Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Indirect adaptive fuzzy-regulated optimal control for unknown continuous-time nonlinear systems*

Haiyun ZHANG[1,2], Deyuan MENG[2], Jin WANG‡[1], Guodong LU[1]

[1]*State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China*

[2]*Department of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China*

E-mail: gray_sun@zju.edu.cn; tinydreams@126.com; dwjcom@zju.edu.cn; lugd@zju.edu.cn

Received Nov. 11, 2019; Revision accepted Mar. 27, 2020; Crosschecked Sept. 28, 2020; Published online Jan. 8, 2021

Abstract: We present a novel indirect adaptive fuzzy-regulated optimal control scheme for continuous-time nonlinear systems with unknown dynamics, mismatches, and disturbances. Initially, the Hamilton-Jacobi-Bellman (HJB) equation associated with its performance function is derived for the original nonlinear systems. Unlike existing adaptive dynamic programming (ADP) approaches, this scheme uses a special non-quadratic variable performance function as the reinforcement medium in the actor-critic architecture. An adaptive fuzzy-regulated critic structure is correspondingly constructed to configure the weighting matrix of the performance function for the purpose of approximating and balancing the HJB equation. A concurrent self-organizing learning technique is designed to adaptively update the critic weights. Based on this particular critic, an adaptive optimal feedback controller is developed as the actor with a new form of augmented Riccati equation to optimize the fuzzy-regulated variable performance function in real time. The result is an online indirect adaptive optimal control mechanism implemented as an actor-critic structure, which involves continuous-time adaptation of both the optimal cost and the optimal control policy. The convergence and closed-loop stability of the proposed system are proved and guaranteed. Simulation examples and comparisons show the effectiveness and advantages of the proposed method.

Key words: Indirect adaptive optimal control; Hamilton-Jacobi-Bellman equation; Fuzzy-regulated critic; Adaptive optimal control actor; Actor-critic structure; Unknown nonlinear systems

https://doi.org/10.1631/FITEE.1900610 **CLC number:** TP13

1 Introduction

In the optimal control framework, Bellman's optimality principle and Pontryagin's minimum principle play the most important roles in finding solutions subject to various dynamics and constraints. Specifically, the optimality principle often leads to the Hamilton-Jacobi-Bellman (HJB) equation offering a necessary and sufficient condition for control optimization (Zak, 2003). However, while solving the optimal control problem for dynamic systems, e.g., with dynamic programming (DP), a main difficulty is encountered that either the HJB equation or the costate equation needs to be developed backward in time. Optimal feedback control by a backward-in-time solving process demands huge computation load and storage complexity, thus leading to the so-called curse of dimensionality (Powell, 2007).

An open-loop optimal control solution is ideal only for systems with a perfectly accurate model at the cost of losing some favorable properties, like system stabilization and robustness to uncertainties and disturbances. Therefore, a forward-in-time

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 51805531 and 51675470), the Natural Science Foundation of Jiangsu Province, China (No. BK20150200), the Key R&D Program of Zhejiang Province, China (No. 2020C01026), and the China Postdoctoral Science Foundation (No. 2020M671706)

ORCID: Haiyun ZHANG, https://orcid.org/0000-0002-2558-1564; Jin WANG, https://orcid.org/0000-0003-3106-021X

method to facilitate the HJB (or costate) equation in feedback control is crucial for real-world applications of optimal theory (Werbos, 2004; Lewis et al., 2012b). Inspired by biological learning and interacting mechanisms, adaptive (approximate) dynamic programming (ADP) methods have been proposed to solve the optimal problems of uncertain systems (Murray et al., 2002; Lee JM and Lee, 2004; Wang et al., 2009; Vamvoudakis and Lewis, 2010).

In general, ADP technology uses an actor-critic architecture of reinforcement learning (RL) to carry out sequential optimization. The actor performs continuous-time control actions, while the critic evaluates the performance and updates the actor's control policy (e.g., the controller's feedback gain) sequentially until the control performance will no longer be improved. Vrabie et al. (2009) proposed a Newtonian method based policy iteration technique for the adaptive optimal control of continuous-time linear systems. Jiang and Jiang (2012) developed a practical computational least square (LS) approach for the online solution of adaptive optimal controllers without knowing the system matrices. In Lee JY et al. (2012), an ε-approximate integral Q-learning and exploring policy iteration method was presented and investigated in terms of persistency of excitation (PE) and exploration. Bian and Jiang (2016) and Yang X et al. (2016) developed data-driven integral RL algorithms to solve input-constrained robust adaptive optimal control problems. Through the RL process, ADP approximately solves Bellman's equation and computes the optimal control policy for a given cost iteratively. Accordingly, it not only enables the forward-in-time computation, but also obtains the optimal feedback controller through a series of policy iterations without using the system dynamics or model (Lewis et al., 2012a; Kiumarsi et al., 2014).

Due to their remarkable abilities and advantages, ADP approaches have become powerful tools for studying control and optimization problems. However, for unknown nonlinear systems with uncertain dynamics, mismatches, and disturbances, where the optimal problem is governed by the nonlinear HJB equation, i.e., nonlinear Lyapunov function, the ADP solution has turned out to be much more computationally complex (Padhi et al., 2006; Wei et al., 2009; Liu and Wei, 2013). Neural networks (NNs) with properties of universal approximation have been introduced into ADP for uncertain nonlinear systems (Liu et al., 2013; Jiang and Jiang, 2014; Song et al., 2014; Lee JY et al., 2015).

Typically, the ADP for uncertain nonlinear systems includes two NNs; an actor NN and a critic NN are used to approximate the control policy and the cost function, respectively. The two-NN-based actor-critic architecture iterates between the steps of performance evaluation and policy improvement with the feedback of reinforcement information, while the optimal control policy approximates the solution of the HJB function. Yang XY et al. (2013) introduced a modified identifier-critic architecture using dual NNs in RL for online optimal control. Liu et al. (2014) proposed an NN-based online HJB solution approach for optimal robust guaranteed cost control by defining an appropriate bounded cost function. Yang X and He (2018) developed a critic NN-based HJB solution and self-learning robust optimal control for input-affine nonlinear systems with an auxiliary optimal control law and a novel concurrent network tuning rule.

Most existing ADP approaches are essentially direct adaptive optimal control methods as the executed control policy approximates and converges to the optimal solution to the HJB equation directly through a sequence of iterations, without an identification procedure or measurements of the state derivative (Lin 2011; Modares and Lewis, 2014; Vamvoudakis, 2017). However, some problems are perplexing and restraining the applicability of ADP, especially the NN-based ADP methods for unknown nonlinear systems: (1) initial admissible (stabilizing) policy which is difficult to fulfill in practice, especially for uncertain dynamic plants; (2) persistency of excitation that brings excessive noise, unpredictable errors and deviations, and performance influence and even degradation; (3) the minimum sampling time which restricts the rapidity and agility of the control response; (4) extra computational load: the use of two separate approximation NNs inevitably increases the expense and complexity of online computation.

In the literature, adaptive fuzzy control has likewise become a productive research field for unknown nonlinear systems because of the powerful modeling and approximating capacities of fuzzy logic systems (FLSs) (Li et al., 2016; Yin et al., 2017; Yu et al., 2018). Ma et al. (2019) proposed an adaptive fuzzy tracking control scheme for a class of uncertain

switched nonlinear systems with multiple constraints based on small-gain approaches. Chang Y et al. (2019) developed an adaptive fuzzy output-feedback control method for switched stochastic nonlinear systems by combining the back-stepping recursive technique and the Lyapunov function approach. Chang XH et al. (2019) designed a synchronous fuzzy output feedback H_∞ controller and quantizer for a continuous nonlinear system using the Takagi-Sugeno fuzzy model representation with both measurement output and control input quantization. Huo et al. (2020) presented an event-triggered adaptive fuzzy output feedback control scheme based on a multi-input and multi-output (MIMO) switched fuzzy observer for nonlinear state estimation and uncertainty approximation.

Motivated by the above discussion, we present a novel indirect adaptive fuzzy-regulated optimal control (IAFOC) scheme for the case of unknown continuous-time nonlinear systems. Within the actor-critic framework, our study formulates and converts the intractable solution of the nonlinear HJB equation into an indirect adaptive optimal balancing control problem. The major contributions of this paper are as follows:

1. Implemented in an adaptive actor-critic architecture, our methodology is extended to an alternative pattern of using the HJB equation aiming to approximate the optimal balancing control solution subject to optimal control cost and system dynamics in a Lyapunov sense.

2. Employing a variable performance function in the control cost as a reinforcement medium, the adaptive fuzzy-regulated critic is endowed with the capacity of simultaneously altering and matching the structures of both the optimal cost and the optimal feedback controller.

3. Different from previous work, this method effectively yields an indirect adaptive optimal balancing control scheme, which finds the adaptive approximation and balance between the optimal cost and optimal control policy in real time, while guaranteeing closed-loop stability. The restrictive conditions including the initial admissible control and the persistence of excitation condition are relaxed using the proposed scheme. This study provides a promising methodology for designing a non-model based robust optimal controller for general continuous-time nonlinear systems with unknown dynamics, mismatches, and disturbances.

2 Nonlinear optimal control formulation

Consider the continuous-time nonlinear systems subject to unknown dynamics described as

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

where $\boldsymbol{x}(t)\in\mathbb{R}^n$, $\boldsymbol{f}(\boldsymbol{x})\in\mathbb{R}^n$, $\boldsymbol{g}(\boldsymbol{x})\in\mathbb{R}^{n\times m}$, and $\boldsymbol{u}(t)\in U\subset\mathbb{R}^m$ represent the system states available for measurement, drift dynamics, input dynamics, and control input, respectively.

Assumption 1 (Vamvoudakis and Lewis, 2010; Bhasin et al., 2013) The unknown dynamic functions $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$ are locally Lipschitz on a compact set $\Omega\subseteq\mathbb{R}^n$ containing the origin $\boldsymbol{f}(\boldsymbol{0})=\boldsymbol{0}$; i.e., there exists a control input $\boldsymbol{u}(t)\in U$ such that system (1) is asymptotically stabilizable on Ω. In addition, there exists an unknown positive constant g_m bounding the input dynamics function such that $0\le\|\boldsymbol{g}(\boldsymbol{x})\|\le g_\mathrm{m}$ for $\boldsymbol{x}\in\mathbb{R}^n$.

Remark 1 Assumption 1 is a general standard assumption that makes sure the solution $\boldsymbol{x}(t)$ of system (1) is unique for any finite initial condition $\boldsymbol{x}_0$. Although the bound assumption on $\boldsymbol{g}(x)$ restricts the class of input dynamics, a wide range of physical or practical nonlinear systems fulfills this property, like robotic systems and aircraft systems (Slotine and Li, 1991; Sastry, 1999).

Define an infinite horizon integral control cost as

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

where $r(\boldsymbol{x},\boldsymbol{u})=\Gamma(\boldsymbol{x})+\boldsymbol{u}^\mathrm{T}\boldsymbol{R}\boldsymbol{u}$ (Here, the performance function $\Gamma(\boldsymbol{x})$ is positive definite; i.e., $\forall\boldsymbol{x}\neq\boldsymbol{0}$, $\Gamma(\boldsymbol{x})>0$, and $\Gamma(\boldsymbol{0})=\boldsymbol{0}$), and $\boldsymbol{R}\in\mathbb{R}^{n\times m}$ is also symmetric positive definite. Reformulate the infinitesimal version of cost (2) as

$$0=r(\boldsymbol{x},\boldsymbol{u}(\boldsymbol{x}))+(\nabla\boldsymbol{V}_{\boldsymbol{x}}^u)^\mathrm{T}[\boldsymbol{f}(\boldsymbol{x})+\boldsymbol{g}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})],\ \boldsymbol{V}^u(\boldsymbol{0})=\boldsymbol{0}, \tag{3}$$

where the column vector $\nabla\boldsymbol{V}_{\boldsymbol{x}}$ denotes the partial derivative (i.e., gradient) of cost (2) with respect to $\boldsymbol{x}$,

and the alternative operator notation $\nabla \equiv \partial/\partial \boldsymbol{x}$. Eq. (3) is also known as a nonlinear Lyapunov equation. Given that $\boldsymbol{u}(\boldsymbol{x}) \in \mu(\Omega)$ is an admissible control policy, if $\boldsymbol{V}^{\mu}(\boldsymbol{x})$ is subject to Eq. (3), while $r(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x})) \geq 0$, then $\boldsymbol{V}^{\mu}(\boldsymbol{x})$ is a Lyapunov function candidate for nonlinear system (1). Define the Hamiltonian function for the nonlinear control problem as

$$\begin{aligned} &H(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{V}_x) \\ &= r(\boldsymbol{x}(t), \boldsymbol{u}(t)) + (\nabla \boldsymbol{V}_x)^{\mathrm{T}}[\boldsymbol{f}(\boldsymbol{x}(t)) + \boldsymbol{g}(\boldsymbol{x}(t))\boldsymbol{u}(t)]. \end{aligned} \tag{4}$$

According to Bellman's optimality principle, the optimal control cost function $\boldsymbol{V}^{*}(\boldsymbol{x})$ satisfies the Hamilton-Jacobi-Bellman (HJB) equation:

$$0 = \min_{u}[H(\boldsymbol{x}, \boldsymbol{u}, \nabla \boldsymbol{V}_x^{*})]. \tag{5}$$

By solving the HJB equation for the cost function $\boldsymbol{V}^{*}(\boldsymbol{x})$ associated with the admissible optimal control policy $\boldsymbol{u}^{*}(\boldsymbol{x})$, the infinite horizon cost (2) for nonlinear system (1) is minimized. According to Assumption 1, the solution to the minimum of HJB equation (5) exists and is unique, and the optimal controller for the given problem is

$$\boldsymbol{u}^{*}(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{R}^{-1}\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x})\nabla \boldsymbol{V}_x^{*}. \tag{6}$$

Applying the optimal control policy (6) to the nonlinear Lyapunov equation (3) yields

$$0 = \Gamma(\boldsymbol{x}) + (\nabla \boldsymbol{V}_x^{*})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}) - \frac{1}{4}(\nabla \boldsymbol{V}_x^{*})^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})\boldsymbol{R}^{-1}\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x})\nabla \boldsymbol{V}_x^{*}. \tag{7}$$

3 Indirect adaptive fuzzy-regulated optimal control design

3.1 Actor-critic architecture

Generally, to obtain the optimal control $\boldsymbol{u}^{*}(\boldsymbol{x})$, one needs to solve the HJB equation (7) first for its associated cost function $\nabla \boldsymbol{V}_x^{*}$ and then apply the result to Eq. (6) for the optimal control input. However, the solution requires complete knowledge or a perfect model of the system dynamics. This is hard to acquire in practice. Thus, the computation of the HJB equation is usually difficult, especially for uncertain nonlinear systems.

To solve this dilemma, the actor-critic architecture of the ADP is proposed for the nonlinear HJB solution problem. As illustrated in Fig. 1, this architecture is composed of two steps, i.e., critic's policy evaluation and actor's control improvement. With a predefined cost function as the reinforcement signal, the involved HJB equation is solved through a sequence of policy iterations until the executed control policy approximates the optimal solution.

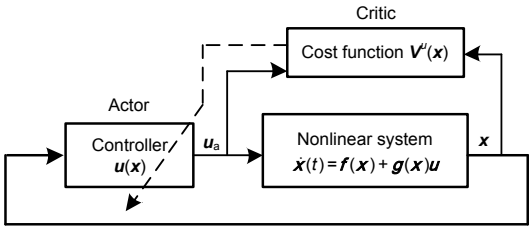


Fig. 1 Actor-critic structure of adaptive optimal control

As discussed earlier, existing ADP methods have encountered certain problems and restrictions. Therefore, within the actor-critic framework, the interest of this study is to find a new adaptive optimal control method to manipulate and use the HJB function arising in the uncertain nonlinear circumstance. As shown in Eq. (7), since the structure of optimal control policy $\boldsymbol{u}^{*}(\boldsymbol{x})$ is determined by Eq. (6), the formulation of the HJB equation is transferred in terms of performance function $\Gamma(\boldsymbol{x})$ and the gradient of control cost $\nabla \boldsymbol{V}_x^{*}$. Consider $\Gamma(\boldsymbol{x})$ and $\nabla \boldsymbol{V}_x^{*}$ as two special manipulated variables. This study transfers a variable performance function $\Gamma(\boldsymbol{x})$ as the reinforcement medium in the actor-critic architecture, so that it will be adapted and manipulated to match the Lyapunov function $\nabla \boldsymbol{V}_x^{*}$ (i.e., optimal control cost) subject to the system dynamics, for the purpose of approximating and balancing the underlying HJB equation (7).

Define a value function associated with the HJB equation as

$$\begin{aligned} \boldsymbol{\psi}(\boldsymbol{x}) = &\Gamma(\boldsymbol{x}) + (\nabla \boldsymbol{V}_x^{*})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}) \\ &- \frac{1}{4}(\nabla \boldsymbol{V}_x^{*})^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})\boldsymbol{R}^{-1}\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x})\nabla \boldsymbol{V}_x^{*}. \end{aligned} \tag{8}$$

Therefore, through the adaption and manipulation of reinforcement medium $\Gamma(\boldsymbol{x})$, our goal is to

derive the adaptive approximation and balance of HJB equation (7) in real time, which can be transferred and expressed as

$$0=\min_{\Gamma(\boldsymbol{x})}[\boldsymbol{\psi}(\boldsymbol{x},\Gamma(\boldsymbol{x}),\nabla\boldsymbol{V}_{\boldsymbol{x}}^{*}(\boldsymbol{x}))]. \tag{9}$$

The performance function $\Gamma(\boldsymbol{x})$ is also referred to as the performance index in cost function (2). Its configuration represents the optimal control strategy mathematically and determines the weighting level on each dynamic variable to be minimized by the optimal control procedure. Instead of using a prescribed form, this study defines the positive definite $\Gamma(\boldsymbol{x})$ ($\forall \boldsymbol{x}\neq\boldsymbol{0}$) as

$$\Gamma(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x})\boldsymbol{x}, \tag{10}$$

where $\boldsymbol{Q}(\boldsymbol{x})=\mathrm{diag}(\boldsymbol{q}(\boldsymbol{x}))$ (Here, $\boldsymbol{q}(\boldsymbol{x})=[q_1(\boldsymbol{x}), q_2(\boldsymbol{x}), \ldots, q_n(\boldsymbol{x})]^{\mathrm{T}}\in\mathbb{R}^n$ denotes a positive vector with weighting elements $\{q_i(\boldsymbol{x})\}_{i=1:n}$ corresponding to system state $\boldsymbol{x}=[x_1, x_2, \ldots, x_n]^{\mathrm{T}}\in\mathbb{R}^n$), and $\boldsymbol{Q}(\boldsymbol{x})\in\mathbb{R}^{n\times n}$ is a symmetric positive matrix function stacking the diagonal elements.

To ensure a desirable feedback structure for the control actor, let the Lyapunov function $\boldsymbol{V}(\boldsymbol{x}(t))$ be

$$\boldsymbol{V}(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}(t)\boldsymbol{x},\ \nabla\boldsymbol{V}(\boldsymbol{x})=2\boldsymbol{P}\boldsymbol{x},\ \boldsymbol{P}=\boldsymbol{P}^{\mathrm{T}}, \tag{11}$$

where $\boldsymbol{P}(t)\in\mathbb{R}^{n\times n}$ is a symmetric positive definite matrix to be calculated. Accordingly, the Lyapunov function (i.e., the optimal control cost) is given by the classic quadratic form, and the feedback structure can be constructed for an optimal controller, as has been performed in Abu-Khalaf and Lewis (2005) and Modares et al. (2013).

3.2 Adaptive fuzzy-regulated critic

Because of their properties of universal approximation, nonlinearity, adaptability, self-learning, and fault tolerance, FLSs are naturally employed as powerful tools to approximate nonlinear functions. Compared to NNs, FLS has outstanding advantages in simplifying the control structure and reducing computational load. Considering the unknown and unpredictable nonlinear characteristics of the controlled system, we introduce a new self-learning fuzzy logic critic structure for the regulation and adaptation of the performance function $\Gamma(\boldsymbol{x})$ to approximately balance and minimize the HJB value function $\boldsymbol{\psi}(\boldsymbol{x})$.

Lemma 1 (Higher-order Weierstrass approximation theorem) Let function $\boldsymbol{h}(\boldsymbol{x})\in C^{m}(\Omega)$. Then there exists a polynomial $\boldsymbol{L}(\boldsymbol{x})$ such that it converges uniformly to $\boldsymbol{h}(\boldsymbol{x})$ and all its partial derivatives up to order m converge uniformly (Finlayson, 1990; Abu-Khalaf and Lewis, 2005; Vamvoudakis and Lewis, 2010; Modares and Lewis, 2014).

Fact 1 According to the optimality conditions, the Lyapunov function $\boldsymbol{V}_{\boldsymbol{x}}^{*}(\boldsymbol{x})$ in Eq. (9) is positive definite, continuous, and smooth, i.e., $\boldsymbol{V}_{\boldsymbol{x}}^{*}(\boldsymbol{x})\in C^{1}(\Omega)$. Assumption 1 guarantees the stabilizable system dynamics (1), and the performance function $\Gamma(\boldsymbol{x})$ is also positive definite and satisfies zero state observability (van der Schaft, 1992) (i.e., $\Gamma(\boldsymbol{x})>0$ when $\boldsymbol{x}\in\Omega\backslash\{0\}$ and $\Gamma(\boldsymbol{0})=\boldsymbol{0}$).

According to Fact 1, the sufficient conditions of the higher-order Weierstrass approximation theorem are fulfilled. There exists a completely independent basis set $\{q_i^{l}(\boldsymbol{x})\}_{l=1}^{N}$, such that the composite performance function $\Gamma(q_i^{l}(\boldsymbol{x}))$ in Eq. (10) can uniformly approximate the solution of Eq. (9), which balances the Lyapunov function $\boldsymbol{V}_{\boldsymbol{x}}^{*}(\boldsymbol{x})\in C^{1}(\Omega)$ and minimizes the value function of $\boldsymbol{\psi}(\boldsymbol{x})\in C^{1,2}(\Omega)$. Within the critic structure, we have

$$\forall\varepsilon>0,\ \exists N,\ \Gamma(q_i^{N}(\boldsymbol{x})),\ \boldsymbol{V}_{N}^{*}:\ \|\boldsymbol{\psi}_{N}(\boldsymbol{x})\|<\varepsilon, \tag{12}$$

where $N\to\infty$, $\Gamma(q_i^{N}(\boldsymbol{x}))\to\Gamma^{*}(\boldsymbol{q}(\boldsymbol{x}))$, $\boldsymbol{V}_{N}^{*}\to\boldsymbol{V}_{\boldsymbol{x}}^{*}$, and $\boldsymbol{\psi}_{N}(\boldsymbol{x})\to\boldsymbol{\psi}^{*}(\boldsymbol{x})\Rightarrow\sup_{\boldsymbol{x}\in\Omega}\|\boldsymbol{\psi}(\boldsymbol{x})\|_{L_2(\Omega)}\to 0$.

Considering the requirements of linear independence and completeness, the weighting elements $q_i(\boldsymbol{x})$ need to satisfy

$$\frac{\partial q_i(\boldsymbol{x})}{\partial x_j}=0,\ \forall i\neq j\ (i,j=1,2,\ldots,n). \tag{13}$$

Namely, the i^{th} weighting element $q_i(\boldsymbol{x})$ is explicitly irrelevant to other system states x_j ($i\neq j$) since the exclusive correlation $q_i(\boldsymbol{x})=q_i(x_i)$ ensures their explicit linear independence.

Based on the above results and conditions, an adaptive fuzzy-regulated critic is designed for performance function $\Gamma(\boldsymbol{x})$, i.e., the weighting elements $\{q_i(\boldsymbol{x})\}_{i=1:n}$. The fuzzy-regulated critic structure

employs the Takagi-Sugeno (T-S) fuzzy inference model, equal-span triangular fuzzifier, max-product inference engine, and height defuzzifier, while the fuzzy rule base subject to Eq. (13) is characterized by a set of if-then rules in the following form:

$$R_i^l:\ \text{if } x_i \text{ is } G_1^{l_i} \text{ and } \Delta x_i \text{ is } G_2^{l_i}, \text{then } y_i \text{ is } q_i^l(\boldsymbol{x}), \quad (14)$$

where l=1, 2, …, L_i, x_i and $\Delta x_i \in \mathbb{R}$ are the inputs of the fuzzy logic structure, y_i is the output of the fuzzy logic structure, and $G_k^{l_i}$ and $q_i^l(\boldsymbol{x})$ are the input fuzzy set and singleton output set, respectively. According to the fuzzy logic design, the output for weighting elements $q_i(\boldsymbol{x})$ can be expressed as

$$y_i(\boldsymbol{x}) = \frac{\sum_{l=1}^{L_i} q_i^l(\boldsymbol{x}) \left(\prod_{k=1}^{2} \mu_{G_k^l}(x_i) \right)}{\sum_{l=1}^{L_i} \prod_{k=1}^{2} \mu_{G_k^l}(x_i)}, \quad (15)$$

where $\mu_{G_k^l}(x_i)$ $(k=1,2)$ is the membership function value of the input variables (x_i and Δx_i). Take l=1, 2, …, L_i, where L_i (i=1, 2, …, n) are the numbers of rules characterized by the distribution of the fuzzy set for the i^{th} weighting element $q_i(\boldsymbol{x})$. With respect to the number of fuzzy rules, $N = \sum_{i=1}^{n} L_i$ is the size of the compact set $\{q_i^l(\boldsymbol{x})\}_{l=1:n}$. Define the fuzzy suitability values and the suitability function vector as

$$\begin{cases} \xi_i^{\mu(l)} = \dfrac{\prod_{k=1}^{2} \mu_{G_k^l}(x_i)}{\sum_{l=1}^{L_i} \prod_{k=1}^{2} \mu_{G_k^l}(x_i)}, \\ \boldsymbol{\varphi}_i = [\xi_i^{\mu(1)}, \xi_i^{\mu(2)}, \ldots, \xi_i^{\mu(L_i)}]^{\mathrm{T}} \in \mathbb{R}^{L_i}. \end{cases} \quad (16)$$

Considering the positive definite property requirement, the weighting element $q_i(\boldsymbol{x})$ is then to be transferred as

$$q_i(\boldsymbol{x}) = | y_i(\boldsymbol{x}) | = | \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{\theta}_i(\boldsymbol{x}) |, \quad (17)$$

where $\boldsymbol{\theta}_i(\boldsymbol{x}) = [q_i^1, q_i^2, \ldots, q_i^{L_i}]^{\mathrm{T}} \in \mathbb{R}^{L_i}$ is a part of the complete compact set of function $\boldsymbol{\phi}_N(\boldsymbol{x}) = [\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_n^{\mathrm{T}}]^{\mathrm{T}} : \mathbb{R}^n \to \mathbb{R}^N$. Rearrange the fuzzy coefficients (i.e., suitability function vectors which are determined in real time by the fuzzy critic structure) as

$$\boldsymbol{S}_\mu^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{\varphi}_1^{\mathrm{T}} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\varphi}_2^{\mathrm{T}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\varphi}_n^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{n \times N}. \quad (18)$$

Then the optimal weighting vector $\boldsymbol{q}(\boldsymbol{x})$ is formulated and obtained as

$$\boldsymbol{q}(\boldsymbol{x}) = | \boldsymbol{S}_\mu^{\mathrm{T}} \boldsymbol{\phi}_N(\boldsymbol{x}) |. \quad (19)$$

To autonomously construct and update the set of critic weights $\{q_i^l(\boldsymbol{x})\}$, a concurrent self-organizing learning algorithm is developed in the fuzzy-regulated critic. With reference to the Levenberg-Marquardt approach (Tao, 2003; Ioannou and Fidan, 2006), the self-organizing learning algorithm is designed as

$$\begin{cases} \Delta q_i^l(\boldsymbol{x}) = \kappa \sigma_i^l \dfrac{\Delta x_i}{x_i^2 + 1}, \\ q_i^l(\boldsymbol{x}_0) = 0, \end{cases} \quad (20)$$

where $\Delta q_i^l(\boldsymbol{x}) = q_i^l(k) - q_i^l(k-1)$ is the discrete correction value for the l^{th} fuzzy weight in $\{q_i^l(\boldsymbol{x})\}$ corresponding to the state variation $\Delta x_i = x_i(k) - x_i(k-1)$ on the time interval $[(k-1)T_{\mathrm{c}}, kT_{\mathrm{c}}]$ characterized by a discrete time sampler T_{c}. $\sigma_i^l = \prod_{k=1}^{2} \mu_{G_k^l}(x_i)$ is the excitation strength of the l^{th} rule. $\kappa > 0$ is the learning rate, and $x_i^2 + 1$ is used for normalization to ensure the bounded gradient of $x_i / (x_i^2 + 1)$. As initial conditions, $q_i^l(\boldsymbol{x}(0)) = 0$ for l=1, 2, …, N.

Using the proposed learning algorithm, the critic monitors and captures the unknown and complicated nonlinear dynamics in real time, adaptively adjusts and updates the set of fuzzy weights $\{q_i^l(\boldsymbol{x})\}$, and improves the precision of the fuzzy-regulated approximation of $\Gamma(\boldsymbol{x})$.

Therefore, given an infinite set $\{q_i^l(\boldsymbol{x})\}_{l=1}^{\infty}$ of linear independent activation functions $\boldsymbol{\phi}_{N \to \infty}(\boldsymbol{x})$ which ensures the completeness property of Lemma 1, then the solution of Eq. (9), i.e., the ideal performance function $\Gamma_\infty^*(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \mathrm{diag}(| (\boldsymbol{S}_{\mu(N)}^{\infty})^{\mathrm{T}} \boldsymbol{\phi}_\infty(\boldsymbol{x}) |) \boldsymbol{x}$, can be represented and approximated by the combination of

a subset of $\boldsymbol{\phi}_N(\boldsymbol{x}) \in C^{1,2}(\Omega)$ $(N=\overline{1,\infty})$. The ideal result indicates the situation where the underlying HJB equation (7) is mathematically balanced using the fuzzy approximation $\min[\psi^*(\boldsymbol{x}, \Gamma^*(\boldsymbol{x}), \nabla \boldsymbol{V}_{\boldsymbol{x}}^*(\boldsymbol{x}))]=0$, and control optimality is substantially achieved without solving the problem. In real circumstances, the fuzzy-regulated critic may not provide perfect representations for the nonlinear control and performance function $\Gamma^*(\boldsymbol{x})=\Gamma_N(\boldsymbol{x})+\delta$ $\left(\| \delta \| \leq \bar{\delta}\right)$.

3.3 Adaptive optimal control actor

According to the development of the fuzzy-regulated critic, the adaptation of performance function $\Gamma(\boldsymbol{x})$ will sequentially adjust and regulate the optimal control policy $\boldsymbol{u}^*(\boldsymbol{x})$. As shown in Eq. (6), the structure of the optimal control policy is given and characterized by the gradient of the Lyapunov function $\nabla \boldsymbol{V}_{\boldsymbol{x}}^*$. To establish the functional relationship of the optimal control policy and the gradient of the Lyapunov function, quick exploration of the system dynamics is necessary, since the controlled nonlinear dynamics of system (1) is completely unknown in the first place.

Assume that the nonlinear dynamics (1) can be rewritten as the system drift dynamics $\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{A}_f(t)\boldsymbol{x}+\Delta\boldsymbol{f}(\boldsymbol{x})$ and input-to-state dynamics $\boldsymbol{g}(\boldsymbol{x})=\boldsymbol{B}_g(t)+\Delta\boldsymbol{g}(\boldsymbol{x})$, where $\boldsymbol{A}_f \in \mathbb{R}^{n\times n}$ and $\boldsymbol{B}_g \in \mathbb{R}^{n\times m}$ are the dynamic matrices and $\Delta\boldsymbol{f}(\boldsymbol{x})$ and $\Delta\boldsymbol{g}(\boldsymbol{x})$ are the residual nonlinear estimation deviations. Different from a system identification technique, only the drift dynamics trend and input-to-state dynamics are going to be estimated in the quick dynamic exploration, while the knowledge regarding them is relatively easy to obtain (Vrabie et al., 2009; Jiang and Jiang, 2012; Lee JY et al., 2015).

During the exploration stage, we temporarily introduce an excitation input $\boldsymbol{u}_{\exp}=\boldsymbol{w}_t$, where $\boldsymbol{w}_t$ is any given non-zero persistent signal (or noise) that is exactly measurable and bounded by $w_{\mathrm{M}}>0$ (i.e., $\|\boldsymbol{w}_t\|<w_{\mathrm{M}}$). For system (1), we have

$$\begin{aligned}\dot{\boldsymbol{x}}(t) &= \boldsymbol{f}(\boldsymbol{x}(t))+\boldsymbol{g}(\boldsymbol{x}(t))\boldsymbol{u}(t)\\ &= [\boldsymbol{A}_f \quad \boldsymbol{B}_g]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{w}_t\end{bmatrix}+(\Delta\boldsymbol{f}+\Delta\boldsymbol{g}),\end{aligned} \tag{21}$$

and define $\boldsymbol{H}_{\mathrm{c}}=[\boldsymbol{A}_f\ \boldsymbol{B}_g]^{\mathrm{T}}=[\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n]\in\mathbb{R}^{(n+m)\times n}$ and $\boldsymbol{z}(t)=[\boldsymbol{x}^{\mathrm{T}} \quad \boldsymbol{w}_{\mathrm{t}}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n+m}$. For the unknown parameter vectors $\boldsymbol{h}_i \in \mathbb{R}^{n+m}$ with i=1, 2, …, n, in this study, we employ the classic LS estimation methodology. Record and denote the data matrices as

$$\begin{cases}\boldsymbol{I}_{\Delta\boldsymbol{x}}=[\bar{x}_\Delta^1, \bar{x}_\Delta^2, \ldots, \bar{x}_\Delta^N]^{\mathrm{T}},\ \bar{x}_\Delta^i=\bar{x}(t_i)-\bar{x}(t_{i-1}),\ t_i=t_{i-1}+T_{\mathrm{e}},\\ \boldsymbol{I}_{\boldsymbol{xw}}=\left[\int_{t_0}^{t_1}\boldsymbol{z}(\tau)\mathrm{d}\tau, \int_{t_1}^{t_2}\boldsymbol{z}(\tau)\mathrm{d}\tau, \ldots, \int_{t_{N-1}}^{t_N}\boldsymbol{z}(\tau)\mathrm{d}\tau\right]^{\mathrm{T}},\end{cases} \tag{22}$$

where T_{e} is the exploration (or estimation) sampling time interval. Considering the number $(n+m)\times n$ of pending parameters in dynamic matrix $\boldsymbol{H}_{\mathrm{c}}$, evaluate and record $N_{\mathrm{e}}\geq(n+m)\times n$ sequent points $\bar{x}_\Delta^i$ in the state space. Then, the following matrix equation is yielded:

$$\begin{cases}\boldsymbol{I}_{\Delta\boldsymbol{x}}^{\mathrm{T}}=\boldsymbol{H}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{I}_{\boldsymbol{xw}}^{\mathrm{T}}=[\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n]^{\mathrm{T}}\boldsymbol{I}_{\boldsymbol{xw}}^{\mathrm{T}},\\ \boldsymbol{I}_{\Delta\boldsymbol{x}}^i=\boldsymbol{I}_{\boldsymbol{xw}}\boldsymbol{h}_i,\end{cases} \tag{23}$$

where $\boldsymbol{I}_{\Delta\boldsymbol{x}}^i \in \mathbb{R}^N$ is the i^{th} column of the corresponding matrix. Solve and estimate LS through Eq. (23) approximately as

$$\begin{cases}\hat{\boldsymbol{h}}_i=(\boldsymbol{I}_{\boldsymbol{xw}}\boldsymbol{I}_{\boldsymbol{xw}}^{\mathrm{T}})^{-1}\boldsymbol{I}_{\boldsymbol{xw}}\boldsymbol{I}_{\Delta\boldsymbol{x}}^i,\\ \hat{\boldsymbol{H}}_{\mathrm{c}}=[\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \ldots, \hat{\boldsymbol{h}}_n]=[\hat{\boldsymbol{A}}_f \quad \hat{\boldsymbol{B}}_g]^{\mathrm{T}}.\end{cases} \tag{24}$$

Then, the preliminary estimation results (24) are further truncated for the drift dynamics trend:

$$\begin{cases}\hat{\boldsymbol{A}}_f :\to \bar{\boldsymbol{A}}=\{\bar{a}_{ij}\}_{i=1:n;j=1:n}=\begin{cases}\mathrm{sign}(\hat{a}_{ij}), & \text{if } |\hat{a}_{ij}|\geq\varepsilon_{\mathrm{e}},\\ 0, & \text{otherwise,}\end{cases}\\ \hat{\boldsymbol{B}}_g :\to \bar{\boldsymbol{B}}=\hat{\boldsymbol{B}}_g,\end{cases} \tag{25}$$

where $\varepsilon_{\mathrm{e}}>0$ is a pspecified exploration threshold, e.g., $\varepsilon_{\mathrm{e}}=0.001$. Finally, the truncated drift dynamics trend $\bar{\boldsymbol{A}}$ and input dynamics $\bar{\boldsymbol{B}}$ of system (1) are obtained for the optimal control policy design. Note that the input-to-state dynamics $\bar{\boldsymbol{B}}$ is important for the calculation of the optimal control input, and there are other accurate estimation or identification methods for the input-to-state dynamics, like the adaptive observer methods and computational adaptive control methods (Jiang and Jiang, 2012; Lee JY et al., 2015).

After the quick dynamic exploration of

initialization, the variations of dynamics need to be further monitored. The new dynamics of $\bar{\boldsymbol{A}}'$ and $\bar{\boldsymbol{B}}'$ are sequentially estimated and evaluated from Eq. (25) over the time period $[t, t+N_eT_e]$ by an analog integration processing unit. To adapt to the unpredictable dynamics, if the new estimation results last consecutive k periods (e.g., k=3 in this study), then it indicates a remarkable system dynamic variation. Thus, the dynamic exploration results are replaced by the new estimates $\bar{\boldsymbol{A}}'$ and $\bar{\boldsymbol{B}}'$ during the monitoring process.

Applying the dynamic exploration results (25) and considering the fuzzy-regulated approximation deviation of $\Gamma(\boldsymbol{x})$, the HJB value function can be rephrased as

$$\begin{aligned}&\min\psi^*(\boldsymbol{x})\\&=\Gamma(\boldsymbol{x})+(\nabla\boldsymbol{V}_x^*)^{\mathrm{T}}[\bar{\boldsymbol{A}}\boldsymbol{x}-\frac{1}{4}\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\nabla\boldsymbol{V}_x^*]+\tilde{\boldsymbol{r}}(\boldsymbol{x})\\&=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}_N(\boldsymbol{x})\boldsymbol{x}+(\nabla\boldsymbol{V}_x^*)^{\mathrm{T}}(\bar{\boldsymbol{A}}\boldsymbol{x}-\frac{1}{4}\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\nabla\boldsymbol{V}_x^*)+\boldsymbol{\varepsilon}_\psi\\&=0,\end{aligned}\tag{26}$$

where the approximation balance error is L_2-norm bounded by an unknown positive (i.e., $\|\boldsymbol{\varepsilon}_\psi\|\leq\bar{\varepsilon}$), as the exploration deviation term $\tilde{\boldsymbol{r}}(\boldsymbol{x})=\Delta\boldsymbol{f}+\Delta\boldsymbol{g}\boldsymbol{u}^*(\boldsymbol{x})\in\mathbb{R}^n$ is likewise within the scope of nonlinear approximation using the fuzzy-regulated critic.

According to Bellman's principle, the performance function (index) $\Gamma(\boldsymbol{x})$ stands for the optimal strategy, and its adaptation correspondingly adjusts and alters the structure and performance of the optimal control policy $\boldsymbol{u}^*(\boldsymbol{x})$ in Eq. (6):

$$\boldsymbol{u}^*(\boldsymbol{x})=-\frac{1}{2}\boldsymbol{R}^{-1}\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x})\nabla\boldsymbol{V}_x^*=-\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}.\tag{27}$$

As shown in Eq. (27), considering the given Lyapunov candidate $\boldsymbol{V}^*(\boldsymbol{x})$ in Eq. (11) characterized by the performance function $\Gamma(\boldsymbol{x})$ in the nonlinear HJB equation (7), the feasible optimal feedback controller is mathematically constructed and determined for the adaptive control actor (Lin, 2011; Modares and Lewis, 2014; Vamvoudakis, 2017). To ensure control optimality, re-taking the partial derivative of the Hamiltonian function (4) under the design and conditions of Eqs. (6), (17), (20), and (26) will yield the following equation for Lyaponov function (2):

$$\begin{aligned}\boldsymbol{0}=&\boldsymbol{P}(t)\bar{\boldsymbol{A}}+\bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{P}(t)-\boldsymbol{P}(t)\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{P}(t)\\&+\boldsymbol{Q}(\boldsymbol{x})+\frac{1}{2}\mathrm{diag}(\boldsymbol{x})\cdot\frac{\partial\boldsymbol{q}(\boldsymbol{x})}{\partial\boldsymbol{x}},\end{aligned}\tag{28}$$

where the diagonal matrix term subject to the object system dynamics can be rewritten using Eq. (13) as

$$\mathrm{diag}(\boldsymbol{x})\cdot\frac{\partial\boldsymbol{q}(\boldsymbol{x})}{\partial\boldsymbol{x}}=\mathrm{diag}\left(\left[x_1\frac{\partial q_1}{\partial x_1},x_2\frac{\partial q_2}{\partial x_2},\ldots,x_n\frac{\partial q_n}{\partial x_n}\right]^{\mathrm{T}}\right).$$

It is noteworthy that an augmented Riccati equation has been newly formulated for the policy of optimal control actor under an adaptive fuzzy-regulated critic. The weighting matrix $\boldsymbol{Q}(\boldsymbol{x})$ along with the diagonal augment term (28) adaptively determines the feedback structure of the optimal control policy through the calculation of $\boldsymbol{P}(t)$ in the Lyapunov function. According to the weighting element configuration $q_i(\boldsymbol{x})$ in Eqs. (19) and (20), for the diagonal special augment matrix, we have

$$\begin{aligned}x_i\frac{\partial q_i}{\partial x_i}&=x_i\boldsymbol{\varphi}_i^{\mathrm{T}}\frac{\partial|\boldsymbol{\theta}_i(\boldsymbol{x})|}{\partial x_i}\\&=\kappa\frac{x_i\,\mathrm{sign}(x_i)}{x_i^2+1}\sum_{l=1}^{L_i}(\sigma_i^l\xi_i^{\mu(l)})\geq 0,\ \forall t,\end{aligned}\tag{29}$$

which ensures the positive definiteness of the augmented performance index (i=1, 2, …, n). Therefore, because of construction similarity, the new augmented Riccati equation (28) subject to the augmented performance index can be solved numerically using current mature algebraic Riccati equation solution algorithms.

Ultimately, the development of the proposed IAFOC scheme is completed within the actor-critic architecture. This scheme as well as the developed online algorithms is presented and depicted in Fig. 2. As shown in Fig. 2, our goals are achieved in the developed actor-critic architecture through two aspects: (1) The adaptive fuzzy critic is designed to regulate a self-learning approximation function $\Gamma(\boldsymbol{x})$ given in Eq. (7) such that it minimizes the value function $\boldsymbol{\psi}(\boldsymbol{x})$ and balance the HJB equation for optimality (9); (2) The adaptation of the optimal feedback control actor $\boldsymbol{u}^*(\boldsymbol{x})$ given in Eq. (6) is constructed using its specific augmented Riccati equation (28) under the adaptive fuzzy critic.

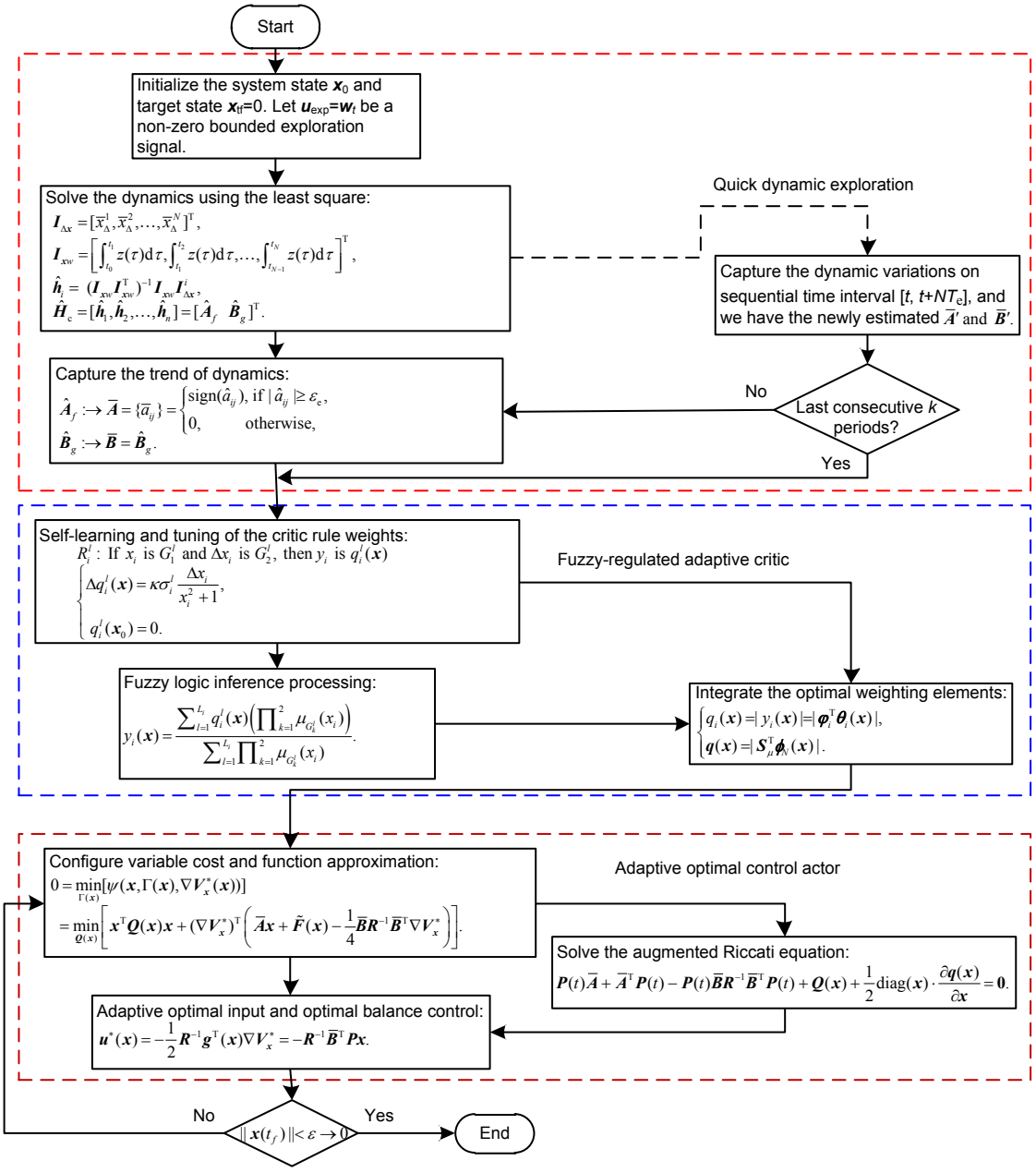


Fig. 2 The proposed IAFOC scheme

4 Stability and convergence of the closed-loop system

Instead of approximating or solving the HJB equation iteratively by the direct ADP method, our methodology is to derive real-time adaptive approximation and balance of the optimal cost and the optimal control policy subject to the HJB equation, and thus guarantees the asymptotic stability of the closed-loop system.

Theorem 1 Consider the nonlinear system (1) satisfying Assumption 1. Let the self-learning and updating law for the adaptive fuzzy-regulated critic be provided by Eqs. (17) and (20). Let $\boldsymbol{u}^*(\boldsymbol{x})$ be the optimal control input (6) and $\boldsymbol{V}_x^*$ be the Lyapunov function (7) for the adaptive control actor characterized by the augmented Riccati equation (28). Based on Lemma 1 and Fact 1, despite the existence of approximation (or balance) error (26), we apply the developed adaptive critic and optimal actor. Then, the

states of the controlled closed-loop system are asymptotically stable and uniformly ultimately bounded in a compact residual set $U(\Omega)\subset\mathbb{R}^n$ and will ultimately converge to zero (the terminal state) (i.e., $\|\boldsymbol{x}(t_f)\|<\varepsilon\to 0$ as $t_f\to\infty$).

Proof Construct the Lyapunov function candidate (11) as

$$\boldsymbol{J}(t)=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}(t)\boldsymbol{x}. \tag{30}$$

Considering the optimal HJB balance error in Eq. (26), the derivative of $\boldsymbol{J}(t)$ is given by

$$\begin{aligned}\dot{\boldsymbol{J}}&=\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{P}\bar{\boldsymbol{A}}+\bar{\boldsymbol{A}}\boldsymbol{P}-2\boldsymbol{P}\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{P})\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\varepsilon}_{\psi}+\boldsymbol{\varepsilon}_{\psi}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}\\&=\boldsymbol{x}^{\mathrm{T}}\left[-\boldsymbol{Q}(\boldsymbol{x})-\frac{1}{2}\mathrm{diag}(\boldsymbol{x})\cdot\frac{\partial\boldsymbol{q}(\boldsymbol{x})}{\partial\boldsymbol{x}}-\boldsymbol{P}\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{P}\right]\boldsymbol{x}\\&\quad+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\varepsilon}_{\psi}+\boldsymbol{\varepsilon}_{\psi}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}.\end{aligned} \tag{31}$$

If the inverse matrix $\boldsymbol{R}^{-1}$ exists and $\boldsymbol{R}=\sqrt{\boldsymbol{R}}\sqrt{\boldsymbol{R}}$, then $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\varepsilon}_{\psi}+\boldsymbol{\varepsilon}_{\psi}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}\le\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}+\boldsymbol{\varepsilon}_{\psi}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}})^{-1}\boldsymbol{R}\boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_{\psi}$. Therefore, we have

$$\begin{aligned}\dot{\boldsymbol{J}}\le\boldsymbol{x}^{\mathrm{T}}&\left[-\boldsymbol{Q}(\boldsymbol{x})-\frac{1}{2}\mathrm{diag}(\boldsymbol{x})\cdot\frac{\partial\boldsymbol{q}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right]\boldsymbol{x}\\&+\boldsymbol{\varepsilon}_{\psi}^{\mathrm{T}}(\bar{\boldsymbol{B}}\boldsymbol{R}^{-1}\bar{\boldsymbol{B}}^{\mathrm{T}})^{-1}\boldsymbol{\varepsilon}_{\psi}.\end{aligned} \tag{32}$$

As the augmented performance index matrix $\boldsymbol{Q}(\boldsymbol{x})+[\mathrm{diag}(\boldsymbol{x})/2]\partial\boldsymbol{q}(\boldsymbol{x})/\partial\boldsymbol{x}$ is positive semi-definite as demonstrated in Eq. (29), inequality (32) guarantees that the Lyapunov candidate $\dot{\boldsymbol{J}}(t)\le 0$ is negative outside the compact set:

$$U(\Omega)=\left\{\|\boldsymbol{x}\|\le\sqrt{\frac{\left\|\boldsymbol{\varepsilon}_{\psi}^{\mathrm{T}}(\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^{\mathrm{T}})^{-1}\boldsymbol{\varepsilon}_{\psi}\right\|^2}{\lambda_{\min}}}\right\}, \tag{33}$$

where $\lambda_{\min}$ is the minimum eigenvalue of the augmented performance index matrix $\boldsymbol{Q}(\boldsymbol{x})+[\mathrm{diag}(\boldsymbol{x})/2]\cdot\partial q(\boldsymbol{x})/\partial\boldsymbol{x}\in\mathbb{R}^{n\times n}$. Therefore, all the closed-loop states and signals $\boldsymbol{x}(t)$ are uniformly convergent and ultimately bounded by the compact set $U(\Omega)$. Furthermore, because of the self-learning adaptive critic design (20), $\lambda_{\min}(\boldsymbol{x})$ is increasingly accumulated as the existence of the system error $x_i\partial q_i/\partial x_i\ge 0$ $(\forall t)$, as demonstrated in Eq. (29). The adaptive adjustment and increase of $\lambda_{\min}(\boldsymbol{x})$ will thus further reduce and contract the domain of $U(\Omega)$ until the system state is ultimately driven to the terminal state $\boldsymbol{x}(t_f)=\boldsymbol{0}$. Thus, it is conclusive that $\forall\varepsilon>0$, $\|\boldsymbol{x}(t_f)\|<\varepsilon\to 0$ as $t_f\to\infty$.

Based on the provided conclusion, the proposed indirect adaptive fuzzy-regulated optimal control scheme will converge to the optimal balance solution of the HJB equation on Ω for the controlled unknown nonlinear system (1).

5 Online implementation and simulation examples

To test the performance and effectiveness of the proposed IAFOC, simulation examples on a power system and Chua's circuit are presented. In both cases, satisfactory control performance and stable convergence are achieved. Table 1 shows the experimental parameters of the proposed controller. An ADP controller as a direct adaptive approach (Vrabie et al., 2009) is introduced for comparison.

5.1 Power system

Consider the continuous-time plant of a power system as

$$\dot{\boldsymbol{x}}=\begin{bmatrix}-0.0665 & 11.5 & 0 & 0\\ 0 & -2.5 & 2.5 & 0\\ -9.5 & 0 & -13.736 & 13.736\\ 0.6 & 0 & 0 & 0\end{bmatrix}\boldsymbol{x}+\begin{bmatrix}0\\0\\13.736\\0\end{bmatrix}\boldsymbol{u}. \tag{34}$$

In this case, Eq. (34) gives a locally linearized model of the real power system characterized by nonlinearities as only the range of the plant parameters can be determined. For the ADP controller, the control cost is prescribed and chosen as identity ($\boldsymbol{Q}=\boldsymbol{I}_4$ and $\boldsymbol{R}=\boldsymbol{I}_1$), and the ADP sampling time interval is set as T_a=0.05 s. For the IAFOC, after the quick dynamic exploration with period of N_eT_e=0.1 s, the estimated system drift dynamics trend $\bar{\boldsymbol{A}}$ and input dynamics $\bar{\boldsymbol{B}}$ for the IAFOC are given as

$$\bar{\boldsymbol{A}}=\begin{bmatrix}-1 & 1 & 0 & 0\\ 0 & -1 & 1 & 0\\ -1 & 0 & -1 & 1\\ 1 & 0 & 0 & 0\end{bmatrix},\quad\bar{\boldsymbol{B}}=\begin{bmatrix}0\\0\\13.7362\\0\end{bmatrix}. \tag{35}$$

Simulation results of case 1 are recorded and illustrated in Figs. 3–5. Comparing the transient responses of the object power system in Fig. 3, the proposed IAFOC (settling time is 4 s) realizes more stable and faster convergence than the ADP controller (settling time is 7 s). Moreover, as shown in Fig. 4, the amplitude range of system input of IAFOC is significantly smaller than that of ADP, which indicates a smoother control process and higher control efficiency. Despite the parametric mismatch of the initial exploration model (35), the IAFOC gradually reduces and minimizes the HJB value function $\boldsymbol{\psi}(\boldsymbol{x})$ (Fig. 5). Through the designed indirect adaptation mechanism, the IAFOC balances and ensures the control optimality of the HJB equation like the direct ADP method.

5.2 Chua's circuit

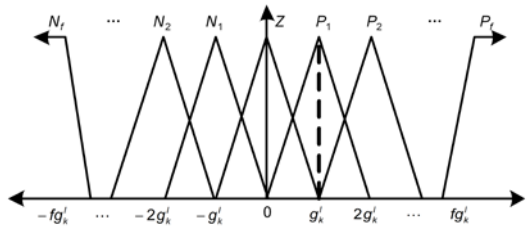
Consider the chaotic Chua circuit with double scroll nonlinearities as

$$\dot{\boldsymbol{x}}=\begin{bmatrix}-1 & 1 & 1\\ 9.78+0.3\cos(\pi t) & -9.78-0.3\cos(\pi t) & 0\\ -14.97-0.4\sin(\pi t) & 0 & 0\end{bmatrix}\boldsymbol{x}+\begin{bmatrix}0\\ -9.78f(x_2)\\ 0\end{bmatrix}+\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\boldsymbol{u}, \tag{36}$$

where $f(x_2)$ is a piecewise nonlinear function expressed as

Table 1 Parameter setup for simulations using IAFOC

System parameter	Numerical value
Starting state $\boldsymbol{x}(t_0)$	$\boldsymbol{x}_0=[0, 0.1, 0, 0]^{\mathrm{T}}$ in case 1; $\boldsymbol{x}_0=[1, -1, 0]^{\mathrm{T}}$ in case 2
Critic rule weights $\{q_i^l(\boldsymbol{x}(t_0))\}$	$\{q_i^l(\boldsymbol{x}(t_0))\}=0,\ l=1,2,\ldots,N$
Exploration input w_{t}	$\boldsymbol{w}_{\mathrm{t}}=[1, 1.2, 1, 1.2]^{\mathrm{T}}d_{\mathrm{w}}(t)$ in case 1; $\boldsymbol{w}_{\mathrm{t}}=[0.3, 0.8, 0.6]^{\mathrm{T}}d_{\mathrm{w}}(t)$ in case 2 ($d_{\mathrm{w}}(t)$ is the random Gaussian white noise with mean 0 and variance 1)
Exploring time interval T_{e}	0.005 s
Input fuzzy sets G_k^l (k=1, 2 for x_i and Δx_i)	$G_k^l=\{N_f,\ldots,N_2,N_1,Z,P_1,P_2,\ldots,P_f\}$ (N: negative; Z: zero; P: positive)
Fuzzy set parameter f and rule number L_i	f=3, $L_i=(2f+1)^2$=49
Membership functions (triangular and equal-span) with $g_k^{l_i}$	$g_1^{l_i}=\{0.5,0.1,0.1,0.1\}$ and $g_2^{l_i}=\{0.01,0.01,0.01,0.01\}$ in case 1; $g_1^{l_i}=\{0.5,0.5,0.5\}$ and $g_2^{l_i}=\{0.01,0.01,0.01\}$ in case 2
Self-learning rate κ_i	$\kappa_i\in\{10, 10, 10, 10\}$ in case 1; $\kappa_i\in\{10, 10, 10\}$ in case 2
Weighting matrix coefficient for control input $\boldsymbol{R}=r\boldsymbol{I}_n$	r=1 in case 1; r=0.1 in case 2
Sampling time interval T_{c}	0.001 s



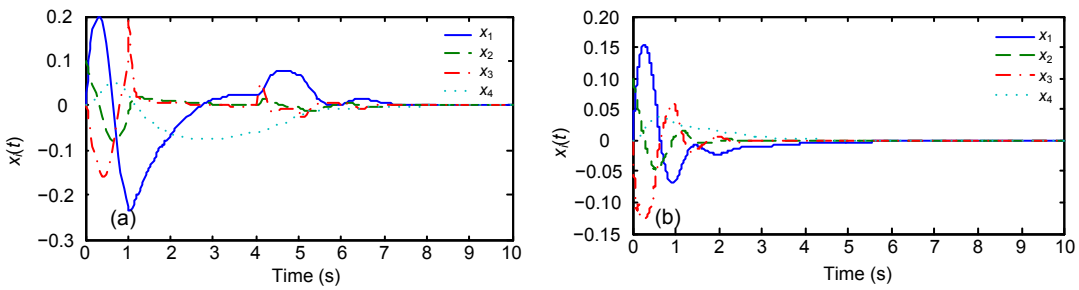


Fig. 3 Power system responses of ADP (a) and IAFOC (b) in case 1

$$f(x_2)=bx_2+0.5(a-b)(|x_2+1|-|x_2-1|) \quad (37)$$

with a=−1.3 and b=−0.75. For the compared ADP controller, the weights of optimal control cost are likewise prescribed as $\boldsymbol{Q}=\boldsymbol{I}_3$ and $\boldsymbol{R}=0.1\boldsymbol{I}_3$ with the ADP sampling time interval T_a=0.05 s. Using the proposed IAFOC, after the quick dynamic exploration with a period of 0.006 s, the system's drift dynamics trend $\bar{\boldsymbol{A}}$ and input dynamics $\bar{\boldsymbol{B}}$ are obtained as

$$\bar{\boldsymbol{A}}=\begin{bmatrix}-1 & 1 & 1\\ 1 & -1 & 0\\ -1 & 0 & 0\end{bmatrix},\ \bar{\boldsymbol{B}}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1.1012 & 0\\ 0 & 0 & 1.0011\end{bmatrix}. \quad (38)$$

Simulation results of case 2 are recorded and illustrated in Figs. 6–8. Note that the case of Chua's circuit is more complex and uncertain than case 1 because of the piecewise nonlinearity, dynamic mismatch, and parametric perturbations. The ADP controller does not provide satisfactory stable performance since the circuit's states are still oscillating persistently, especially for $x_3(t)$ (Fig. 6a). Along with the circuit state deviations, the control input of ADP is likewise not convergent (Fig. 7a). In contrast, the proposed IAFOC achieves straightforward swift and stable convergence of the circuit system (Fig. 6b) and control input (Fig. 7b) (settling time is 0.4 s). As shown in Fig. 8, the ADP controller reduces only the HJB value function $\psi(\boldsymbol{x})$ to a reasonable range, and generates a residual optimality deviation corresponding to its transient responses. As comparison, the proposed IAFOC minimizes and balances the HJB equation smoothly and swiftly, ensures control optimality in the meantime, and preserves advantageous control performance and robustness to dynamic mismatch, uncertainty, and perturbation.

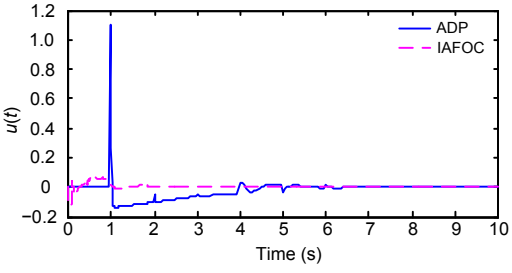


Fig. 4 Optimal control input in case 1

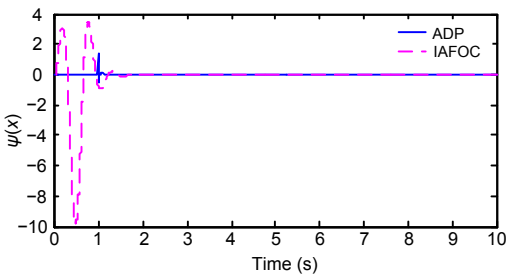


Fig. 5 Evolution of HJB value function $\psi(x)$ in case 1

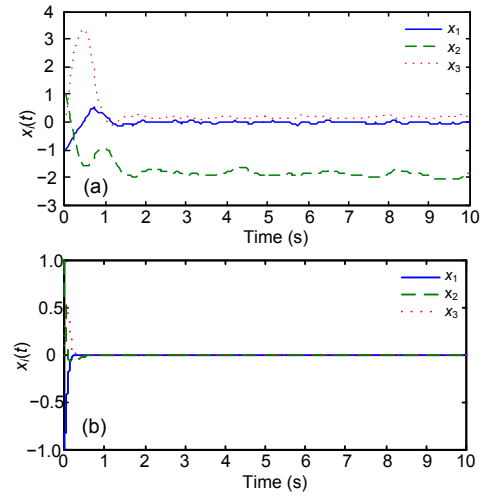


Fig. 6 Chua's circuit responses of ADP (a) and IAFOC (b) in case 2

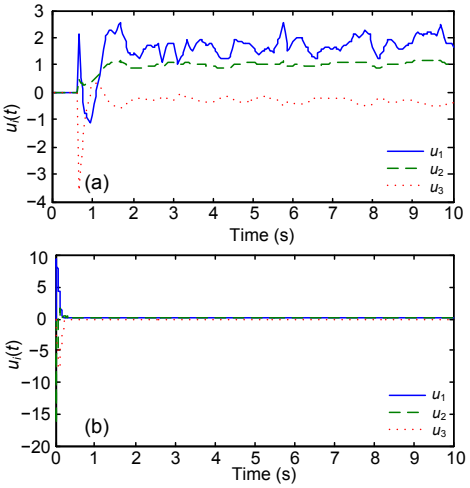


Fig. 7 Optimal control inputs of ADP (a) and IAFOC (b) in case 2

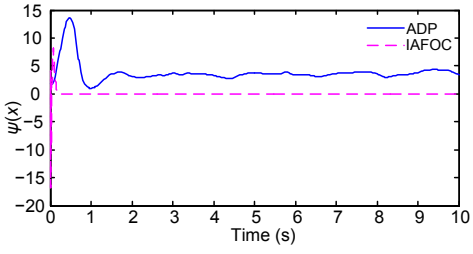


Fig. 8 Evolution of HJB value function $\psi(x)$ in case 2

Considering the results in two diverse cases, in comparison with ADP, the proposed IAFOC approach demonstrates outstanding performance and better transient control response, and guarantees the asymptotic stability of the unknown nonlinear system without requiring persistent excitation or any initial stabilizing conditions. Importantly, considering the indirect adaptation principle as demonstrated in the evolution of the HJB value function $\psi(\boldsymbol{x})$ (Figs. 5 and 8), the proposed IAFOC uniformly reduces the value function to zero after a transient learning and adjustment period (about 1.2 s in case 1 and 0.2 s in case 2). The results indicate that the involved HJB equation has been approximated and balanced autonomously, while the control optimality is adaptively achieved and finally ensured.

For the implementation of IAFOC, quick dynamic exploration is needed to primarily derive the system's drift dynamics trend and the input-to-state dynamics as the initial conditions. The accuracy of dynamic exploration will influence control performance as the estimation results are essentially required in the optimal control calculation. Thus, our future work will include the design of quick, precise, and easily implemented dynamic estimation (exploration) techniques for more complicated situations.

6 Conclusions

In this study, we have presented the development of an online indirect adaptive fuzzy-regulated optimal control (IAFOC) scheme for continuous-time nonlinear systems with unknown dynamics. The distinctive features and performance of the IAFOC scheme have been studied through theoretical analysis and simulations. Conclusions are as follows:

1. Without iteratively solving the HJB equation, a novel indirect adaptation methodology of approximately minimizing and balancing the nonlinear HJB equation has been proposed for the first time using the actor-critic architecture in the proposed scheme.

2. Using a fuzzy-regulated performance function as reinforcement medium, the supervisory adaptive critic associated with the self-organizing weight learning law ensured online approximation and simultaneous adjustment of both the optimal cost and the optimal control policy.

3. For the control actor, an adaptive optimal feedback controller has been constructed with a new augmented Riccati governing equation, and fast and stable convergence of nonlinear system states and parameters has been achieved autonomously under diverse conditions.

4. Compared with the direct ADP method, the proposed IAFOC provided not only better control performance and superior transient responses, but also advantageous robustness to dynamic mismatch, uncertainty, and perturbations.

On the other hand, mathematical proof showed the effectiveness and closed-loop stability of the control scheme in the Lyapunov sense. Through the designed indirect adaptation mechanism, the IAFOC effectively minimized and balanced the HJB equation, and therefore guaranteed control optimality and asymptotic stability. How to design a quick and precise dynamic estimation technique for the control initialization is one of our future research directions.

Contributors

Haiyun ZHANG and Jin WANG designed and conducted the research. Deyuan MENG processed the data. Haiyun ZHANG drafted the manuscript. Guodong LU helped organize the manuscript. Haiyun ZHANG and Jin WANG revised and finalized the paper.

Compliance with ethics guidelines

Haiyun ZHANG, Deyuan MENG, Jin WANG, and Guodong LU declare that they have no conflict of interest.

References

Abu-Khalaf M, Lewis FL, 2005. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 41(5):779-791. https://doi.org/10.1016/j.automatica.2004.11.034

Bhasin S, Kamalapurkar R, Johnson M, et al., 2013. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica*, 49(1):82-92. https://doi.org/10.1016/j.automatica.2012.09.019

Bian T, Jiang ZP, 2016. Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design. *Automatica*, 71:348-360. https://doi.org/10.1016/j.automatica.2016.05.003

Chang XH, Yang C, Xiong J, 2019. Quantized fuzzy output feedback H_∞ control for nonlinear systems with adjustment of dynamic parameters. *IEEE Trans Syst Man Cybern Syst*, 49(10):2005-2015. https://doi.org/10.1109/TSMC.2018.2867213.

Chang Y, Wang YQ, Alsaadi FE, et al., 2019. Adaptive fuzzy output-feedback tracking control for switched stochastic

pure-feedback nonlinear systems. *Int J Adapt Contr Signal Process*, 33(10):1567-1582. https://doi.org/10.1002/acs.3052

Finlayson BA, 1990. The Method of Weighted Residuals and Variational Principles. Academic Press, New York, USA.

Huo X, Ma L, Zhao XD, et al., 2020. Event-triggered adaptive fuzzy output feedback control of MIMO switched nonlinear systems with average dwell time. *Appl Math Comput*, 365:124665. https://doi.org/10.1016/j.amc.2019.124665

Ioannou PA, Fidan B, 2006. Advances in Design and Control. Adaptive Control Tutorial. SIAM, Philadelphia, USA.

Jiang Y, Jiang ZP, 2012. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, 48(10):2699-2704. https://doi.org/10.1016/j.automatica.2012.06.096

Jiang Y, Jiang ZP, 2014. Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. *IEEE Trans Neur Netw Learn Syst*, 25(5):882-893. https://doi.org/10.1109/TNNLS.2013.2294968

Kiumarsi B, Lewis FL, Modares H, et al., 2014. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. *Automatica*, 50(4):1167-1175. https://doi.org/10.1016/j.automatica.2014.02.015

Lee JM, Lee JH, 2004. Approximate dynamic programming strategies and their applicability for process control: a review and future directions. *Int J Contr Autom Syst*, 2(3):263-278.

Lee JY, Park JB, Choi YH, 2012. Integral Q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems. *Automatica*, 48(11):2850-2859. https://doi.org/10.1016/j.automatica.2012.06.008

Lee JY, Park JB, Choi YH, 2015. Integral reinforcement learning for continuous-time input-affine nonlinear systems with simultaneous invariant explorations. *IEEE Trans Neur Netw Learn Syst*, 26(5):916-932. https://doi.org/10.1109/TNNLS.2014.2328590

Lewis FL, Vrabie DL, Syrmos VL, 2012a. Optimal Control (3rd Ed.). Wiley, Hoboken, USA.

Lewis FL, Vrabie D, Vamvoudakis KG, 2012b. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. *IEEE Contr Syst Mag*, 32(6):76-105. https://doi.org/10.1109/MCS.2012.2214134

Li YM, Tong SC, Li TS, 2016. Hybrid fuzzy adaptive output feedback control design for uncertain MIMO nonlinear systems with time-varying delays and input saturation. *IEEE Trans Fuzzy Syst*, 24(4):841-853. https://doi.org/10.1109/TFUZZ.2015.2486811

Lin WS, 2011. Optimality and convergence of adaptive optimal control by reinforcement synthesis. *Automatica*, 47(5):1047-1052. https://doi.org/10.1016/j.automatica.2011.01.060

Liu DR, Wei QL, 2013. Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems. *IEEE Trans Cybern*, 43(2):779-789. https://doi.org/10.1109/TSMCB.2012.2216523

Liu DR, Yang X, Li HL, 2013. Adaptive optimal control for a class of continuous-time affine nonlinear systems with unknown internal dynamics. *Neur Comput Appl*, 23(7):1843-1850. https://doi.org/10.1007/s00521-012-1249-y

Liu DR, Wang D, Wang FY, et al., 2014. Neural-network-based online HJB solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems. *IEEE Trans Cybern*, 44(12):2834-2847. https://doi.org/10.1109/TCYB.2014.2357896

Ma L, Huo X, Zhao XD, et al., 2019. Adaptive fuzzy tracking control for a class of uncertain switched nonlinear systems with multiple constraints: a small-gain approach. *Int J Fuzzy Syst*, 21(8):2609-2624. https://doi.org/10.1007/s40815-019-00708-9

Modares H, Lewis FL, 2014. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica*, 50(7):1780-1792. https://doi.org/10.1016/j.automatica.2014.05.011

Modares H, Naghibi Sistani MB, Lewis FL, 2013. A policy iteration approach to online optimal control of continuous-time constrained-input systems. *ISA Trans*, 52(5):611-621. https://doi.org/10.1016/j.isatra.2013.04.004

Murray JJ, Cox CJ, Lendaris GG, et al., 2002. Adaptive dynamic programming. *IEEE Trans Syst Man Cybern Part C*, 32(2):140-153. https://doi.org/10.1109/TSMCC.2002.801727

Padhi R, Unnikrishnan N, Wang XH, et al., 2006. A Single Network Adaptive Critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems. *Neur Netw*, 19(10):1648-1660. https://doi.org/10.1016/j.neunet.2006.08.010

Powell WB, 2007. Approximate Dynamic Programming: Solving the Curses of Dimensionality. Wiley, New York, USA.

Sastry SS, 1999. Nonlinear Systems: Analysis, Stability, and Control. Springer-Verlag, New York, USA.

Slotine JE, Li W, 1991. Applied Nonlinear Control. Prentice Hall, Englewood Cliffs, NJ, USA.

Song RZ, Xiao WD, Zhang HG, et al., 2014. Adaptive dynamic programming for a class of complex-valued nonlinear systems. *IEEE Trans Neur Netw Learn Syst*, 25(9):1733-1739. https://doi.org/10.1109/TNNLS.2014.2306201

Tao G, 2003. Adaptive Control Design and Analysis. In: Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley-Interscience, Hoboken, NJ, USA.

Vamvoudakis KG, 2017. Q-learning for continuous-time linear systems: a model-free infinite horizon optimal control approach. *Syst Contr Lett*, 100:14-20. https://doi.org/10.1016/j.sysconle.2016.12.003

Vamvoudakis KG, Lewis FL, 2010. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 46(5):878-888. https://doi.org/10.1016/j.automatica.2010.02.018

van der Schaft AJ, 1992. L_2-gain analysis of nonlinear systems and nonlinear state-feedback H_1 control. *IEEE Trans Autom Contr*, 37(6):770-784. https://doi.org/10.1109/9.256331

Vrabie D, Pastravanu O, Abu-Khalaf M, et al., 2009. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477-484. https://doi.org/10.1016/j.automatica.2008.08.017

Wang FY, Zhang HG, Liu DR, 2009. Adaptive dynamic programming: an introduction. *IEEE Comput Intell Mag*, 4(2):39-47. https://doi.org/10.1109/MCI.2009.932261

Wei QL, Zhang HG, Dai J, 2009. Model-free multiobjective approximate dynamic programming for discrete-time nonlinear systems with general performance index functions. *Neurocomputing*, 72(8-9):1839-1848. https://doi.org/10.1016/j.neucom.2008.05.012

Werbos P, 2004. ADP: goals, opportunities and principles. In: Si J, Barto A, Powell W, et al. (Eds.), Handbook of Learning and Approximate Dynamic Programming. Institute of Electrical and Electronics Engineers, New York, USA, p.3-44. https://doi.org/10.1002/9780470544785.ch1

Yang X, He HB, 2018. Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances. *Neur Netw*, 99:19-30. https://doi.org/10.1016/j.neunet.2017.11.022

Yang X, Liu DR, Luo B, et al., 2016. Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning. *Inform Sci*, 369:731-747. https://doi.org/10.1016/j.ins.2016.07.051

Yang XY, Liu DR, Huang YZ, 2013. Neural-network-based online optimal control for uncertain non-linear continuous-time systems with control constraints. *IET Contr Theory Appl*, 7(17):2037-2047. https://doi.org/10.1049/iet-cta.2013.0472

Yin YF, Zhao XD, Zheng XL, 2017. New stability and stabilization conditions of switched systems with mode-dependent average dwell time. *Circ Syst Signal Process*, 36(1):82-98. https://doi.org/10.1007/s00034-016-0306-7

Yu ZX, Yang YK, Li SG, et al., 2018. Observer-based adaptive finite-time quantized tracking control of nonstrict-feedback nonlinear systems with asymmetric actuator saturation. *IEEE Trans Syst Man Cyber Syst*, 50(11): 545-4556. https://doi.org/10.1109/TSMC.2018.2854927

Zak SH, 2003. Systems and Control. Oxford University Press, New York, USA.