



Survey

A survey of adaptive optimal control theory

Xiaoxuan Pei, Kewen Li and Yongming Li*

College of Science, Liaoning University of Technology, Jinzhou 121001, China

* **Correspondence:** Email: l_y_m_2004@163.com.

Abstract: This paper makes a survey about the recent development of optimal control based on adaptive dynamic programming (ADP). First of all, based on DP algorithm and reinforcement learning (RL) algorithm, the origin and development of the optimization idea and its application in the control field are introduced. The second part introduces achievements in the optimal control direction, then we classify and summarize the research results of optimization method, constraint problem, structure design in control algorithm and practical engineering process based on optimal control. Finally, the possible future research topics are discussed. Through a comprehensive and complete investigation of its application in many existing fields, this survey fully demonstrates that the optimal control algorithms via ADP with critic-actor neural network (NN) structure, which also have a broad application prospect, and some developed optimal control design algorithms have been applied to practical engineering fields.

Keywords: optimal control; ADP; backstepping design; neural networks; application

1. Research and development of optimization control algorithms

1.1. Origin and development of optimal control design

The research on optimal control for nonlinear systems plays a significant role in industry and military fields. Due to the influence of the environment and the limitations of the engineering system, it is very tough to maximize or minimize the performance index of the controlled system in practical. Therefore, optimization problem is a difficult problem in current control field, and has gradually been the focus of attention. The optimal control problems of nonlinear systems are finally transformed into the solutions of Hamilton-Jacobi-Bellman (HJB) partial differential equations.

However, because the HJB equation is a nonlinear partial differential equation, it is difficult to obtain an analytical solution. Therefore, how to obtain the analytical solution of HJB equation then realize the optimization performance index of the system, is the key points to resolve the optimization issue.

In order to avoid the problems encountered in solving HJB equation, Kalman [1] proposed the inverse-optimal-based control method for the first time. On the basis of [1], Freeman and Kokotovic [2] studied inverse optimization control of nonlinear systems. The basic idea of inverse optimization control is not to minimize the cost function by designing the controller, but to minimize the cost function by designing the appropriate control Lyapunov function (CLF). Therefore, the solution of HJB equation comes down to seeking the CLF of the controlled system, thus avoiding the shortcoming of directly solving HJB equation.

Besides, aiming at the above problems, Bellman [3] proposed the theory of dynamic programming (DP). However, the issue of “dimension disaster” would be caused in the process of DP control design, that is, the complexity of space storage and computation increases exponentially with the increase of the dimension of control vectors and states. Therefore, in order to overcome the phenomenon of “dimension disaster” in the process of optimal control design, an adaptive optimal control design method combining NNs is proposed by Werbos [4], which is called RL or adaptive/approximate dynamic programming (ADP). In the control field, RL can effectively solve the “dimension disaster” problem in DP. In [5], Werbos retrospected the classic econometric approach and proposed a robust method. In [6], Werbos defined a more limited design called “brain-like intelligent control”, it discusses the brain as a member of intelligent control, which implies a property to be sought in future research.

1.2. Development of optimization control algorithms

Since then, inspired by [4], large amounts of optimal control methods via APD have been developed, see [7, 8]. Among them, for continuous-time (CT) systems, in offline situations, Abu-Khalaf and Lewis [9] presented an offline algorithm via RL to solve the optimal control issue of CT nonlinear systems. Since the offline control algorithms cannot be adopted to adjust online in real time, thus, to overcome this disadvantage, Vamvoudak and Lewis [10] proposed an online adaptive method via policy iteration. In [11], Li et al. investigated the Lyapunov stability problem for impulsive systems via event-triggered impulsive control. In [12], Li et al. considered a class of nonlinear impulsive systems with delayed impulses, based on impulsive control theory and the ideas of average dwell-time (ADT), a set of Lyapunov-based sufficient conditions for globally exponential stability were obtained. However, we need to know the accurate knowledge of CT nonlinear systems in [9, 10]. Since nonlinear systems usually contain uncertain nonlinear functions, it is difficult to acquire the analytical solution of the HJB equation.

In order to solve this problem, by choosing an appropriate cost function to reflect uncertainty regulation, the authors in [13] proposed a robust optimization controller design strategy based on an online strategy iterative algorithm for a class of continuous nonlinear systems with nonlinearities. Zhang et al. [14] designed a new data-driven robust identified optimization tracking controller via the acquired data-driven model for a kind of nonlinear CT systems.

2. Development of the adaptive optimal control for affine nonlinear systems

2.1. Design of optimal control for affine nonlinear systems

Consider a class of affine nonlinear system as:

$$\dot{x}(t) = g(x(t))u(t) + f(x(t)), x(0) = x_0 \quad (2.1)$$

where $f(x)$ and $g(x)$ are the uncertain smooth functions, which satisfy that $f(0) = 0$, $g(0) = 0$. u is the control input, $x \in \mathbf{R}^n$ is the state vector. For the above system, some adaptive optimization control strategies have been proposed.

In [15], for a class of affine nonlinear CT systems with unknown internal dynamics, Liu and Wang et al. developed an online method based on ADP, which constructed a critic neural network to facilitate the solution of the modified HJB equation. In [16], Liu et al. developed an online optimization control algorithm for CT affine nonlinear systems with infinite horizon cost. And in [17], Wen and Chen et al. studied an adaptive optimized tracking control method via RL algorithm and NNs.

For [16, 17], value function is selected as:

$$V(z) = \int_t^\infty r(z(r), u(z)) d\tau \quad (2.2)$$

where $r(z, u) = z^T(t)Q(x)z(t) + u^T u$ is the value function, and $Q(x) = q(x)q^T(x) \in \mathbf{R}^{n \times n}$ is a positive definite matrix. The HJB equation is defined as:

$$\begin{aligned} H(z, u, V_z) &= V_z^T(z)z(t) + r(z, u) \\ &= V_z^T(f(x) + g(x)u - y_d(t)) + z^T Q(x)z(t) + u^T u \end{aligned} \quad (2.3)$$

where $V_z \in \mathbf{R}^n$ is the partial gradient of V_z , $y_d(t) \in \mathbf{R}_n$ is the ideal tracking trajectory, $z(t) = x(t) - y_d(t)$ is the tracking error. When considering the input constraint control and saturation constraints control issue, Liu and Yang developed a robust optimal adaptive optimal control method via RL for a kind of uncertain nonlinear systems. In [18, 19], there exists a symmetric definite matrix Q , and the value function is selected as:

$$V(x(t)) = \int_t^\infty [x^T Qx + \varpi(u)] ds, (s \geq t) \quad (2.4)$$

where $\varpi(u)$ is positive. For the sake of solving the constraint control issue, define $\varpi(u)$ as:

$$\begin{aligned} \varpi(u) &= 2\kappa \int_0^u (\psi^{-1}(v/\kappa))^T R dv \\ &= 2\kappa \sum_{i=1}^m \int_0^u (\psi^{-1}(v_i/\kappa))^T R_i dv \end{aligned} \quad (2.5)$$

where $R = \text{diag}[r_1, \dots, r_n]$ with $r_i > 0$, ($i = 1, \dots, m$), $\psi(\cdot)$ is a bounded one-to-one function with $|\psi(\cdot)| \leq 1$, $\psi \in \mathbf{R}^m$, $\psi^{-1} = (\psi^{-1})^T$, $\psi^{-1}(v/\kappa) = [\psi^{-1}(v_1/\kappa), \dots, \psi^{-1}(v_m/\kappa)]^T$, $u(x) \in \Xi$, $\Xi = \{u|u \in \mathbf{R}^m, |u_i| \leq \kappa, i = 1, 2, \dots, m\}$, $\kappa > 0$ is a constant. Define the HJB equation and the value function as:

$$H(x, V_x, u) = V_x^T(f(x) + g(x)u) + r(x, u) \quad (2.6)$$

where $V_x \in \mathbf{R}^n$ is the partial derivative of $V(x)$ with respect to x .

In the previous article, since there exist the unknown nonlinear functions $f(x)$ and $g(x)$, the analytic solution of the equation cannot be received when resolving the HJB equation. Because of their properties and fault tolerance, attributes of nonlinearity, adaptivity, the identified solution of the HJB equation can be obtained symmetric via NNs.

2.2. Development of identifier-actor-critic-based optimization control

Because the system (2.1) contains unknown dynamics, we can identify the system for receiving the optimal control. In [20], Yang et al. presented identifier-actor-critic (IAC) structure, where the actor NN is carried out control actions, and critic NN is employed to estimated these actions, and then returns the evaluations to actor, and the dynamics of uncertain system robust dynamic can be approximate by NN identifiers. From system (2.1), we have that:

$$\dot{x} = g(x)u + f(x) = g(x)u + Ax + F(x) \quad (2.7)$$

where $A \in \mathbf{R}^{n \times n}$ is a certain constant matrix, $F(x) = f(x) - Ax$.

A NN is applied to identify $F(x)$ as follows:

$$F(x) = W_1^T \sigma(x) + \varepsilon_1(x) \quad (2.8)$$

where $\varepsilon_1(x) \in \mathbf{R}^n$ is the NN function reestablishment error, $\sigma(x)$ is the activation function, $W_1^T \in \mathbf{R}^{n \times n}$ is the NN weight. By using (2.8), (2.7) can be developed by:

$$\dot{x}(t) = g(x)u + \varepsilon_1(x) + Ax + W_1^T \sigma(x) \quad (2.9)$$

The NN identifier is designed as:

$$\hat{\dot{x}}(t) = g(\hat{x})u + v(t) + A\hat{x} + \hat{W}_1^T \sigma(\hat{x}) \quad (2.10)$$

where $\hat{x} \in \mathbf{R}^n$ is the identifier NN state, $\hat{W}_1 \in \mathbf{R}^{n \times n}$ is weight estimation, and $v(t)$ is the robust feedback term.

The optimal value function can be expressed by NN as:

$$V^*(\hat{x}) = W^T \phi(\hat{x}) + \varepsilon_v(\hat{x}) \quad (2.11)$$

The optimal control can be expressed by NN as:

$$u^*(\hat{x}) = -\frac{1}{2}R^{-1}g^T(\hat{x})(\phi'(\hat{x})^T W + \varepsilon'_v(\hat{x})^T) \quad (2.12)$$

where $\varepsilon_v(\cdot) \in R$ is the function reestablishment error, $\phi(\hat{x}) = [\phi_1(\hat{x}), \phi_2(\hat{x}), \dots, \phi_N(\hat{x})]^T \in \mathbf{R}^N$, $\phi'(\hat{x}) = \frac{\Delta \phi(\hat{x})}{\Delta \hat{x}}$ and $W \in \mathbf{R}^N$ are uncertain desired NN weights, N is the number of neurons.

The critic-actor $\hat{V}(\hat{x})$ and \hat{u} , which can learn the optimization value function and adjust the optimization control online, is expressed as:

$$\hat{V}(\hat{x}) = \hat{W}_c^T \phi(\hat{x}) \quad (2.13)$$

$$\hat{u}(\hat{x}) = -\frac{1}{2}R^{-1}g^T(\hat{x})\phi'(\hat{x})\hat{W}_a \quad (2.14)$$

where $\hat{W}_c(t) \in \mathbf{R}^N$ and $\hat{W}_a(t) \in \mathbf{R}^N$ estimate the ideal weights of the critic-actor NNs. Whereas the system dynamics are estimated online by using the identification error $\tilde{x}(t) = x(t) - \hat{x}(t)$. The overall planning diagram of the control algorithm is given in Figure 1.

can ensure that each subsystem can be optimized. Based on [23], for a kind of nonlinear large-scale systems with strict-feedback structure, Tong et al. [24] proposed the fuzzy decentralized adaptive optimal control, and used FLS to identify the uncertain nonlinear function of the systems. And in [25], for a quarter of the car active electric suspension systems, Li et al. addressed the output-feedback adaptive NN optimization control issue.

Because there are unknown nonlinear functions, the updating laws and learning laws designed for the above systems are very complex. In order to solve this problem, in [26], Wen et al. proposed a simplified RL algorithm, which generates a negative gradient of a simple positive function from the partial derivative of HJB equation, and derives a new law from the negative gradient.

Define the Hamiltonian's approximation error as:

$$\begin{aligned} E &= H(\hat{z}, u, \hat{V}_{\hat{z}}^*) - H(z, u^*, V_{\hat{z}}^*) \\ &= H(\hat{z}, u, \hat{V}_{\hat{z}}^*) \end{aligned} \quad (3.2)$$

where $V_{\hat{z}}^*(\hat{z})$ is the gradient of $V^*(\hat{z})$, u^* is the optimal control. Since $V_{\hat{z}}^*(\hat{z})$ and u^* contain the unknown part $V_{\hat{z}}^0(\hat{z})$, which can be approximated on a compact set by NNs as:

$$V_{\hat{z}}^0(\hat{z}) = \Theta_V^{*T} \varphi_V(\hat{z}) + \varepsilon_V(\hat{z}) \quad (3.3)$$

Since Θ_V^* is an uncertain constant vector, it is not available in practical control, RL algorithm is implemented by both critic-actor NNs.

The learning law of critic NN is designed as:

$$\dot{\hat{\Theta}}_{Vc}(t) = -k_c \varphi_V(\hat{z}) \varphi_V^T(\hat{z}) \hat{\Theta}_{Vc}(t) \quad (3.4)$$

where k_c is the critic network learning rate, $\hat{\Theta}_{Vc}(t)$ is the critic NN weight.

The learning law of actor NN is designed as:

$$\dot{\hat{\Theta}}_{Va}(t) = -\varphi_V(\hat{z}) \varphi_V^T(\hat{z}) (k_a (\hat{\Theta}_{Va}(t) - \hat{\Theta}_{Vc}(t)) + k_c \hat{\Theta}_{Vc}(t)) \quad (3.5)$$

where k_a is the actor network learning rate, $\hat{\Theta}_{Va}(t)$ is the actor NN weight, $k_a > k_c > 0$.

In accordance with the above description, the optimal solution $\hat{a}(\hat{z})$ is supposed to meet $E(t) = H(\hat{z}, u, \hat{V}_{\hat{z}}^*) \rightarrow 0$.

If $H(\hat{z}, u, \hat{V}_{\hat{z}}^*)$ is held and exist the unique solution, then it is equivalent to the following equation holds:

$$\frac{\partial H(\hat{z}, u, \hat{V}_{\hat{z}}^*)}{\partial \hat{\Theta}_{Va}} = \varphi_V \varphi_V^T (\hat{\Theta}_{Va}^T(t) - \hat{\Theta}_{Vc}^T(t)) = 0 \quad (3.6)$$

The positive definite function is designed as:

$$P(t) = (\hat{\Theta}_{Va}(t) - \hat{\Theta}_{Vc}(t))^T (\hat{\Theta}_{Va}(t) - \hat{\Theta}_{Vc}(t)) \quad (3.7)$$

Clearly, the Eq (3.6) is the equivalent to $P(t) = 0$. Since $\frac{\partial P(t)}{\partial \hat{\Theta}_{Va}(t)} = -\frac{\partial P(t)}{\partial \hat{\Theta}_{Vc}(t)} = 2(\hat{\Theta}_{Va}(t) - \hat{\Theta}_{Vc}(t))$, we

can get

$$\begin{aligned}
 \frac{\partial P(t)}{dt} &= \frac{\partial P(t)}{\partial \hat{\Theta}_{Vc}(t)} \cdot \hat{\Theta}_{Vc}(t) + \frac{\partial P(t)}{\partial \hat{\Theta}_{Va}(t)} \cdot \hat{\Theta}_{Va}(t) \\
 &= -k_c \frac{\partial P(t)}{\partial \hat{\Theta}_{Vc}(t)} \varphi_V \varphi_V^T \hat{\Theta}_{Vc}^T(t) \\
 &\quad - \frac{\partial P(t)}{\partial \hat{\Theta}_{Vc}(t)} \varphi_V \varphi_V^T [k_a (\hat{\Theta}_{Va}(t) - \hat{\Theta}_{Vc}(t)) + k_c \hat{\Theta}_{Vc}(t)] \\
 &= -\frac{k_a}{2} \frac{\partial P(t)}{\partial \hat{\Theta}_a(t)} \varphi_V \varphi_V^T \frac{\partial P(t)}{\partial \hat{\Theta}_a(t)} \leq 0
 \end{aligned} \tag{3.8}$$

In [27], for nonlinear lithium battery systems, Pei et al. addressed adaptive NN output feedback optimization control problem, and the stability of the nonlinear lithium battery is proved.

On the basis of [26], under the frame of backstepping control, some simplified-based adaptive optimization control algorithms have been proposed, which require construct all intermediate control functions and the actual control function of backstepping to be the optimization controls, hence, RL is performed in each subsystem (see Figure 2).

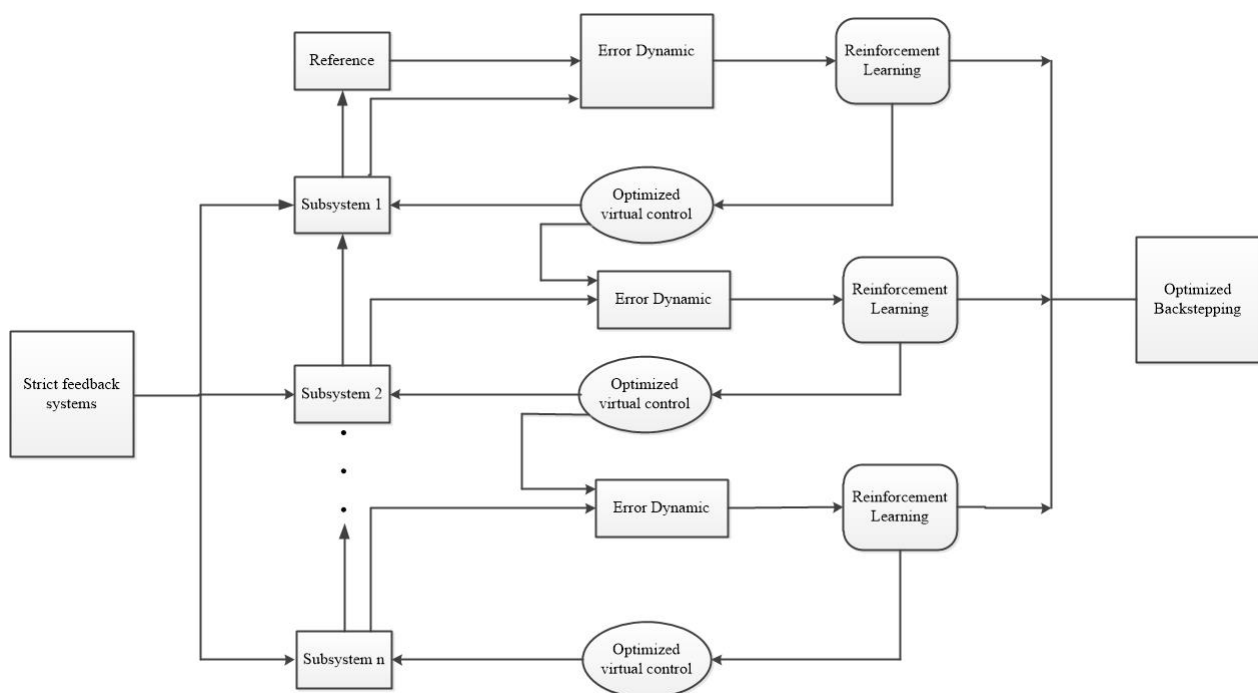


Figure 2. The block diagram of optimization control method based on backstepping.

In [28], Wen et al. addressed optimization control method for nonlinear strict-feedback systems with unknown functions. In [29], for second-order unknown nonlinear multiagent systems, Lan et al. proposed a distributed time-varying optimization formation protocol based on an adaptive NN state observer. In [30], Xiao et al. addressed the distributed optimization containment control issue for multiple nonholonomic mobile robots differential game.

3.2. State-constrained optimal control based on backstepping

It is worth mentioning that system states usually need to be confined within some preselected compact sets due to the physical limitations of actual systems. For real systems, in [31], Jiang and Lou considered the input-to-state stability (ISS) of delayed systems with bounded-delay impulses. In [32], for a hydraulic servo actuator (HSA) with sensor faults, Vladimir and Ljubisa investigated the mechanism for the fault estimation (FE) problem. However, methods in [29, 30] could not solve the actual constraint problem. To solve this problem, various state-constrained control methodologies is discussed.

Aiming at strict-feedback nonlinear systems, which contain immeasurable states and internal dynamics, Li et al. [33] proposed an output-feedback adaptive NN optimization control design. Under the backstepping control design, there will be coupling terms or cross terms at each step, which will lead to that each subsystem is not optimal. Therefore, state constraints should be introduced to make the coupling terms bounded to ensure that each subsystem is optimal. And all the states are limited in the compact sets, that is, $|x_i| < k_{ci}$, where $k_{ci} > 0$.

The neoteric barrier optimization performance index functions for subsystems are designed to ensure that the system state does not violate the constraint bounds and achieves the optimization control objective, which is selected as:

$$J(z(t)) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_t^\tau q(z(t), \alpha(z)) dz \quad (3.9)$$

where τ is the terminal time, $q(z, \alpha) = \xi \log[k_b^4 / (k_b^4 - z^4)] + r(\alpha)^2$, $\xi > 0$ is a constant, α is the intermediate control function, The following Hamiltonian can be derived as:

$$H(\hat{z}, u, \hat{V}_{\hat{z}}^*) = \xi \log \frac{k_b^4}{(k_b^4 - z^4)} + r(\alpha)^2 + \frac{dV^*(z)}{dz} (\alpha^* + g(x) - y_r) \quad (3.10)$$

According to the algorithm presented in [34], for power systems with stochastic character, Li et al. designed the adaptive NN optimal tracking control to resolve the issue of state constraints and uncertain nonlinear dynamics. In [35], Li et al. put forward an adaptive NN optimized output-feedback control method to solve the issue of unknown nonlinear dynamics and input saturation. In [36], for uncertain nonlinear systems with time-varying full state constraints, input saturation and unknown control direction, Wu and Xie employed asymmetric barrier Lyapunov functions, the auxiliary subsystem and the Nussbaum gain technique.

Based on the above published works, some adaptive optimal control methods via backstepping control have also been applied to practical systems, for example, see [37,38]. In [37], Li et al. presented an adaptive NN optimized control strategy for full vehicle active suspension system. And in [38], Li et al. studied adaptive optimal formation control approach for second-order stochastic multi-agent system, which contains unknown nonlinear dynamics.

3.3. Inverse optimization control based on backstepping

Based on the inverse optimization control method in [1], Ezal et al. [38] proposed a new robust backstepping inverse optimal control design, which achieved both local optimization and global inverse

optimization. For a class of nonlinear uncertain strict feedback systems, Li et al. [39] designed adaptive fuzzy inverse optimization control by establishing an equivalent system and an auxiliary system.

System (2.1) can be rewritten as the following nonlinear system:

$$\dot{x} = G(x)u + F(x) + q(x) \quad (3.11)$$

where $u \in \mathbf{R}$ is the control input, x is the state vector, $x = 0$ is the equilibrium point of system. $q(x)$ is an uncertain bounded function vector, $G(x)$ and $F(x)$ are smooth function vectors.

Define γ is a class K_∞ function, then the derivative of γ exists and it is also a class K_∞ function. An auxiliary system is constructed for the nonlinear system (2.1):

$$\dot{x} = l\gamma(2|L\Delta V|R(x)) \times \frac{R^{-2}(x)(L\Delta V)^T}{(L\Delta V)^2} + F(x) + G(x)u \quad (3.12)$$

where $V(x)$ is the control Lyapunov function. $L \Delta V = \partial \Delta V / \partial x$, $L_F V_n = \partial V_n / \partial x F(x)$, $L_G V_n = \partial V_n / \partial x G(x)$.

The cost functional is selected as:

$$J(u) = \sup_{d \in D} \{ \lim_{t \rightarrow \infty} [\int_0^t (l(x)) + u^T R u - \gamma(d) d\tau + E(x)] \} \quad (3.13)$$

where D is a set of locally bounded functions of x , $R(x)$ is matrix-valued function, which satisfied that $R(x) = R(x)^T > 0$. $E(x)$ and $l(x)$ are positive definite radially unbounded functions.

The fuzzy adaptive inverse optimization control structure is shown in Figure 3.

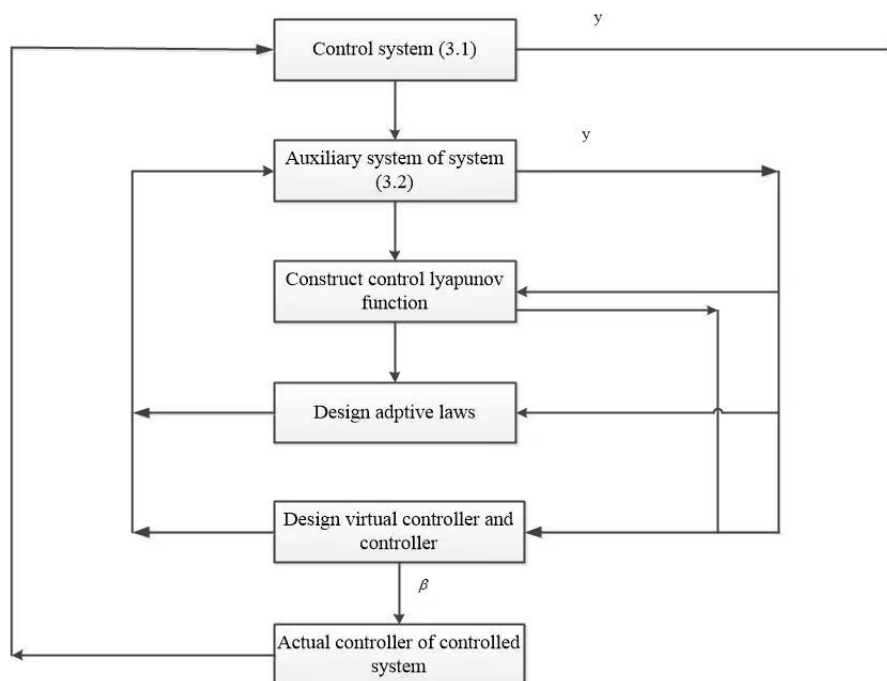


Figure 3. Block diagram of the inverse optimization control structure.

And in [40], Li et al. studied a fuzzy inverse optimization fuzzy adaptive output feedback control method based on observer for a class of nonlinear strict feedback systems. In [41], Lu et al. addressed a fuzzy adaptive inverse optimization control issue, and a switching inverse optimization controller is constructed by using a single parameter learning mechanism, which confirmed that the method guarantees the input-to-state stability of the control systems.

Inspired by the above theory, the inverse optimization theory is also widely applied to some practical systems. In [42], for vehicle active suspension system with unknown nonlinear dynamics, Li et al. designed an adaptive fuzzy inverse optimal control method via state observer. Long et al. [43] proposed an inverse optimal fuzzy adaptive control approach for the system of flexible spacecraft system with fault-free actuator, which is subjected to input saturation, uncertain parameter and external disturbances.

In addition, people hope the practical engineering will reach the stable in finite time, and use the less control energy when achieving the satisfactory performance indicators simultaneously. Thus, how to achieve the effective balance between control quality and control energy has become a hot research issue. In [44], for nonlinear impulsive systems, Li and Ho studied the problem of finite-time stability (FTS). In [45], Li and Yang developed the Lyapunov–Razumikhin method for finite-time stability (FTS) and finite-time contractive stability (FTCS) of time-delay systems. In [46], via the power integral control approach and backstepping control method, Yang designed a semi-global real finite time controller. Then, according to the basic idea of inverse optimization, an appropriate objective functional is constructed, and the constructed objective functional is minimized by adjusting the parameters of semi-global real finite time controller.

Consider the Lyapunov function as follows:

$$V_1 = \left[\frac{r_1}{2\nu - \tau} \right] x^{\frac{(2\nu - \tau)}{r_1}} + \frac{1}{2} \bar{\omega}_1^2 \quad (3.14)$$

where $\bar{\omega}_1 = \omega_1^* - \hat{\omega}_1$, $\hat{\omega}_1$ is the estimation of the unknown parameter ω_1^* , $\nu = \max\{r_1, p_1 r_2\}$, $j = 1, 2, \dots, n$, $p_j r_{j+1} = r_j + \tau$, $\nu = \max_{1 \leq j \leq n} \{r_j, p_j r_{j+1}\}$, $j = 1, 2, \dots, n$. r_j and p_j is the ratio of two positive odd numbers. $r_1 = 1$, τ is the design parameter.

Based on [46], for a class of interlinked nonlinear systems with powers of positive odd rational numbers, Li et al. [47] developed a series of homogeneous controllers, which are capable of guaranteeing the local finite-time stability of the closed-loop systems by using the adding one power integrator approach and backstepping technique.

Most of the existing optimization finite-time control methods are limited by complicated design and updating process, which vastly affect the ideal property of optimization finite-time control. In order to solve this issue, in [48], Lu et al. first proposed an immediate fuzzy adaptive inverse optimization approach to receive a switching-type inverse optimization controller and a one parameter learning mechanism. The inverse optimal stabilization is solvable, and there exists a matrix-valued function $P(x)$, which satisfied that $P(x) = P(x)^T > 0$, then the cost function is defined as:

$$J(u) = \lim_{t \rightarrow \infty} \int_0^t [L(x) + \hbar(|P(x)^{\frac{1}{2}} u|)] d\tau \quad (3.15)$$

where \hbar and its derivative \hbar' are K_∞ functions, $L(x)$ is positive functions, $u(x)$ is away from the origin in succession with $u(x) = 0$.

For a kind of robotic manipulator system, which contains uncertain dynamics and input saturation, the authors in [49] proposed a fixed-time trajectory tracking control approach based on RL. For the sake of guaranteeing that e_1 and e_2 convergence to diminutive neighborhood around 0 in a uniformly bounded convergence time T_s , where T_s stands alone with the original states. A noval nonsingular fixed-time fast terminal sliding mode is proposed as:

$$s = K(e_1 e_1) + sig^{v_1(e_2)} \quad (3.16)$$

where $Ke_1 = diag[k_{e11}, k_{e12}, \dots, k_{e1n}]$ is a diagonal matrix. $k_{e1i}, i = 1, 2, \dots, n$, are designed as:

$$k_{e1i} = (\alpha|e_{1i}|^{p-1/(kv_1)} + \beta|e_{1i}|^{g-1/(kv_1)})^{kv_1} \quad (3.17)$$

where p and g are positive scalars with $gk > 1$ and $1/v_1 < pk < 1$, $\alpha > 0, \beta > 0, k > 1, v_1 > 1$.

Obviously, the above developed control method can effectively solve the finite/fixed-time optimal control problems and can make the minimize the cost function. Besides, in [50], Hu et al. considered the fixed-time stability of delayed neural networks with impulsive perturbations.

4. Conclusions

It can be seen from this review that optimization control design for unknown nonlinear systems via RL and ADP has been diffusely studied in control area and has achieved fruitful results. The origin and the development of optimization algorithms have been introduced, the research results of optimization control of affine nonlinear systems have been summarized. Then, under the frame of backstepping control, the adaptive optimal control, finite-time inverse optimal control, constraint control have also been described for strict-feedback nonlinear systems. At the same time, we have summarized the applications development of adaptive optimization control methods. In addition, as a novel hot issue in this field, finite/fixed-time optimal control via backstepping and RL/ADP for nonlinear systems have attracted considerable attentions, both theory and practical applications also need to be further studied in the future.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 61822307.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. R. E. Kalman, When is a linear control system optimal, *J. Basic Eng.*, **86** (1964), 51–60. <https://doi.org/10.1115/1.3653115>
2. R. A. Freeman, P. V. Kokotovic, Inverse optimality in robust stabiliztion, *SIAM J. Control Optim.*, **34** (1998). <https://doi.org/10.1137/S0363012993258732>

3. R. Bellman, Dynamic programming, *Science*, **153** (1966), 34–37. <https://doi.org/10.1126/science.153.3731.34>
4. P. J. Werbos, *New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph.D thesis, Harvard University, 1974.
5. P. J. Werbos, Advanced forecasting methods for global crisis warning and models of intelligence, *Gen. Syst.*, **1977** (1977), 25–38. <https://doi.org/10.1086/292050>
6. P. J. Werbos, Optimization methods for brain-like intelligent control, in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, **1** (1977), 579–584. <https://doi.org/10.1109/CDC.1995.478957>
7. G. A. Rovithakis, M. A. Christodoulou, Adaptive control of unknown plants using dynamical neural networks, *IEEE Trans. Syst. Man Cybern.*, **24** (1994), 400–412. <https://doi.org/10.1109/21.278990>
8. J. J. Murray, C. J. Cox, G. G. Lendaris, R. Saeks, Adaptive dynamic programming, *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, **32** (2002), 140–153. <https://doi.org/10.1109/TSMCC.2002.801727>
9. M. Abu-Khalaf, F. L. Lewis, Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach, *Automatica*, **41** (2010), 779–791. <https://doi.org/10.1016/j.automatica.2004.11.034>
10. K. G. Vamvoudakis, F. L. Lewis, Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem, *Automatica*, **46** (2010), 878–888. <https://doi.org/10.1016/j.automatica.2010.02.018>
11. X. D. Li, D. X. Peng, J. D. Cao, Lyapunov stability for impulsive systems via event-triggered impulsive control, *IEEE Trans. Autom. Control*, **65** (2020), 4908–4913. <https://doi.org/10.1109/TAC.2020.2964558>
12. X. D. Li, S. J. Song, J. H. Wu, Exponential stability of nonlinear systems with delayed impulses and applications, *IEEE Trans. Autom. Control*, **64** (2019), 4024–4034. <https://doi.org/10.1109/TAC.2019.2905271>
13. D. Wang, D. R. Liu, H. L. Li, Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems, *IEEE Trans. Autom. Sci. Eng.*, **11** (2014), 627–632. <https://doi.org/10.1109/TASE.2013.2296206>
14. H. G. Zhang, L. L. Cui, X. Zhang, Y. H. Luo, Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method, *IEEE Trans. Neural Networks*, **22** (2011), 2226–2236. <https://doi.org/10.1109/TNN.2011.2168538>
15. D. R. Liu, D. Wang, F. Y. Wang, H. L. Li, X. Yang, Neural-network-based online HJB solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems, *IEEE Trans. Cybern.*, **44** (2014), 2834–2847. <https://doi.org/10.1109/TCYB.2014.2357896>
16. D. R. Liu, X. Yang, H. L. Li, Adaptive optimal control for a class of continuous-time affine nonlinear systems with unknown internal dynamics, *Neural Comput. Appl.*, **23** (2013), 1843–1850. <https://doi.org/10.1007/s00521-012-1249-y>

17. G. X. Wen, C. L. Philip Chen, S. Z. Sam Ge, H. L. Yang, X. G. Liu, Optimized adaptive nonlinear tracking control using actor-critic reinforcement learning strategy, *IEEE Trans. Ind. Inf.*, **15** (2019), 4969–4977. <https://doi.org/10.1109/TII.2019.2894282>
18. X. Yang, D. R. Liu, Y. Z. Huang, Neural-network-based online optimal control for uncertain non-linear continuous-time systems with control constraints, *IET Control Theory Appl.*, **7** (2013), 2037–2047. <https://doi.org/10.1049/iet-cta.2013.0472>
19. D. R. Liu, X. Yang, D. Wang, Q. L. Wei, Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints, *IEEE Trans. Cybern.*, **45** (2015), 1372–1385. <https://doi.org/10.1109/TCYB.2015.2417170>
20. X. Yang, D. R. Liu, Q. L. Wei, Online approximate optimal control for affine non-linear systems with unknown internal dynamics using adaptive dynamic programming, *IET Control Theory Appl.*, **8** (2014), 1676–1688. <https://doi.org/10.1049/iet-cta.2014.0186>
21. S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, W. E. Dixon, A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems, *Automatica*, **49** (2013), 82–92. <https://doi.org/10.1016/j.automatica.2012.09.019>
22. M. Krstic, P. V. Kokotovic, I. Kanellakopoulos, *Nonlinear and Adaptive Control Design*, John Wiley & Sons, 1995.
23. G. X. Wen, S. Z. Sam Ge, F. W. Tu, Optimized backstepping for tracking control of strict-feedback systems, *IEEE Trans. Neural Networks Learn. Syst.*, **29** (2018), 3850–3862. <https://doi.org/10.1109/TNNLS.2018.2803726>
24. S. C. Tong, K. K. Sun, S. Sui, Observer-based adaptive fuzzy decentralized optimal control design for strict-feedback nonlinear large-scale systems, *IEEE Trans. Fuzzy Syst.*, **26** (2017), 569–584. <https://doi.org/10.1109/TFUZZ.2017.2686373>
25. Y. M. Li, T. C. Wang, W. Liu, S. C. Tong, Neural network adaptive output-feedback optimal control for active suspension systems, *IEEE Trans. Syst. Man Cybern.: Syst.*, **52** (2021), 4021–4032. <https://doi.org/10.1109/TSMC.2021.3089768>
26. G. X. Wen, C. L. Philip Chen, W. N. Li, Simplified optimized control using reinforcement learning algorithm for a class of stochastic nonlinear systems, *Inf. Sci.*, **517** (2020), 230–243. <https://doi.org/10.1016/j.ins.2019.12.039>
27. X. X. Pei, Y. M. Li, S. D. Yi, Adaptive neural network optimal control of hybrid electric vehicle power battery, *J. Jilin Univ. (Eng. Technol. Edition)*, **2021** (2021). <https://doi.org/10.13229/j.cnki.jdxbgxb20211422>
28. G. X. Wen, C. L. Philip Chen, S. Z. Sam Ge, Simplified optimized backstepping control for a class of nonlinear strict-feedback systems with unknown dynamic functions, *IEEE Trans. Cybern.*, **51** (2020), 4567–4580. <https://doi.org/10.1109/TCYB.2020.3002108>
29. J. Lan, Y. J. Liu, D. X. Yu, G. X. Wen, S. C. Tong, L. Liu, Time-varying optimal formation control for second-order multiagent systems based on neural network observer and reinforcement learning, *IEEE Trans. Neural Networks Learn. Syst.*, **2022** (2022), 1–12. <https://doi.org/10.1109/TNNLS.2022.3158085>

30. W. B. Xiao, Q. Zhou, Y. Liu, H. Y. Li, R. Q. Lu, Distributed reinforcement learning containment control for multiple nonholonomic mobile robots, *IEEE Trans. Circuits Syst. I Regul. Pap.*, **69** (2021), 896–907. <https://doi.org/10.1109/TCSI.2021.3121809>
31. B. X. Jiang, Y. J. Lou, J. Q. Lu, Input-to-state stability of delayed systems with bounded-delay impulses, *Math. Modell. Control*, **2** (2022), 44–54. <https://doi.org/10.3934/mmc.2022006>
32. V. Djordjevic, L. Dubonjic, M. M. Morato, D. Prsic, V. Stojanovic, Sensor fault estimation for hydraulic servo actuator based on sliding mode observer, *Math. Modell. Control*, **2** (2022), 34–43. <https://doi.org/10.3934/mmc.2022005>
33. Y. M. Li, Y. J. Liu, S. C. Tong, Observer-based neuro-adaptive optimized control of strict-feedback nonlinear systems with state constraints, *IEEE Trans. Neural Networks Learn. Syst.*, **33** (2022), 3131–3145. <https://doi.org/10.1109/TNNLS.2021.3051030>
34. Y. M. Li, Y. L. Fan, K. W. Li, W. Liu, S. C. Tong, Adaptive optimized backstepping control-based RL algorithm for stochastic nonlinear systems with state constraints and its application, *IEEE Trans. Cybern.*, **2021** (2021), 1–14. <https://doi.org/10.1109/TCYB.2021.3069587>
35. Y. M. Li, J. X. Zhang, W. Liu, S. C. Tong, Observer-based adaptive optimized control for stochastic nonlinear systems with input and state constraints, *IEEE Trans. Neural Networks Learn. Syst.*, **2021** (2021), 1–15. <https://doi.org/10.1109/TNNLS.2021.3087796>
36. Y. Wu, X. J. Xie, Robust adaptive control for state-constrained nonlinear systems with input saturation and unknown control direction, *IEEE Trans. Syst. Man Cybern.: Syst.*, **51** (2019), 1192–1202. <https://doi.org/10.1109/TSMC.2019.2895048>
37. Y. M. Li, J. X. Zhang, S. C. Tong, Fuzzy adaptive optimized leader-following formation control for second-order stochastic multi-agent systems, *IEEE Trans. Ind. Inf.*, **18** (2021), 6026–6037. <https://doi.org/10.1109/TII.2021.3133927>
38. K. Ezal, Z. G. Pan, P. Kokotovic, Locally optimal and robust backstepping design, *IEEE Trans. Autom. Control*, **45** (2000), 260–271. <https://doi.org/10.1109/9.839948>
39. Y. M. Li, X. Min, S. C. Tong, Adaptive fuzzy inverse optimal control for uncertain strict-feedback nonlinear systems, *IEEE Trans. Fuzzy Syst.*, **28** (2019), 2363–2374. <https://doi.org/10.1109/TFUZZ.2019.2935693>
40. Y. M. Li, X. Min, S. C. Tong, Observer-based fuzzy adaptive inverse optimal output feedback control for uncertain nonlinear systems, *IEEE Trans. Fuzzy Syst.*, **29** (2020), 1484–1495. <https://doi.org/10.1109/TFUZZ.2020.2979389>
41. K. X. Lu, Z. Liu, C. L. Philip Chen, Y. N. Wang, Y. Zhang, Inverse optimal design of direct adaptive fuzzy controllers for uncertain nonlinear systems, *IEEE Trans. Fuzzy Syst.*, **30** (2022), 1669–1682. <https://doi.org/10.1109/TFUZZ.2021.3064678>
42. X. Min, Y. M. Li, S. C. Tong, Adaptive fuzzy output feedback inverse optimal control for vehicle active suspension systems, *Neurocomputing*, **403** (2020), 257–267. <https://doi.org/10.1016/j.neucom.2020.04.096>
43. H. H. Long, J. K. Zhao, J. Q. Lai, H_∞ inverse optimal adaptive fault-tolerant attitude control for flexible spacecraft with input saturation, *J. Shanghai Jiaotong Univ. (Sci.)*, **20** (2015), 513–527. [10.1007/s12204-015-1659-y](https://doi.org/10.1007/s12204-015-1659-y)

44. X. D. Li, D. W. C. Ho, J. D. Cao, Finite-time stability and settling-time estimation of nonlinear impulsive systems, *Automatica*, **99** (2019), 361–368. <https://doi.org/10.1016/j.automatica.2018.10.024>
45. X. D. Li, X. Y. Yang, S. J. Song, Lyapunov conditions for finite-time stability of time-varying time-delay systems, *Automatica*, **103** (2019), 135–140. <https://doi.org/10.1016/j.automatica.2019.01.031>
46. Y. M. Li, T. T. Yang, S. C. Tong, Adaptive neural networks finite-time optimal control for a class of nonlinear systems, *IEEE Trans. Neural Networks Learn. Syst.*, **31** (2019), 4451–4460. <https://doi.org/10.1109/TNNLS.2019.2955438>
47. Y. M. Li, T. T. Yang, L. Liu, G. Feng, S. C. Tong, Finite-time optimal control for interconnected nonlinear systems, *Int. J. Robust Nonlinear Control*, **30** (2020), 3451–3470. <https://doi.org/10.1002/rnc.4944>
48. K. X. Lu, Z. Liu, H. Y. Yu, C. L. Philip Chen, Y. Zhang, Adaptive fuzzy inverse optimal fixed-time opntrol of uncertain nonlinear systems, *IEEE Trans. Fuzzy Syst.*, **45** (2000), 260–271. <https://doi.org/10.1109/TFUZZ.2021.3132151>
49. S. J. Cao, L. Sun, J. J. Jiang, Z. Y. Zuo, Reinforcement learning-based fixed-time trajectory tracking control for uncertain robotic manipulators with input saturation, *IEEE Trans. Neural Networks Learn. Syst.*, **2021** (2021), 1–12. <https://doi.org/10.1109/TNNLS.2021.3116713>
50. J. T. Hu, G. X. Sui, X. X. Lv, X. D. Li, Fixed-time control of delayed neural networks with impulsive perturbations, *IEEE Trans. Neural Networks Learn. Syst.*, **23** (2018), 904–920. <https://doi.org/10.15388/NA.2018.6.6>



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)