rho_i(0)|$, and the steady-state error is bounded by $[-\rho_{i\infty}, \rho_{i\infty}]$. Moreover, before the tracking error convergence to the residual set, the convergence rate (in the sense of exponential convergence) is always larger than l_i.

Based on the PPF, we define $q_i(t) \triangleq z_{1i}(t)/\rho_i(t)$ and consider the following transformation law:

$$e_i(t) \triangleq \ln \frac{c_\rho + l_{ei}(1-c_\rho) + q_i(t)}{1 - l_{ei}(1-c_\rho) - q_i(t)} \tag{14}$$

It can be readily checked that e_i, $i = 1, 2, ..., n$, is well-defined when $z_{1i}(t) \in \mathfrak{D}_z$.

Eq. (14), in turn, leads to $z_{1i} = \hbar_i(e_i, \rho_i)$ with

$$\hbar_i(e_i, \rho_i) \triangleq [(1+c_\rho)\mathrm{sig}(e_i) - c_\rho - l_{ei}(1-c_\rho)]\rho_i \tag{15}$$

and here $\mathrm{sig}(e_i) \triangleq 1/(1+\mathrm{e}^{-e_i})$ denotes the sigmoid function. Thus, by employing (14), the tracking error z_1 in the domain $\mathfrak{D}_z$ is transformed to an auxiliary tracking error $e = [e_1, e_2, ..., e_n]$ defined on $\mathbb{R}^n$. This is the most essential property of PPF.

Taking the time derivative of (14), one has

$$\begin{aligned} \dot{e}_i &= \frac{2(z_{2i} - z_{1i}\dot{\rho}_i/\rho_i)}{[1 - l_{ei}(1-c_\rho) - q_i][c_\rho + l_{ei}(1-c_\rho) + q_i]\rho_i} \\ &= r_i[z_{2i} - \frac{\dot{\rho}_i}{\rho_i}\hbar_i(e_i, \rho_i)] \end{aligned} \tag{16}$$

where $r_i \triangleq 2/[(1 - l_{ei}(1-c_\rho) - q_i)(c_\rho + l_{ei}(1-c_\rho) + q_i)\rho_i]$. Then one has the following transformed error model:

$$\begin{aligned} \dot{e} &= \Upsilon[z_2 - \mathcal{P}z_1] \\ \dot{z}_2 &= M^{-1} \cdot [Y(z_1, z_2, x_{r1}, x_{r2})\theta + \mu] \end{aligned} \tag{17}$$

where $z_1 = [\hbar_1, \hbar_2, .., \hbar_n]^{\mathrm{T}}$ can be represented by e and ρ with $\rho = [\rho_1, \rho_2, ..., \rho_n]^{\mathrm{T}}$. Besides, we denote $\Upsilon \triangleq \mathrm{diag}\{r_1, r_2, ..., r_n\}$, and $\mathcal{P} \triangleq \mathrm{diag}\{\dot{\rho}_1/\rho_1, \dot{\rho}_2/\rho_2, ..., \dot{\rho}_n/\rho_n\}$. Accordingly, an augmented model can be organized as follows.

$$X = F(X) + G(X)\mu \tag{18}$$

where $X \triangleq [e^{\mathrm{T}}, z_2^{\mathrm{T}}, x_{r1}^{\mathrm{T}}, x_{r2}^{\mathrm{T}}, \rho^{\mathrm{T}}, \dot{\rho}^{\mathrm{T}}]^{\mathrm{T}}$ is the augmented state vector, and

$$\begin{aligned} F(X) &\triangleq \begin{bmatrix} \Upsilon(z_2 - \mathcal{P}z_1) \\ M^{-1} \cdot [Y(z_1, z_2, x_{r1}, x_{r2})\theta] \\ x_{r2} \\ f_r(x_r) \\ \dot{\rho} \\ -L\dot{\rho} \end{bmatrix}, \\ G(X) &\triangleq \left[0_{n\times n}, M^{-1}, 0_{n\times n}, 0_{n\times n}, 0_{n\times n}, 0_{n\times n}\right]^{\mathrm{T}}. \end{aligned}$$

and here $L \triangleq \mathrm{diag}\{l_1, l_2, ..., l_n\}$.

D. Optimal Control Formulation

By employing the state transformation law in (14), the prescribed-performance tracking control can be achieved through guaranteeing the boundedness and convergence of e and z_2. Moreover, to reduce the control cost, we aim to find a control policy μ that minimizes the following cost index:

$$J \triangleq \int_0^\infty [r(e(\tau), z_2(\tau)) + \mu^{\mathrm{T}}(\tau)R\mu(\tau)]\mathrm{d}\tau \tag{19}$$

where $r(e, z_2) \triangleq e^{\mathrm{T}}Q_e e + z_2^{\mathrm{T}}Q_z z_2$, and Q_e, Q_z and R are positive-definite matrices. One can see that, by employing the error states and also the virtual control signal μ, J is well-defined and the optimal tracking control problem of the original system is transformed to a stationary optimal problem.

Based on (19), one can construct the following cost functional:

$$V(X) \triangleq \int_t^\infty [r(e(\tau), z_2(\tau)) + \mu^{\mathrm{T}}(\tau)R\mu(\tau)]\mathrm{d}\tau \tag{20}$$

Moreover, substituting (18) into the time derivative of (20) yields the following Hamiltonian:

$$\nabla_X^{\mathrm{T}} V \cdot (F + G\mu) + r + \mu^{\mathrm{T}} R\mu = 0 \tag{21}$$

We use μ^* to denote the optimal control policy and V^* to denote the corresponding optimal cost functional. Then, by taking partial differential for both sides of Eq. (21) with respect to μ, one can obtain a closed-form expression of μ^* as follows

$$\mu^* = -\frac{1}{2}R^{-1}G^{\mathrm{T}}\nabla_X V^* \tag{22}$$

Then substituting μ^* back into Eq. (21) results in the well-known HJB equation in terms of $\nabla_X V^*$:

$$r + \nabla_X^{\mathrm{T}} V^* F - \frac{1}{4}\nabla_X^{\mathrm{T}} V^* G R^{-1} G^{\mathrm{T}} \nabla_X V^* = 0 \tag{23}$$

However, it is very hard to analytically solve (23), and the existence of parameter uncertainties also increases the technical difficulties of this optimal control problem. To address these issues, a parameter estimator and a critic-only ADP controller are designed in the following section to compensate the uncertainties and approximate the optimal control policy.

III. Design of An Estimator-based Critic-Only ADP Controller with Prescribed Performance

A. Design of a Constrained Parameter Estimator

As the essential part of the whole control strategy, a novel constrained parameter estimator is proposed in this subsection to estimate θ, which can guarantee the exponential convergence of the error $\tilde{\theta} \triangleq \hat{\theta} - \theta$ subject to the satisfaction of FE conditions, and here $\hat{\theta}$ denotes the estimate of θ. First, we construct the following filtered states for the original system in (3) and (4):

$$\dot{x}_{f2}(t) = -\alpha x_{f2}(t) + x_2(t),\ x_{f2}(0) = (1/\alpha)x_2(0) \tag{24}$$

$$\dot{u}_f(t) = -\alpha u_f(t) + u(t),\ u_f(0) \in \mathbb{R}^n \tag{25}$$

$$\dot{W}_f(t) = -\alpha W_f(t) + W_r(t),\ W_f(0) \in \mathbb{R}^{n\times m} \tag{26}$$

where $\alpha > 0$ is a user-defined filtering gain. The new regressor matrix W_r satisfies $W_r\theta = \dot{M}\dot{x}_{f2} - C(x_1, x_2)x_2 - g(x_1, x_2)$. We emphasize again that, based on the parameter-affine property of the EL system, the regressor matrix W_r is available for the estimator design, and its specific expression can be obtained by taking Jacobian with respect to θ.

Then substituting Eqs. (24)-(26) into (4) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(M\dot{x}_{f2} - W_f\theta - u_f) = -\alpha(M\dot{x}_{f2} - W_f\theta - u_f) \tag{27}$$

which renders $M\dot{x}_{f2} = W_f\theta + u_f + \gamma$, where $\gamma(t) = \gamma(0)\mathrm{e}^{-\alpha t}$. By the initial conditions given in (24)-(26), we have $\gamma(0) = 0_n$. Therefore,

$$M\dot{x}_{f2} = W_f\theta + u_f \tag{28}$$

An important feature of the filtered dynamics in (28) is that $\dot{x}_{f2}$ is an available signal (note that $\dot{x}_2$ usually cannot be measured). From (28), one can see that $u_f = W_I\theta$, where W_I is another auxiliary regression matrix satisfying $W_I\theta = M\dot{x}_{f2} - W_f\theta$. This indicates the filtered control input u_f actually contains the information of the uncertain parameter vector θ.

As mentioned in the introduction, the parameter estimator needs to obey some specific constraints. In this paper, we consider the situation that the uncertain parameters lie within certain bounds: $\theta_k \in (\theta_{k,\min}, \theta_{k,\max})$, $k = 1, 2, ..., m$, where θ_k denotes the entry of θ, and $\theta_{k,\min}$ and $\theta_{k,\max}$ are respectively the lower and upper bounds of θ_k. Then, consider the following projection law:

$$\theta_k = (\theta_{k,\max} - \theta_{k,\min})\mathrm{sig}(\psi_k) + \theta_{k,\min} \tag{29}$$

where $\mathrm{sig}(\psi_k) \triangleq 1/(1 + \mathrm{e}^{-\psi_k})$. One can see that, under this projection, the constrained parameter estimation problem of θ_k on the interval $(\theta_{k,\min}, \theta_{k,\max})$ is transformed to the unconstrained one of ψ_k on $\mathbb{R}$, for $k = 1, 2, ..., m$. Then, a novel parameter estimator is proposed in the following theorem.

Theorem 1: Considering the EL system described by (3) and (4), the filtered states defined in (24)-(26), and the projection law in Eq. (29), design the following estimator

$$\begin{aligned}\dot{\hat{\psi}}(t) = &-k_1[W_I^{\mathrm{T}}(t)W_I(t)\hat{\theta}(t) - W_I^{\mathrm{T}}(t)u_f(t)]\\ &- k_2 Y_W \sum_{i=1}^{q}[W_I^{\mathrm{T}}(t_i)W_I(t_i)\hat{\theta}(t) - W_I^{\mathrm{T}}(t_i)u_f(t_i)]\end{aligned} \tag{30}$$

where $\hat{\psi} = [\hat{\psi}_1, \hat{\psi}_2, ..., \hat{\psi}_m]^{\mathrm{T}}$ is the estimate of $\psi = [\psi_1, \psi_2, ..., \psi_m]^{\mathrm{T}}$; t_i denotes a set of past time indexes with $0 \le t_i \le t$, $i = 1, 2, ...q$, and q is a constant which denotes the total number of historical data points; Y_W is defined by

$$Y_W = \begin{cases} Y_\theta^{-1}, & \text{if } Y_\theta \text{ is full-rank} \\ k_\lambda I_{m\times m}, & \text{otherwise} \end{cases} \tag{31}$$

and here $Y_\theta \triangleq \sum_{i=1}^{q} W_I^{\mathrm{T}}(t_i)W_I(t_i)$ is employed for ease of notation; k_I, k_1, k_2, and k_λ are user-defined positive constants. Set the estimate of every θ_k to be:

$$\hat{\theta}_k = (\theta_{k,\max} - \theta_{k,\min})\mathrm{sig}(\hat{\psi}_k) + \theta_{k,\min} \tag{32}$$

Then one has

1) $\tilde{\psi}, \tilde{\theta} \in \mathcal{L}_\infty$, where $\tilde{\psi} \triangleq \hat{\psi} - \psi$;
2) $\forall t \ge 0$, $\hat{\theta}_k(t) \in (\theta_{k,\min}, \theta_{k,\max})$;
3) If Y_θ is full-rank, $\tilde{\theta}(t)$ exponentially converges to zero.

Proof: Consider the following storage function,

$$\begin{aligned}V_I \triangleq \sum_{k=1}^{m}\{(\theta_{k,\max} - \theta_{k,\min})[\tilde{\psi}_k + \ln(1 + \mathrm{e}^{-\tilde{\psi}_k - \psi_k})\\ - \tilde{\psi}_k\mathrm{sig}(\psi_k) - \ln(1 + \mathrm{e}^{-\psi_k})]\}\end{aligned} \tag{33}$$

where $\tilde{\psi}_k \triangleq \hat{\psi}_k - \psi_k$ denotes the entry of $\tilde{\psi}$. Then it can be readily verified that $V_I \to +\infty$ when $\tilde{\psi} \to \pm\infty$, and $\tilde{\psi} = 0_m$ is the unique and global minimizer of V_I. Thus V_I is a valid Lyapunov function candidate of $\tilde{\psi}$. Taking the time derivative for both side of (33) yields

$$\dot{V}_I = \frac{\partial V_I}{\partial \tilde{\psi}}\dot{\tilde{\psi}} = \sum_{k=1}^{m}(\theta_{k,\max} - \theta_{k,\min})(\mathrm{sig}(\hat{\psi}_k) - \mathrm{sig}(\psi_k))\dot{\hat{\psi}}_k \tag{34}$$

By (29) and (32) , one has $\tilde{\theta}_k = \hat{\theta}_k - \theta_k = (\theta_{k,\max} - \theta_{k,\min})(\mathrm{sig}(\hat{\psi}_k) - \mathrm{sig}(\psi_k))$. Thus

$$\dot{V}_I = \tilde{\theta}^{\mathrm{T}}\dot{\hat{\psi}} \tag{35}$$

Recall the fact that $u_f = W_I\theta$, $\dot{\hat{\psi}}$ in (30) is equivalent to

$$\dot{\hat{\psi}}(t) = -k_1 W_I^{\mathrm{T}}(t)W_I(t)\tilde{\theta}(t) - k_2 Y_W Y_\theta \tilde{\theta}(t) \tag{36}$$

Thus

$$\dot{V}_I = -k_1\|W_I(t)\tilde{\theta}(t)\|^2 - k_2\tilde{\theta}^{\mathrm{T}}(t)Y_W Y_\theta\tilde{\theta}(t) \tag{37}$$

Eq. (37) shows $V_I \in \mathcal{L}_\infty$, and this ensures $\tilde{\psi} \in \mathcal{L}_\infty$. Then, based on the projection law in (32), we have $\tilde{\theta} \in \mathcal{L}_\infty$ and $\hat{\theta}_k(t) \in (\theta_{k,\min}, \theta_{k,\max})$ for all $t \ge 0$, $k = 1, 2, ..., m$. Thus the first and second statements in the theorem are proved.

Then assume that adequate data is collected in Y_θ after $t \ge t^*$, i.e. Y_θ is full-rank. To show the exponential convergence of $\tilde{\theta}$ under this condition, first we state that V_I satisfies the following property:

$$c_{\min}\|\tilde{\theta}(t)\|^2 \le V_I(t) \le c_{\max}\|\tilde{\theta}(t)\|^2 \tag{38}$$

where

$$c_{\min} \triangleq \inf_{t\ge 0, k=1,2,...,m}\left[\frac{(1 + \mathrm{e}^{-\tilde{\psi}_k(t)-\psi_k})^2}{(\theta_{k,\max} - \theta_{k,\min})\mathrm{e}^{-\tilde{\psi}_k(t)-\psi_k}}\right],$$

$$c_{\max} \triangleq \sup_{t\ge 0, k=1,2,...,m}\left[\frac{(1 + \mathrm{e}^{-\tilde{\psi}_k(t)-\psi_k})^2}{(\theta_{k,\max} - \theta_{k,\min})\mathrm{e}^{-\tilde{\psi}_k(t)-\psi_k}}\right].$$

Based on the result that $\tilde{\psi} \in \mathcal{L}_\infty$, one has $c_{\min}$ and $c_{\max}$ are bounded and positive constants. Then, defining an auxiliary variable as $K(\tilde{\psi}_k) \triangleq c_{\max}\tilde{\theta}_k^2 - (\theta_{k,\max} - \theta_{k,\max})[\tilde{\psi}_k + \ln(1 + \mathrm{e}^{-(\tilde{\psi}_k + \psi_k)}) - \tilde{\psi}_k\mathrm{sig}(\psi_k) - \ln(1 + \mathrm{e}^{-\psi_k})]$, one has

$$\frac{\partial K}{\partial \tilde{\psi}_k} = \tilde{\theta}_k\left[\frac{2c_{\max}(\theta_{k,\max} - \theta_{k,\min})\mathrm{e}^{-\tilde{\psi}-\psi_k}}{(1 + \mathrm{e}^{-\tilde{\psi}-\psi_k})^2} - 1\right] \tag{39}$$

Eq. (39) shows that when $\tilde{\psi}_k \le 0$, $\partial K/\partial\tilde{\psi}_k \le 0$ and vice versa. Thus $\tilde{\psi} = 0$ is the global minimizer of $K(\tilde{\psi}_k)$, with the minimal value $K(0) = 0$. This result verifies the right-hand side of (38) (through summing up $K(\tilde{\psi}_k)$ for all k), and the left-hand side of (38) can also be proved by similar analysis.

Thus, when Y_θ is full-rank, Eqs. (31), (37) and (38) lead to

$$\dot{V}_I = -k_1\|W_I(t)\tilde{\theta}(t)\|^2 - k_2 k_\lambda\|\tilde{\theta}(t)\|^2 \le -\frac{k_2 k_\lambda}{c_{\max}}V_I \tag{40}$$

Therefore,

$$V_I \le V_I(0)\mathrm{e}^{-k_2 k_\lambda (t-t^*)/c_{\max}} \tag{41}$$

Recall (38), one has

$$\|\tilde{\theta}\|^2 \le \frac{V_I(0)}{c_{\min}}\mathrm{e}^{-k_2 k_\lambda (t-t^*)/c_{\max}} \tag{42}$$

So $\tilde{\theta}$ exponential converges to zero. The proof is complete.

Remark 1: Without employing the past measurements (which are introduced by the matrix Y_θ), one can only guarantee $\lim_{t\to\infty} W_I(t)\tilde{\theta}(t) = 0_n$ by analyzing $\dot{V}_I$. Then, to show the convergence of $\tilde{\theta}$, W_I is required to satisfy the PE condition as in the definition 1. In contrast, in our design, we ensures the exponential convergence of $\tilde{\theta}$ once Y_θ is full-rank. This only requires W_I to satisfy the FE condition, which is a significant relaxation when compared with PE conditions that are usually required in the conventional estimator/identifier design.

Remark 2: It should be emphasized that the idea to employ both real-time data and past measurements in the estimator (30) is inspired by the CL technique [38], [39], [40]. To ensure Y_θ is full-rank if W_I satisfies the FE condition, a simplest way is to add all incoming data of W_I into Y_θ until $\text{rank}(Y_\theta) = m$. A more sophisticated method is to design a selection algorithm for Y_θ, some examples are shown in [38], [39], [40].

Remark 3: Compared with the CL technique, the estimator proposed in the present paper has two main distinctions. First, by introducing a special projection law, the estimates are restricted to pre-determined bounds during the whole identification process. Second, by employing the filtered states and a two-layer regression structure, the proposed estimator does not require the information of any immeasurable derivatives.

Remark 4: As discussed in the introduction, the proposed estimator, to some extent, ensures the prescribed performance of $\tilde{\theta}$. This is from two aspects, as illustrated in Fig. 2. First, the estimator ensures $\hat{\theta}_k(t) \in (\theta_{k,\min}, \theta_{k,\max})$, $k = 1, 2, ..., m$, for all $t \geq 0$. Second, from Eq. (42), we have $\|\tilde{\theta}(t)\| \leq \kappa_2 \mathrm{e}^{-\kappa_1 t}$, where $\kappa_1 = k_2 k_\lambda/(2c_{\max})$ and $\kappa_2 = \sqrt{V_I(0)/c_{\min}}\mathrm{e}^{-\kappa_1 t^*/2}$. One can see that the convergence rate κ_1 can be increased by k_2 and k_λ, while a limitation is that κ_2 cannot be pre-specified since $V_I(0)$ is unknown.

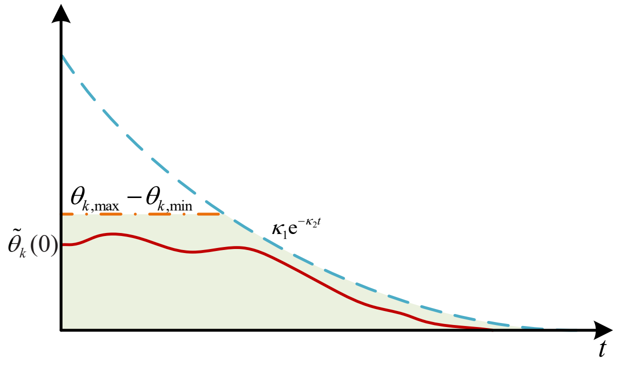

Figure 2: Prescribed performance illustration of the estimator.

B. Design of a Critic-Only ADP with Prescribed Performance

To approximate the optimal cost function V^* and the corresponding optimal control policy, a critic-only ADP controller is designed in this subsection. Based on the Weierstrass approximation theorem [43], [44], V^* can be reconstructed by

$$V^*(X) = w^{\mathrm{T}}\sigma(X) + \epsilon(X) \tag{43}$$

for $X \in \mathcal{X}$, where $\mathcal{X}$ is a compact set. Here $\sigma(X) = [\sigma_1(X), \sigma_2(X), ..., \sigma_p(X)]^{\mathrm{T}} \in \mathbb{R}^p$, and σ_i denotes a set of basis functions with $\sigma_i(0) = 0$ and $\dot{\sigma}_i(0) = 0$, $i = 1, 2, ..., p$, $w \in \mathbb{R}^p$ is a unknown weigh vector, and $\epsilon(X)$ is the reconstruction error. This kind of reconstruction method has been commonly employed in relevant results [20], [29], [45].

Then the optimal policy follows

$$\mu^* = -\frac{1}{2}R^{-1}G^{\mathrm{T}}(\nabla_X\sigma \cdot w + \nabla_X\epsilon) \tag{44}$$

Since w is unknown, an auxiliary weight vector $\hat{w}$ is employed to denote the estimate of w, and the corresponding approximation of V^* and μ^* are

$$V = \hat{w}^{\mathrm{T}}\sigma \tag{45}$$

$$\mu = -\frac{1}{2}R^{-1}\hat{G}^{\mathrm{T}}\nabla_X\sigma \cdot \hat{w} \tag{46}$$

where $\hat{G}$ denotes the estimate of G (in which uncertain parameters are replaced by their estimates). Then we consider the following Bellman error:

$$\delta_b \triangleq \nabla_X^{\mathrm{T}} V[\hat{F} + \hat{G}\mu] + r + \mu^{\mathrm{T}}R\mu \tag{47}$$

where $\hat{F}$ is the estimate of F (uncertain parameters are replaced by their estimates). Then recall (21), one has

$$\begin{aligned}\delta_b =& \nabla_X^{\mathrm{T}} V[\hat{F} + \hat{G}\mu] + \mu^{\mathrm{T}}R\mu \\ &- \nabla_X^{\mathrm{T}} V^*[F + G\mu^*] - \mu^{*\mathrm{T}}R\mu^* \\ =& \varpi^{\mathrm{T}}\tilde{w} + \epsilon_1 + \epsilon_H\end{aligned} \tag{48}$$

where $\tilde{w} \triangleq \hat{w} - w$ is the weight estimation error, $\epsilon_1 = w\nabla_X^{\mathrm{T}}\sigma\tilde{F} - 0.25w^{\mathrm{T}}\tilde{D}w$ is the error induced by the parameter estimation process, ϵ_H is a residual error defined same with [31], [46], [47]. Besides, $\varpi = \nabla_X^{\mathrm{T}}\sigma(\hat{F} - 0.5\hat{P}\nabla_X\sigma \cdot \hat{w})$, $\hat{P} \triangleq \hat{G}R^{-1}\hat{G}^{\mathrm{T}}$ and $\hat{D} \triangleq \nabla_\eta^{\mathrm{T}}\sigma\hat{P}\nabla_\eta\sigma$ are employed for ease of notation, and $\tilde{F} \triangleq \hat{F} - F$ and $\tilde{D} \triangleq \hat{D} - D$ with $D \triangleq \nabla_\eta^{\mathrm{T}} GR^{-1}G^{\mathrm{T}}\nabla_\eta\sigma$. Since δ_b contains the information of $\tilde{w}$, it has been commonly-used in the update law of $\dot{\hat{w}}$. Before proposing the specific design of $\dot{\hat{w}}$, the following assumptions are made for $X \in \mathcal{X}$.

Assumption 1: The reconstruction error, its gradient, and the residual error are bounded, such that

$$\|\epsilon\| \leq b_\epsilon, \ \|\nabla_X\epsilon\| \leq b_{\epsilon_x}, \|\epsilon_H\| \leq b_{\epsilon_H} \tag{49}$$

Assumption 2: The basis functions and their gradient are bounded and satisfy

$$\|\sigma\| \leq b_\sigma, \ \|\nabla_X\sigma\| \leq b_{\sigma_x} \tag{50}$$

Assumption 3: Under the condition $\tilde{\theta} \in \mathcal{L}_\infty$, the error induced by the parameter estimator satisfies

$$\|\varpi\epsilon_1/(1 + \varpi^{\mathrm{T}}\varpi)^2\| \leq b_{\epsilon_1} \tag{51}$$

Here b_ϵ, b_{ϵ_x}, b_{ϵ_H}, b_σ, b_{σ_x}, b_{ϵ_1} are all positive constants.

Although the Bellman error is usually employed to update $\hat{w}$, the precise estimation requires that ϖ satisfies the PE condition in conventional ADP methods [22], [24], [25], [26]. This condition is quite strong and almost infeasible in real engineering applications, especially for online cases. Motivated by the results given in [30], [31], [32], past measurements (which are collected online) are introduced into the critic update law in this paper to relax the excitation condition. For ease of notations, We denote $\varsigma = \varpi^{\mathrm{T}}\varpi + 1$, $\zeta = \varpi/(\varpi^{\mathrm{T}}\varpi + 1)$, $\varphi_1 = \zeta\zeta^{\mathrm{T}}$ and $\varphi_2 = \zeta(r + \mu^{\mathrm{T}}R\mu)/\varsigma$, and also consider two time indexes t_{w1} and t_{w2} with $0 \leq t_{w1} \leq t_{w2} \leq t$. Then we design the following auxiliary variables:

$$\dot{\xi}_1(t, t_{w1}) = -\kappa\xi_1(t, t_{w1}) + \varphi_1(t), \quad \xi_1(t_{w1}) = 0_{p\times p} \tag{52}$$

$$\dot{\xi}_2(t, t_{w1}) = -\kappa\xi_2(t, t_{w1}) + \varphi_2(t), \quad \xi_2(t_{w1}) = 0_{p\times 1} \tag{53}$$

where $\kappa > 0$ is a user-defined constant. We also define $\Xi(t, t_{w2}, t_{w1}) \triangleq \xi_1(t_{w2}, t_{w1})\hat{w}(t) + \xi_2(t_{w2}, t_{w1})$. Then, following (52) and (53), one has

$$\Xi(t, t_{w2}, t_{w1}) = \mathrm{e}^{-\kappa t_{w2}} \int_{t_{w1}}^{t_{w2}} \mathrm{e}^{\kappa\tau}(\varphi_1(\tau)\hat{w}(t) + \varphi_2(\tau))\mathrm{d}\tau \tag{54}$$

By the definition of φ_1 and φ_2, one has

$$\begin{aligned}\Xi(t, t_{w2}, t_{w1}) =& \mathrm{e}^{-\kappa t_{w2}} \int_{t_{w1}}^{t_{w2}} \frac{\mathrm{e}^{\kappa\tau}\zeta(\tau)}{\varsigma(\tau)} \cdot \\ & \left(\varpi^{\mathrm{T}}(\tau)\hat{w}(t) + r(\tau) + \mu^{\mathrm{T}}(\tau)R\mu(\tau)\right)\mathrm{d}\tau \\ =& \left(\int_{t_{w1}}^{t_{w2}} \mathrm{e}^{-\kappa(t_{w2}-\tau)}\zeta\zeta^{\mathrm{T}}\mathrm{d}\tau\right)\tilde{w}(t) + \bar{\epsilon}_1 \\ =& \xi_1(t_{w2}, t_{w1})\tilde{w}(t) + \bar{\epsilon}_1\end{aligned} \tag{55}$$

where

$$\bar{\epsilon}_1 \triangleq \int_{t_{w1}}^{t_{w2}} \frac{\mathrm{e}^{-\kappa(t_{w2}-\tau)}\varpi(\tau)[\epsilon_1(\tau) + \epsilon_H(\tau)]}{[\varpi^{\mathrm{T}}(\tau)\varpi(\tau) + 1]^2}\mathrm{d}\tau$$

is an integral residual error, and $\bar{\epsilon}_1 \in \mathbb{R}^p$.

Based on these designs, an approximate optimal control strategy with prescribed performance is summarized in the following theorem.

Theorem 2: Consider the augmented model in (18), the parameter estimator in (30) subject to the satisfaction of a FE condition of W_I, and the critic-only ADP control structure in Eqs. (45) and (46). Assume ζ satisfies a FE condition, i.e., there exist t_{w1}, t_{w2}, c_w with $0 \leq t_{w1} \leq t_{w2} \leq t$ and $c_w > 0$ such that

$$\int_{t_{w1}}^{t_{w2}} \zeta(\tau)\zeta^{\mathrm{T}}(\tau)\mathrm{d}\tau \geq c_w I_{p\times p} \tag{56}$$

Design the update law for the critic NN to be

$$\dot{\hat{w}}(t) = -c_1\frac{\varpi(t)\delta_b(t)}{(\varpi^{\mathrm{T}}(t)\varpi(t) + 1)^2} - c_2\Xi(t, t_{w2}, t_{w1}) \tag{57}$$

where c_1 and c_2 are positive constants. Then, if the initial state is in the admissible domain $\mathfrak{D}_z$, the weight estimation error $\tilde{w}$ and the transformed tracking error e and z_2 are UUB, and all prescribed specifications on z_1 are guaranteed.

Proof : Consider the following storage function

$$L \triangleq V^* + \frac{\rho_1}{2}\tilde{w}^{\mathrm{T}}\tilde{w} \tag{58}$$

where ρ_1 is a positive constant employed just for stability analysis. The time derivative of L is analyzed in (59), in which the arithmetic-geometric mean inequality is employed. Then by setting the auxiliary parameters ρ_1 to satisfy:

$$\rho_1 > \frac{2\mathrm{e}^{t_{w2}-t_{w1}} \cdot \max_{t\geq 0}\{\|D\|\}}{c_2 c_w} \tag{60}$$

one has

$$\dot{L} \leq -e^{\mathrm{T}}Q_e e - z_2^{\mathrm{T}}Q_z z_2 - \frac{\rho_2}{2}\|\tilde{w}\|^2 - \rho_1 c_1\|\zeta^{\mathrm{T}}\tilde{w}\| + \rho_3 \tag{61}$$

where $\rho_2 = \|\rho_1 c_2 c_w \mathrm{e}^{-(t_{w2}-t_{w1})}I_{p\times p} - D\|$, and $\rho_3 = \|(w^{\mathrm{T}}\nabla_X^{\mathrm{T}}\sigma + \nabla_X^{\mathrm{T}}\epsilon)GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta}\| + 0.5\nabla_X^{\mathrm{T}}\epsilon GR^{-1}G^{\mathrm{T}}\nabla_X\epsilon + \rho_1 c_1\|\epsilon_1 + \epsilon_H\|^2/[2(\varpi^{\mathrm{T}}\varpi + 1)^2] + \rho_1 c_2\|\bar{\epsilon}_1\|^2\mathrm{e}^{(t_{w2}-t_{w1})}/(2c_w)$. By the assumptions 1-3 and the condition given in (60), one has $\rho_2(t) > 0$ for $t \geq 0$, and $\rho_3 \in \mathcal{L}_\infty$. This directly guarantees the UUB of e, z_2, and $\tilde{w}$. Finally, the fact $e \in \mathcal{L}_\infty$ ensures that all prescribed performance specifications are guaranteed. This completes the proof.

Remark 5: By utilizing past measurements and introducing the vector Ξ into $\dot{\hat{w}}$, the stability of the closed-loop system is guaranteed under the satisfaction of the FE condition of ζ, which is much weaker than the conventional PE conditions. It should be emphasized that, also for PE relaxation purpose, CL-based ADP methods as in [30], [31], [32] employ discrete historical data stack in the update laws of NN weights: $\sum_{i=1}^{l}(\zeta(i)/\varsigma(i))\hat{\delta}_b(i)$, where $\hat{\delta}_b(i)$ denotes the Bellman error at a past (discrete) time point i while replacing $\hat{w}(i)$ with its real-time counterpart $\hat{w}(t)$. This kind of method, however, requires complicated online selection algorithms to update the data stack. A commonly-used example of such algorithms maximizes the minimum singular value of $\sum_{i=1}^{l}\zeta(i)\zeta^{\mathrm{T}}(i)$, by swapping the most recent incoming data with every recorded data and then comparing the corresponding minimum eigenvalues (with a complexity of $\mathcal{O}(l \cdot p^3)$ at every discrete time point, where l is the length of the history stack, and p denotes the dimension of $\sum_{i=1}^{l}\zeta(i)\zeta^{\mathrm{T}}(i)$). In contrast, the method proposed in this paper takes advantage of all incoming data by designing a special integral-form information matrix ($\Xi(t, t_{w2}, t_{w1})$), this new approach is arguably easier to implement and is computationally cheaper.

Remark 6: Note that the time interval $[t_{w1},\ t_{w2}]$ can be changed during the control process. A straightforward and easily implementable strategy is that: 1) First set $t_{w1} = 0$, and once ζ is FE on $[0, t]$, fix t_{w2} to this time point and accordingly construct $\Xi(t, t_{w2}, 0)$. 2) As time goes on, if ζ satisfies a FE condition on a new time interval $[\bar{t}_{w1},\ \bar{t}_{w2}]$, one can replace $\Xi(t, t_{w2}, 0)$ with $\Xi(t, \bar{t}_{w2}, \bar{t}_{w1})$ to potentially reduce the residual set caused by $\bar{\epsilon}_1$.

Remark 7: It is noteworthy that an adaptive tracking controller was proposed in [15] for Euler-Lagrange systems, which also guarantees the prescribed performance of the coordinate tracking error, and a NN is employed to approximate the drift dynamics. A limitation of this elegant result is that its controller requires the accurate information of the inertia matrix M. In contrast, by fully employing the parameter affine property of Euler-Lagrange systems, all system parameters can be uncertain for the constrained estimator and the estimator-based ADP controller proposed in the present paper. More importantly, our control scheme has the ability to make a trade-off between performance and control cost, reducing the control cost while strictly guaranteeing the performance specifications. These advantages are also illustrated by numerical simulations.

$$
\begin{aligned}
\dot{L} =&\nabla_X^{\mathrm{T}} V^*[F + G\mu - \frac{1}{2}GR^{-1}G^{\mathrm{T}}(\nabla_X\sigma \cdot w + \nabla_X\epsilon) + \frac{1}{2}GR^{-1}G^{\mathrm{T}}(\nabla_X\sigma \cdot w + \nabla_X\epsilon) + GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta}] + \rho_1 \tilde{w}^{\mathrm{T}}\dot{\hat{w}} \\
=&\nabla_X^{\mathrm{T}} V^*(F + G\mu^*) + (w^{\mathrm{T}}\nabla_X^{\mathrm{T}}\sigma + \nabla_X^{\mathrm{T}}\epsilon)[-\frac{1}{2}GR^{-1}G^{\mathrm{T}}\nabla_X\sigma\tilde{w} + \frac{1}{2}GR^{-1}G^{\mathrm{T}}\nabla_X\epsilon + GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta}] + \rho_1\tilde{w}^{\mathrm{T}}\dot{\hat{w}} \\
=&-r - \frac{1}{4}w^{\mathrm{T}}Dw - \frac{1}{2}w^{\mathrm{T}}D\tilde{w} - \frac{1}{2}\nabla_X^{\mathrm{T}}\epsilon GR^{-1}G^{\mathrm{T}}\nabla_X\sigma\tilde{w} + \frac{1}{4}\nabla_X^{\mathrm{T}}\epsilon GR^{-1}G^{\mathrm{T}}\nabla_X\epsilon \\
&-(w^{\mathrm{T}}\nabla_X^{\mathrm{T}}\sigma + \nabla_X^{\mathrm{T}}\epsilon)GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta} + \rho_1\tilde{w}^{\mathrm{T}}\dot{\hat{w}} \\
\leq&-r + \frac{1}{2}\tilde{w}^{\mathrm{T}}D\tilde{w} + \frac{1}{2}\nabla_X^{\mathrm{T}}\epsilon GR^{-1}G^{\mathrm{T}}\nabla_X\epsilon + \|(w^{\mathrm{T}}\nabla_X^{\mathrm{T}}\sigma + \nabla_X^{\mathrm{T}}\sigma)GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta}\| + \rho_1\tilde{w}^{\mathrm{T}}\dot{\hat{w}} \\
\leq&-r - \tilde{w}^{\mathrm{T}}[\rho_1c_2\xi_1(t_{w1}, t_{w2}) + \rho_1c_1\zeta\zeta^{\mathrm{T}} - D]\tilde{w} + \frac{1}{2}\nabla_X^{\mathrm{T}}\epsilon GR^{-1}G^{\mathrm{T}}\nabla_X\epsilon + \|(w^{\mathrm{T}}\nabla_X^{\mathrm{T}}\sigma + \nabla_X^{\mathrm{T}}\epsilon)GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta}\| \\
&+ \frac{\rho_1c_1\tilde{w}^{\mathrm{T}}\varpi(\epsilon_1 + \epsilon_H)}{(\varpi^{\mathrm{T}}\varpi + 1)^2} + \rho_1c_2\tilde{w}^{\mathrm{T}}\bar{\epsilon_1} \\
\leq&-e^{\mathrm{T}}Q_ee - z_2^{\mathrm{T}}Q_zz_2 - \frac{1}{2}\tilde{w}^{\mathrm{T}}[\rho_1c_2c_w\mathrm{e}^{-(t_{w2}-t_{w1})}I_{p\times p} + \rho_1c_1\zeta\zeta^{\mathrm{T}} - 2D]\tilde{w} + \|(w^{\mathrm{T}}\nabla_X^{\mathrm{T}}\sigma + \nabla_X^{\mathrm{T}}\epsilon)GH(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta}\| \\
&+ \frac{1}{2}\nabla_X^{\mathrm{T}}\epsilon GR^{-1}G^{\mathrm{T}}\nabla_X\epsilon + \frac{\rho_1c_1\|\epsilon_1 + \epsilon_H\|^2}{2(\varpi^{\mathrm{T}}\varpi + 1)^2} + \frac{\rho_1c_2\|\bar{\epsilon}_1\|^2\mathrm{e}^{(t_{w2}-t_{w1})}}{2c_w}
\end{aligned}
\tag{59}
$$

IV. Numerical Simulations

To verify the effectiveness of the proposed controller, the model of a typical two-link robot manipulator as in [48] is employed to carry out numerical simulation tests, which is defined by

$$M\ddot{q} + (V_m + F_d)\dot{q} + F_s = u \tag{62}$$

where $q = [q_1, q_2]^{\mathrm{T}} \in \mathbb{R}^2$ and $\dot{q} = [\dot{q}_1, \dot{q}_2]^{\mathrm{T}} \in \mathbb{R}^2$ are respectively the angular positions and velocities,

$$M = \begin{bmatrix} p_1 + 2p_3\cos(q_2) & p_2 + p_3\cos(q_2) \\ p_2 + p_3\cos(q_2) & p_2 \end{bmatrix},$$

$$V_m = \begin{bmatrix} -p_3\sin(q_2)\dot{q}_2 & -p_3\sin(q_2)(\dot{q}_1 + \dot{q}_2) \\ p_3\sin(q_2)\dot{q}_1 & 0 \end{bmatrix},$$

$$F_d = \begin{bmatrix} p_4 & 0 \\ 0 & p_5 \end{bmatrix},\ F_s = \begin{bmatrix} p_6\tanh(\dot{q}_1) \\ p_7\tanh(\dot{q}_2) \end{bmatrix},$$

and $p_1 = 3.473\text{kg·m}^2$, $p_2 = 0.196\text{kg·m}^2$, $p_3 = 0.242\text{kg·m}^2$, $p_4 = 5.3\text{Nm·s}$, $p_5 = 1.1\text{Nm·s}$, $p_6 = 8.45\text{Nm}$, and $p_7 = 2.35\text{Nm}$ are system parameters. Unlike [15], [48], we assume that all these parameters are unknown, thus $\theta = [p_1, p_2, ..., p_7]^{\mathrm{T}}$. The estimation bounds of $\hat{\theta}$ are $\theta_{\max} = [4.5, 0.25, 0.3, 7, 1.5, 10, 2.8]^{\mathrm{T}}$ and $\theta_{\min} = [3, 0, 0, 4.5, 0.5, 8, 1.8]^{\mathrm{T}}$. The reference trajectory is designed as $\dot{x}_r(t) = [\sin(0.5t), -0.25\cos(t), 0.5\cos(0.5t), 0.25\sin(t)]^{\mathrm{T}}$ with $x_r(0) = [0, -0.25, 0.5, -0.5]^{\mathrm{T}}$. The parameters of the proposed approximate optimal controller (denote as "AOC") are set to be $Q_e = 2I_{2\times2}$, $Q_z = I_{2\times2}$, $R = I_{2\times2}$, $\alpha = 0.1$, $k_1 = 0.5$, $k_2 = 10$, $k_\lambda = 1$, $c_1 = 2$, $c_2 = 1$, $\kappa = 0.02$, $t_{w1} = 0$, $t_{w2} = 2$. Following [31], [32], [48], [44], [49], polynomial basis functions are chosen to construct the critic NN:

$$
\begin{aligned}
\sigma(X) = [&e_1z_{21}, e_2z_{22}, 0.5z_{21}^2, 0.5z_{22}^2, e_1z_{22} \\
&z_{21}z_{22}, 0.5e_1^2z_{21}^2, 0.5e_2^2z_{22}^2, 0.5z_{21}^2z_{22}^2 \\
&0.5z_{21}^2x_{r21}^2, 0.5z_{22}^2x_{r22}^2]^{\mathrm{T}}
\end{aligned}
$$

and the initial weight is set to be $\hat{w}(0) = [2, 2, 10, 10, 0_{1\times7}]^{\mathrm{T}}$. This design renders

$$\mu(0) = -\hat{M}^{-1}(0)e(0) - 5\hat{M}^{-1}(0)z_2(0) \tag{63}$$

where $\hat{M}$ is the estimate of M. Please refer to Appendix A for the motivation of why such an initial controller is selected in this case study. Other initial conditions are set to be $q(0) = [3, -3]^{\mathrm{T}}$, $\dot{q}(0) = [0.1, 0.1]^{\mathrm{T}}$, $\hat{\theta}(0) = [4, 0.1, 0.1, 6, 0.8, 9, 2]^{\mathrm{T}}$. Performance specifications are designed as: $\rho_{1,0} = \rho_{2,0} = 4$, $\rho_{1\infty} = \rho_{2\infty} = 0.1$, and $c_\rho = 1$.

Besides, it should be emphasized that the control method proposed in this paper is an online ADP method, which does not require any offline or other pre-gathered data sets. While most online data is employed to calculate the control signal directly, a small part of it is fed into the information matrices in both the parameter estimator and the ADP controller, i.e. Y_θ and Ξ. The simulation runs under the fixed-step solver mode with sample time as 0.01s, and the solver is ode4 (Runge-Kutta). Under these settings, on the one hand, the data in the first 2s is employed to build Ξ ($t_{w1} = 0$s and $t_{w2} = 2$s), and it is released (set Ξ to be a zero matrix) after 10s to reduce the residual error; on the other hand, the data in the first 5s is employed by Y_θ and kept until the end of simulation.

Moreover, to show the advantages of the approximate optimal controller (denote as "AOC") proposed in the present paper, two other controllers are also employed in simulations.

1) *The immersion and invariance-based controller in [3] (denoted as I&I).* As a typical non-CE adaptive control method, I&I-based controllers have shown superior closed-loop performance than conventional adaptive controllers. The I&I-based controller proposed in [3] can also achieve tracking control objectives for Euler-Lagrange systems under system uncertainties. The specific design of this controller can be found in [3], and the control gains are chosen to be $\gamma = 2$, $k_p = 0.5$, $k_v = 2$ and $\alpha = 0.5$.

2) *The adaptive neural control with prescribed performance in [15] (denoted as PPC).* As mentioned in the Remark 7, the controller in [15] is also capable of guaranteeing prescribed performance of the coordinate tracking error. The control parameters of PPC are set to be $\Gamma = 2$ and $k = 10$.

Note that the control gains for all the three controllers have been carefully adjusted to achieve satisfactory transient performance under similar control input levels.

The time responses of $\tilde{\theta}(t)$ under AOC is shown in Fig. 3. The switch law in (31) is triggered after $t = 5$s to reduce truncation errors. Fig. 3 shows that the estimation error $\|\tilde{\theta}\|$ rapidly converge to zero under the proposed estimator. To show the constraint handling ability of the estimator, the simulation results of $\hat{\theta}_2$ and $\hat{\theta}_3$ with and without employing the projection law are illustrated in Fig. 4. One can see that, the estimator with projection renders a superior estimation performance and can restrict all estimates to the required bounds. Otherwise, the estimates go out the feasible region.

Then, the time responses of e_1 and e_2 under different controllers are demonstrated in Figs. 5 and 6, respectively. And the simulation results of z_2 are given in Fig. 7. Although all the three control approaches achieve the tracking control objective, the I&I controller cannot meet the performance specifications. Compared with the other two controllers, the PPC method in [15] leads to a rapid convergence of all tracking errors (this may benefit from its inner sliding-mode structure), while the AOC proposed in the present paper has smoother trajectories and renders fewer fluctuations. Under AOC, there is almost no overshoot and the transient trajectories are strictly within the admissible region specified by PPFs. Then the control cost ($\mu^{\mathrm{T}} R \mu$) of these controllers is given in Fig. 8. One can see that, PPC has a higher control cost than the other two methods. In contrast, AOC can make a trade-off and significantly reduces the control cost.

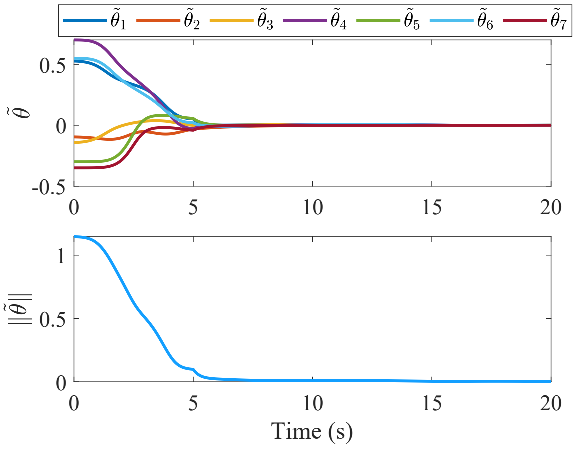

Figure 3: Time responses of $\tilde{\theta}$ under AOC.

Under the AOC, the weight vector of the critic NN converges to

$$\hat{w}_f = [4.4314, 2.6588, 9.9185, 9.8388, 0.3171, 0.5705, \\ -0.3684, -1.1564, -0.2285, -0.0282, -0.0005]^{\mathrm{T}}$$

Since the HJB equation in the simulation case cannot be precisely solved, we cannot compare $\hat{w}$ with its unknown

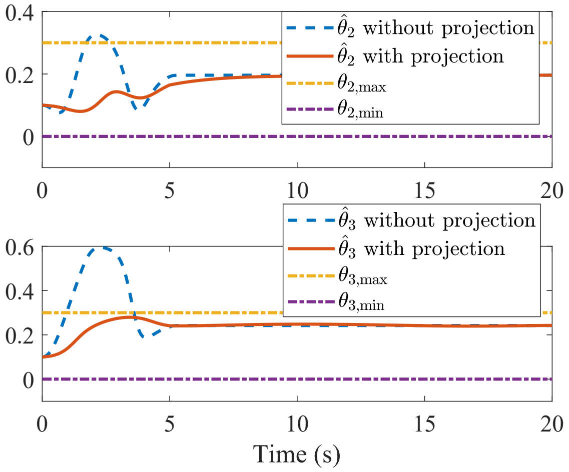

Figure 4: Time responses of $\hat{\theta}_2$ and $\hat{\theta}_3$ with and without projection.

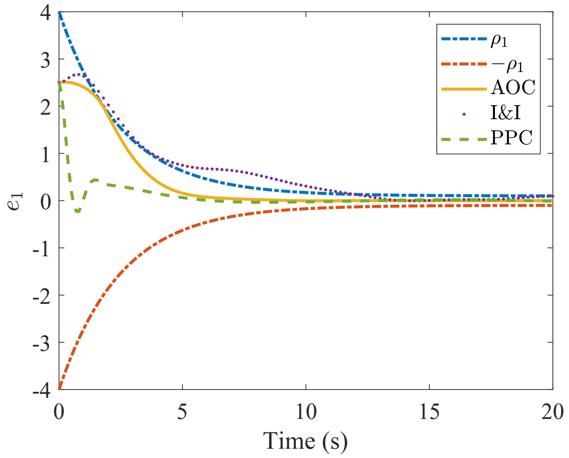

Figure 5: Time responses of e_1 under different controllers.

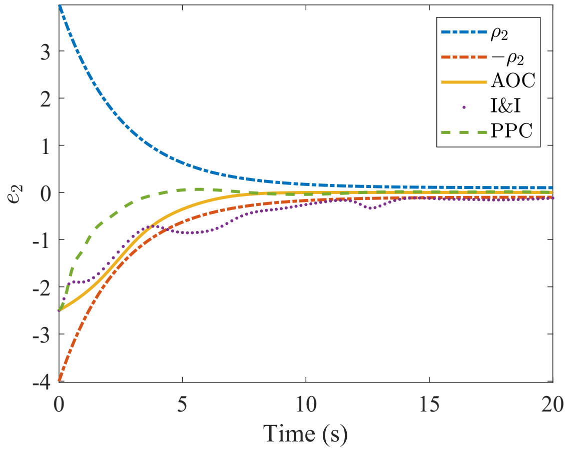

Figure 6: Time responses of e_2 under different controllers.

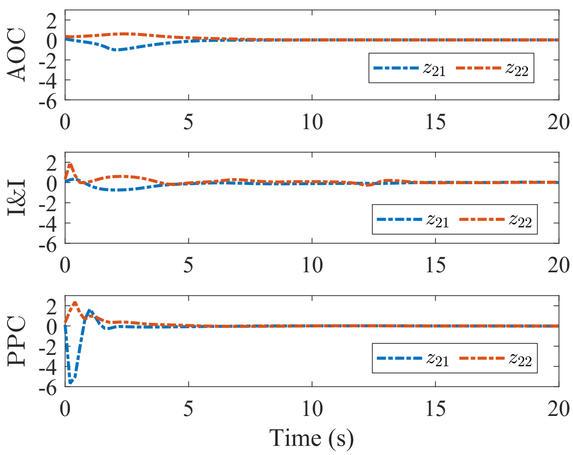

Figure 7: Time responses of z_2 under different controllers.

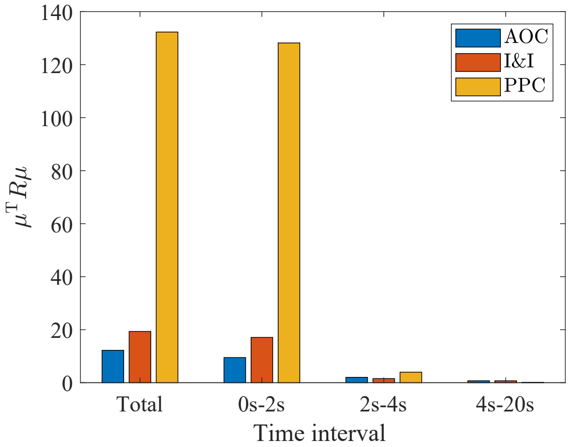

Figure 8: Control cost under different controllers.

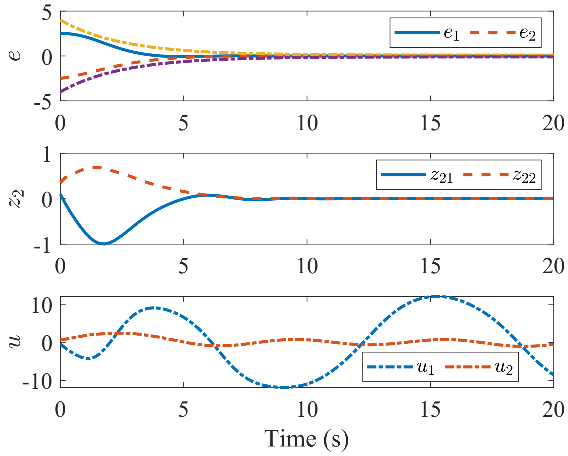

Figure 9: Simulation results under the fixed NN weights.

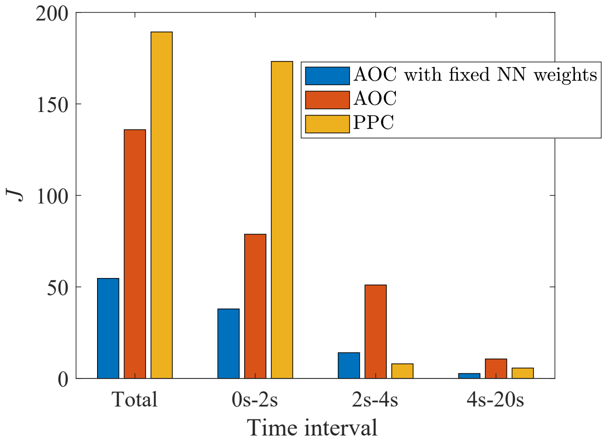

Figure 10: Simulation results of the cost index J.

true values. Instead, we carry out the simulation again while fixing the NN weights to be the estimated value $\hat{w}_f$. The corresponding results of e, z_2 and u are given in Fig. 9, and its overall cost (J) compared to the original online-tuning AOC and the PPC are shown in Fig. 10. One can see that, the finalized AOC controller with fixed weights estimated by the proposed method clearly has a superior performance than the PPC and the original online-tuning AOC, and renders a much less cost. These facts prove the approximate optimal property of the proposed control scheme in the present paper.

The computational cost of the proposed controller has also been analyzed. Theoretically, this mainly comes from two aspects. The first aspect is related to the reconstruction of the cost function V^*. Recall (43), one can see that a linear combination of basis/activation functions is employed to reconstruct the cost function. This indicates that the most intensive algebraic operation involved in the design is calculating the gradient of $\sigma(X)$ with respect to X (as shown in (44)). The second aspect is related to the information matrices in the parameter estimator and the controller, which requires integral operations but just for fixed time intervals. All the other operations involved in the proposed controller are basic addition/subtraction and multiplication/division operations, and different specifications of the prescribed performance function (which decides the transient and steady-state performance of the coordinate tracking error) have negligible influence on the computational complexity. For the simulation case considered in the paper, the proposed controller needs 5.104499 seconds to complete a simulation of 100 seconds (under a computer with Intel Core i7-9700K CPU @ 3.60GHz, 32GB RAM), while the I&I can finish the simulation in 3.603280 seconds. These analyses and results show that though the computational complexity of the proposed controller is higher than the nonlinear adaptive controller, the additional computational cost is limited and acceptable. Especially considering the fact that the proposed controller can achieve approximate optimal control (significantly reduced the control cost than advanced adaptive controllers, such as I&I) with prescribed performance. For cases when the computational resources are extremely limited, one can reduce the computational complexity of the proposed controller by reducing the number of basis functions and the data collecting time in information matrices. At the same time, a trade-off is required since these approaches may result in potential performance depredation.

To sum up, simulation results verify the effectiveness of the approximate optimal tracking controller proposed in the present paper, and the prescribed performance specifications are strictly guaranteed during the whole control process.

V. Conclusion

A novel RL-based approximate optimal tracking control scheme for uncertain Euler-Lagrange systems was proposed in this paper. Specifically, an estimator-based critic-only approximate optimal tracking controller was designed to precisely estimate unknown parameters and approximate the solution of the HJB equation. Moreover, the proposed control scheme is capable of guaranteeing prescribe performance. Uniformly ultimately bounded stability was rigorously guaranteed via Lyapunov-based analysis under relaxed excitation conditions (FE conditions). Future work would consider more general constraints on system states.

Appendix

A. An Example of the Initial Control Design

This appendix provides details regarding the selection of the initial controller in the simulation.

We aim to ensure that this initial controller can guarantee the UUB of z_1, z_2 and e. The following Lemma shows that a PD-like controller could satisfy this requirement.

Lemma 1: Consider the tracking error model of the Euler-Lagrange system in (10) and (16), design the following PD-like controller:

$$u = -k_p \hat{M}^{-1} e - k_v \hat{M}^{-1} z_2 + H(x_{r1}, x_{r2}, f_r(x_r))\hat{\theta} \tag{64}$$

and here $\hat{\theta}$ and $\hat{M}$ are updated by the parameter estimator proposed in Theorem 1. Then for all $z_1(0) \in \mathfrak{D}_z$ and $z_2(0) \in \mathbb{R}^n$, one has $e, z_1, z_2 \in \mathcal{L}_\infty$ and $z_1(t) \in \mathfrak{D}_z$ for all $t \geq 0$, subject to the satisfaction of the conditions as given in the proof.

Proof: To show the boundedness of the closed-loop system, consider the following storage function:

$$V_{pd} = e^{\mathrm{T}} e + c z_2^{\mathrm{T}} M z_2 + e^{\mathrm{T}} z_2 \tag{65}$$

where c is a positive constant. Then one has

$$V_{pd} = \left[\begin{array}{cc} e^{\mathrm{T}} & z_2^{\mathrm{T}} \end{array}\right] A \left[\begin{array}{c} e \\ z_2 \end{array}\right] \tag{66}$$

where

$$A = \left[\begin{array}{cc} I_{n\times n} & 0.5 I_{n\times n} \\ 0.5 I_{n\times n} & cM I_{n\times n} \end{array}\right]$$

Thus by setting $c > 0.25/m_m$, one has $A > 0$ and V_{pd} is a valid storage function. Then the time derivative of V_{pd} is analyzed in (67). Rearranging the last inequality in (67), one has:

$$\dot{V}_{pd} \leq -\left[\begin{array}{cc} e^{\mathrm{T}} & z_2^{\mathrm{T}} \end{array}\right] B_1 \left[\begin{array}{c} e \\ z_2 \end{array}\right] + B_2 \left[\begin{array}{c} e \\ z_2 \end{array}\right] \tag{68}$$

where

$$B_1 = \left[\begin{array}{cc} -\frac{k_p}{m_M \bar{m}_M} I_{n\times n} & \frac{b_{y2}\bar{m}_m\|\theta\|+k_v}{2m_m\bar{m}_m} + \frac{ck_p}{\bar{m}_m} + r_M \\ \frac{b_{y2}\bar{m}_m\|\theta\|+k_v}{2m_m\bar{m}_m} + \frac{ck_p}{\bar{m}_m} + r_M & -2c(\frac{k_v}{\bar{m}_M} - b_{q2}\|\theta\|) - r_M \end{array}\right]$$

and

$$B_2 = \left[\begin{array}{c} 2b_p\rho_M + (b_{y1}\rho_M\|\theta\| + b_{y3}\|\theta\| + b_H b_{\tilde{\theta}})/m_m \\ b_p\rho_M + 2c(b_{q1}\rho_M\|\theta\| + b_{q3}\|\theta\| + b_H b_{\tilde{\theta}}) \end{array}\right]^{\mathrm{T}},$$

and here $b_p = \max_{t\geq 0}\{\|\mathcal{P}(t)\|\}$, $\rho_M = \max_{i=1,2,\ldots,n}\{\rho_{i0}\}$, $b_H = \max_{t\geq 0}\{H[x_{r1}(t), x_{r2}(t), f_r(x_r(t))]\}$, $r_M = \max_{t\geq 0}\{\Upsilon(t)\}$, and $b_{\tilde{\theta}} = \max_{t\geq 0}\{\|\tilde{\theta}(t)\|\}$. Besides, $\bar{m}_m$ and $\bar{m}_M$ respectively denote the minimum and maximum eigenvalues of $\hat{M}$. Moreover, the regressor matrix $Q(z_1, z_2, x_{r1}, x_{r2})$ is employed for ease of notation. The definition of it is very similar to $Y(z_1, z_2, x_{r1}, x_{r2})$, in which only the term $C(x_1, x_2)x_2$ is replaced by $C(x_1, x_2)x_{r2}$. We assume that both Y and Q are local Lipschitz (the example system in the simulation satisfies this assumption). This indicates that there exists positive constants b_{y1}, b_{y2}, b_{y3}, b_{q1}, b_{q2}, and b_{q3}, such that $\|Y(z_1, z_2, x_{r1}, x_{r2})\| \leq b_{y1}\|z_1\| + b_{y2}\|z_2\| + b_{y3}$ and $\|Q(z_1, z_2, x_{r1}, x_{r2})\| \leq b_{q1}\|z_1\| + b_{q2}\|z_2\| + b_{q3}$. It is noteworthy that the results given in (67) and (68) rely on the following facts: 1) All the properties of the Euler-Lagrange system considered in the paper, as discussed in Sec.II.B, particularly the fact that $(M - 2C)$ is an anti-symmetric matrix. 2) $H(x_{r1}, x_{r2}, f_r(x_r))$ is bounded, since $x_{r1}, x_{r2}, f_r(x_r) \in \mathcal{L}_\infty$. 3) The parameter estimation error $\tilde{\theta}$ is bounded, as shown in Theorem 1. 4) It can be readily proved that $\rho, \dot{\rho}, \mathcal{P} \in \mathcal{L}_\infty$. 5) $\Upsilon(t)$ is always positive-definite and bounded if $z_1(t) \in \mathfrak{D}$ and $X \in \mathcal{X}$. 6) When $z_1 \in \mathfrak{D}_z$, one has $\|z_1(t)\| \leq \rho_M$. 7) We assume $\hat{M}$ is positive-definite since $\tilde{\theta}$ exponentially converges to zero.

Eq. (68) shows that, by setting c, k_p, and k_v to ensure $B_1 > 0$, one has e, z_1 and z_2 are UUB. Based on the definition of e, one has $z_1(t) \in \mathfrak{D}_z$ for all $t \geq 0$. The proof is complete.

We emphasize that the requirement $B_1 > 0$ is a sufficient condition which is deduced by considering extreme situations. In practice, the margins of the initial control parameters are actually much larger and can be set empirically.

A remaining issue is representing the initial controller by a set of basis functions. Since the term introduced by the parameter estimation error $(H(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta})$ has been considered in Theorem 2, one can see that designing an initial controller following (64) is equivalent to design μ as:

$$\mu = -k_p \hat{M}^{-1} e - k_v \hat{M}^{-1} z_2 \tag{69}$$

For the specific simulation scenario in this paper, the basis functions are set to be

$$\begin{aligned} \sigma(X) = [&e_1 z_{21}, e_2 z_{22}, 0.5z_{21}^2, 0.5z_{22}^2, e_1 z_{22} \\ &z_{21} z_{22}, 0.5e_1^2 z_{21}^2, 0.5e_2^2 z_{22}^2, 0.5z_{21}^2 z_{22}^2 \\ &0.5z_{21}^2 x_{r21}^2, 0.5z_{22}^2 x_{r22}^2]^{\mathrm{T}} \end{aligned}$$

with initial guess of w to be $\hat{w}(0) = [2, 2, 10, 10, 0_{1\times 7}]^{\mathrm{T}}$. This leads to

$$\mu(0) = -\hat{M}^{-1}(0)e(0) - 5\hat{M}^{-1}(0)z_2(0) \tag{70}$$

$$\begin{aligned}\dot{V}_{pd} =&2e^{\mathrm{T}}\dot{e} + 2cz_2^{\mathrm{T}}M\dot{z}_2 + cz_2^{\mathrm{T}}\dot{M}z_2 + \dot{e}^{\mathrm{T}}z_2 + e^{\mathrm{T}}M^{-1}M\dot{z}_2 \\ =&2e^{\mathrm{T}}(\Upsilon z_2 - \mathcal{P}z_1) + 2cz_2^{\mathrm{T}}[Q(z_1, z_2, x_{r1}, x_{r2})\theta + H(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta} - k_p\hat{M}^{-1}e - k_v\hat{M}^{-1}z_2] \\ &+ (\Upsilon z_2 - \mathcal{P}z_1)^{\mathrm{T}}z_2 + e^{\mathrm{T}}M^{-1}[Y(z_1, z_2, x_{r1}, x_{r2})\theta + H(x_{r1}, x_{r2}, f_r(x_r))\tilde{\theta} - k_p\hat{M}^{-1}e - k_v\hat{M}^{-1}z_2] \\ \le&2r_M\|e\|\|z_2\| + 2b_p\rho_M\|e\| + 2c\|z_2\|(b_{q1}\rho_M\|\theta\| + b_{q2}\|z_2\|\|\theta\| + b_{q3}\|\theta\| + b_Hb_{\tilde{\theta}} + k_p\|e\|/\bar{m}_m - k_v\|z_2\|/\bar{m}_M) \\ &+ r_M\|z_2\|^2 + b_p\rho_M\|z_2\| + \|e\|(b_{y1}\rho_M\|\theta\| + b_{y2}\|z_2\|\|\theta\| + b_{y3}\|\theta\| + b_Hb_{\tilde{\theta}} + k_v\|z_2\|/\bar{m}_m)/m_m - k_p\|e\|^2/(m_M\bar{m}_M)\end{aligned} \tag{67}$$

which is consistent with the one in Eq. (69) (under the condition $k_p = 1$ and $k_v = 5$).

Finally, it should be emphasized that any controllers that can ensure the UUB of e z_1, and z_2 is eligible to be employed as the initial control policy for the proposed method, as long as they can be reconstructed by proper basis functions. The one here can be regarded as a simple but effective example for the specific simulation scenario considered in this paper.

References

[1] P. Ioannou and J. Sun, *Robust Adaptive Control*, Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[2] D. Seo and M. R. Akella, “High-performance spacecraft adaptive attitude-tracking control through attracting-manifold design,” *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 4, pp. 884–891, 2008.

[3] ——, “Non-certainty equivalent adaptive control for robot manipulator systems,” *Systems & Control Letters*, vol. 58, no. 4, pp. 304–308, 2009.

[4] A. Astolfi and R. Ortega, “Immersion and invariance: A new tool for stabilization and adaptive control of nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 590–606, 2003.

[5] R. Ortega, L. Hsu, and A. Astolfi, “Immersion and invariance adaptive control of linear multivariable systems,” *Systems & Control Letters*, vol. 49, no. 1, pp. 37–47, 2003.

[6] M. Mattioni, S. Monaco, and D. Normand-Cyrot, “Immersion and invariance stabilization of strict-feedback dynamics under sampling,” *Automatica*, vol. 76, pp. 78–86, 2017.

[7] H. Dong, Q. Hu, M. R. Akella, and H. Yang, “Composite adaptive attitude-tracking control with parameter convergence under finite excitation,” *IEEE Transactions on Control Systems Technology*, 2019, Early Access.

[8] C. P. Bechlioulis and G. A. Rovithakis, “Robust adaptive control of feedback linearizable mimo nonlinear systems with prescribed performance,” *IEEE Transactions on Automatic Control*, vol. 53, no. 9, pp. 2090–2099, 2008.

[9] ——, “Adaptive control with guaranteed transient and steady state tracking error bounds for strict feedback systems,” *Automatica*, vol. 45, no. 2, pp. 532–538, 2009.

[10] ——, “Prescribed performance adaptive control for multi-input multi-output affine in the control nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 55, no. 5, pp. 1220–1226, 2010.

[11] A. K. Kostarigka and G. A. Rovithakis, “Adaptive dynamic output feedback neural network control of uncertain mimo nonlinear systems with prescribed performance,” *IEEE transactions on neural networks and learning systems*, vol. 23, no. 1, pp. 138–149, 2011.

[12] S. El-Ferik, H. A. Hashim, and F. L. Lewis, “Neuro-adaptive distributed control with prescribed performance for the synchronization of unknown nonlinear networked systems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2135–2144, 2017.

[13] C. Wei, J. Luo, H. Dai, and G. Duan, “Learning-based adaptive attitude control of spacecraft formation with guaranteed prescribed performance,” *IEEE transactions on cybernetics*, vol. 49, no. 11, pp. 4004–4016, 2019.

[14] Q. Hu, X. Shao, and L. Guo, “Adaptive fault-tolerant attitude tracking control of spacecraft with prescribed performance,” *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 331–341, 2017.

[15] M. Wang and A. Yang, “Dynamic learning from adaptive neural control of robot manipulators with prescribed performance,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2244–2255, 2017.

[16] L. Liu, Y.-J. Liu, and S. Tong, “Fuzzy-based multierror constraint control for switched nonlinear systems and its applications,” *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 8, pp. 1519–1531, 2018.

[17] L. Liu, Y.-J. Liu, A. Chen, S. Tong, and P. Chen, “Integral barrier lyapunov function based adaptive control for switched nonlinear systems,” *SCIENCE CHINA Information Sciences*, vol. 63, 2020, Paper No. 132203.

[18] P. J. Werbos, “Intelligence in the brain: A theory of how it works and how to build it,” *Neural Networks*, vol. 22, no. 3, pp. 200–212, 2009.

[19] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[20] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming with Applications in Optimal Control*, Springer, 2017.

[21] B. Luo, D. Liu, T. Huang, and D. Wang, “Model-free optimal tracking control via critic-only Q-learning,” *IEEE transactions on neural networks and learning systems*, vol. 27, no. 10, pp. 2134–2144, 2016.

[22] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, “Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers,” *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 76–105, 2012.

[23] H. Modares, F. L. Lewis, and M.-B. Naghibi-Sistani, “Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 24, no. 10, pp. 1513–1525, 2013.

[24] H. Zhang, L. Cui, X. Zhang, and Y. Luo, “Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method,” *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2226–2236, 2011.

[25] Y. Jiang and Z.-P. Jiang, “Robust adaptive dynamic programming and feedback stabilization of nonlinear systems,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 882–893, 2014.

[26] D. Liu and Q. Wei, “Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 621–634, 2014.

[27] D. Liu, D. Wang, and X. Yang, “An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs,” *Information Sciences*, vol. 220, pp. 331–342, 2013.

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

[29] Y. Jiang and Z.-P. Jiang, *Robust adaptive dynamic programming*. John Wiley & Sons, 2017.

[30] R. Kamalapurkar, P. Walters, and W. E. Dixon, “Model-based reinforcement learning for approximate optimal regulation,” *Automatica*, vol. 64, pp. 94–104, 2016.

[31] K. G. Vamvoudakis, M. F. Miranda, and J. P. Hespanha, “Asymptotically stable adaptive–optimal control algorithm with saturating actuators and relaxed persistence of excitation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2386–2398, 2016.

[32] R. Kamalapurkar, L. Andrews, P. Walters, and W. E. Dixon, “Model-based reinforcement learning for infinite-horizon approximate optimal tracking,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 753–758, 2017.

[33] D. Görges, “Relations between model predictive control and reinforcement learning,” in *2017 IFAC World Congress*. IFAC, 2017, pp. 4920–4928.

[34] J. Na, B. Wang, G. Li, S. Zhan, and W. He, “Nonlinear constrained optimal control of wave energy converters with adaptive dynamic

programming," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 7904–7915, 2019.

[35] W. He, Y. Dong, and C. Sun, "Adaptive neural impedance control of a robotic manipulator with input saturation," *IEEE Transactions on Systems Man Cybernetics: Systems*, vol. 46, no. 3, pp. 334–344, 2016.

[36] M. Jin, S. H. Kang, P. H. Chang, and J. Lee, "Robust control of robot manipulators using inclusive and enhanced time delay control," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 2141–2152, 2017.

[37] B. Xiao, Q. Hu, and Y. Zhang, "Finite-time attitude tracking of spacecraft with fault-tolerant capability," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 4, pp. 1338–1350, 2015.

[38] G. Chowdhary and E. Johnson, "Concurrent learning for convergence in adaptive control without persistency of excitation," in *50th IEEE Conference on Decision and Control*, IEEE, Atlanta, GA, USA, 2011.

[39] ——, "Theory and flight-test validation of a concurrent-learning adaptive controller," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 2, pp. 592–607, 2012.

[40] G. Chowdhary, M. Muhlegg, and E. Johnson, "Exponential parameter and tracking erorr convergence guarantees for adaptive controllers without persistency of excitation," *International Journal of Control*, vol. 87, no. 8, pp. 1583–1603, 2014.

[41] R. Kamalapurkar, B. Reish, G. Chowdhary, and W. E. Dixon, "Concurrent learning for parameter estimation using dynamic state-derivative estimators," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3594–3601, 2017.

[42] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[43] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network hjb approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[44] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[45] L. Liu, Y.-J. Liu, and S. Tong, "Neural networks-based adaptive finite-time fault-tolerant control for a class of strict-feedback switched nonlinear systems," *IEEE transactions on cybernetics*, vol. 49, no. 7, pp. 2536–2545, 2018.

[46] D. Wang, C. Mu, D. Liu, and H. Ma, "On mixed data and event driven design for adaptive-critic-based nonlinear h_∞ control," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 4, pp. 993–1005, 2017.

[47] S. Xue, B. Luo, and D. Liu, "Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, Early Access.

[48] R. Kamalapurkar, H. Dinh, S. Bhasin, and W. E. Dixon, "Approximate optimal trajectory tracking for continuous-time nonlinear systems," *Automatica*, vol. 51, pp. 40–48, 2015.

[49] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, 2013.

Hongyang Dong is currently a Research Fellow in machine learning and intelligent control at the School of Engineering, University of Warwick, Coventry, UK. He obtained his Ph.D. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2018. From 2015-2017, he was a joint Ph.D. student at the Cockrell School of Engineering, University of Texas at Austin, Texas, USA. His current research interests include reinforcement learning, deep learning, intelligent control, and adaptive control.

Xiaowei Zhao is Professor of Control Engineering and an EPSRC Fellow at the School of Engineering, University of Warwick, Coventry, UK. He obtained his PhD degree in Control Theory from Imperial College London in 2010. After that he worked as a postdoctoral researcher at the University of Oxford for three years before joining Warwick in 2013. His main research areas are control theory with applications on offshore renewable energy systems, local smart energy systems, and autonomous systems.

Biao Luo (M'15-SM'18) received the Ph.D. degree in control science and engineering from Beihang University, Beijing, China, in 2014. He is currently a Professor with the School of Automation, Central South University (CSU), Changsha, China. Before joining CSU, he was an Associate Professor and Assistant Professor with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2014 to 2018. His current research interests include distributed parameter systems, intelligent control, reinforcement learning, deep learning, and computational intelligence. Dr. Luo was a recipient of the Chinese Association of Automation Outstanding Ph.D Dissertation Award in 2015. He serves as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, the *Artificial Intelligence Review*, the *Neurocomputing*, and the *Journal of Industrial and Management Optimization*. He is a Senior Member of the IEEE, and the Secretariat of Adaptive Dynamic Programming and Reinforcement Learning Technical Committee, Chinese Association of Automation.