SCIENCE CHINA
Information Sciences

• HIGHLIGHT •

May 2017, Vol. 60 058201:1–058201:3
doi: 10.1007/s11432-016-9022-1

Developing nonlinear adaptive optimal regulators through an improved neural learning mechanism

Ding WANG[1,2†] & Chaoxu MU[2*†]

[1] *The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing* 100190*, China;*
[2] *Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin* 300072*, China*

Received October 22, 2016; accepted November 29, 2016; published online March 17, 2017

Citation Wang D, Mu C X. Developing nonlinear adaptive optimal regulators through an improved neural learning mechanism. Sci China Inf Sci, 2017, 60(5): 058201, doi: 10.1007/s11432-016-9022-1

Optimal feedback design of dynamical systems is a significant topic in automatic control community and information science. As for nonlinear systems, optimal control design always leads to coping with the nonlinear Hamilton-Jacobi-Bellman equation. Nevertheless, it is intractable to acquire the analytic solution of the nonlinear Hamilton-Jacobi-Bellman equation for general nonlinear systems. As a result, some promising iterative methods have been established to deal with the optimal control problems in recent years. Among them, adaptive/approximate dynamic programming [1] is regarded as a typical method for designing optimal control adaptively and forward-in-time [2, 3]. In the last two decades, the methodology of adaptive/approximate dynamic programming has progressed significantly in the aspect of optimal control for complex nonlinear systems [4–7]. This considerably promotes the development of the adaptive critic control designs of complex nonlinear systems. However, the traditional adaptive critic control design always depends on the choice of an initial stabilizing control, which is considerably difficult to determine in control practices. This highlight focuses on developing nonlinear adaptive optimal regulators through an improved neural learning mechanism. In this highlight, $\mathbb{R}$ stands for the set of all real numbers. $\mathbb{R}^n$ is the Euclidean space of all n-dimensional real vectors. $\mathbb{R}^{n\times m}$ is the space of all $n \times m$ real matrices. Let Ω be a compact subset of $\mathbb{R}^n$ and $\mathscr{A}(\Omega)$ be the set of admissible control laws on Ω. Superscript "T" is considered for representing the transpose operation and $\nabla(\cdot) \triangleq \partial(\cdot)/\partial x$ is employed to denote the gradient operator.

Problem description. Consider a class of nonlinear continuous-time systems described by

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

where $x(t) \in \Omega \subset \mathbb{R}^n$ is the state vector, $u(t) \in \Omega_u \subset \mathbb{R}^m$ is the control vector, and system functions $f(\cdot)$ and $g(\cdot)$ are differentiable in the arguments satisfying $f(0) = 0$. We let the initial state at $t = 0$ be $x(0) = x_0$ and $x = 0$ be the equilibrium point of the controlled plant. The internal system function $f(x)$ is assumed Lipschitz continuous on a set Ω in $\mathbb{R}^n$ and containing the origin. Generally, the nonlinear plant (1) is assumed controllable.

For designing an infinite horizon optimal feedback control law $u(x)$, we let $Q(x) > 0$ for any

[*] Corresponding author (email: cxmu@tju.edu.cn)
[†] Equal contribution
The authors declare that they have no conflict of interest.

$x \neq 0$ and $Q(0) = 0$, let $R \in \mathbb{R}^{m\times m}$ be a positive definite matrix with appropriate dimension, use $U(x(\tau), u(\tau)) = Q(x(\tau)) + u^{\mathrm{T}}(\tau)Ru(\tau)$ to represent the utility function, and then define the infinite horizon cost function as

$$J(x(t), u) = \int_t^{\infty} U(x(\tau), u(\tau))\mathrm{d}\tau. \tag{2}$$

For simplicity, cost $J(x(t), u)$ is often written as $J(x(t))$ or $J(x)$ in the sequel. Our concern is always the cost function starting from $t = 0$, which is denoted as $J(x(0)) = J(x_0)$. For an admissible control law $u \in \mathscr{A}(\Omega)$, if the cost function (2) with respect to it is continuously differentiable, the related infinitesimal version is the nonlinear Lyapunov equation:

$$0 = U(x, u) + (\nabla J(x))^{\mathrm{T}}[f(x) + g(x)u],\ \ J(0) = 0.$$

Define the Hamiltonian of system (1) as

$$H(x,u,\nabla J(x)) = U(x, u) + (\nabla J(x))^{\mathrm{T}}[f(x) + g(x)u].$$

In light of Bellman's optimality principle, the optimal cost function $J^*(x)$, specifically defined as

$$J^*(x) = \min_{u\in\mathscr{A}(\Omega)} \int_t^{\infty} U(x(\tau), u(\tau))\mathrm{d}\tau,$$

ensures the Hamilton-Jacobi-Bellman equation

$$\min_u H(x, u, \nabla J^*(x)) = 0$$

holds. Based on optimal control theory, the optimal feedback control law is formulated as follows:

$$\begin{aligned} u^*(x) &= \arg\min_u H(x, u, \nabla J^*(x)) \\ &= -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla J^*(x). \end{aligned} \tag{3}$$

Using the optimal feedback control expression (3), the Hamilton-Jacobi-Bellman equation turns to be

$$0 = U(x, u^*) + (\nabla J^*(x))^{\mathrm{T}}[f(x) + g(x)u^*]. \tag{4}$$

Note that Eq. (4) is actually $H(x, u^*, \nabla J^*(x)) = 0$ and it is difficult to get the solution of $J^*(x)$ in theory. Hence, obtaining the optimal control law (3) for general nonlinear systems is not easy. This inspired us to devise an approximate control strategy to overcome the difficulty mentioned below.

Design method. The major contribution of this highlight lies in that it constructs a simple reinforced structure to achieve the nonlinear optimal regulation design adaptively, without requiring the initial stabilizing controller.

By incorporating a neural network architecture, the adaptive-critic-based design provides an important idea to approximate the optimal controller of general nonlinear systems [1–3, 5, 6]. During the neural network implementation, we denote l_c as the number of neurons in the hidden layer. Considering the universal approximation property, the optimal cost function $J^*(x)$ can be expressed by a neural network with a single hidden layer on a compact set Ω as $J^*(x) = \omega_c^{\mathrm{T}}\sigma_c(x) + \varepsilon_c(x)$, where $\omega_c \in \mathbb{R}^{l_c}$ is the ideal weight vector being upper bounded, $\sigma_c(x) \in \mathbb{R}^{l_c}$ is the activation function, and $\varepsilon_c(x) \in \mathbb{R}$ is the reconstruction error. As the ideal weight is unknown, a critic neural network is developed to approximate the optimal cost function as $\hat{J}^*(x) = \hat{\omega}_c^{\mathrm{T}}\sigma_c(x)$, where $\hat{\omega}_c \in \mathbb{R}^{l_c}$ denotes the estimated weight vector. Then, we derive the gradient vector as $\nabla\hat{J}^*(x) = (\nabla\sigma_c(x))^{\mathrm{T}}\hat{\omega}_c$. The approximate optimal feedback control law is

$$\hat{u}^*(x) = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)(\nabla\sigma_c(x))^{\mathrm{T}}\hat{\omega}_c. \tag{5}$$

Then, the approximate Hamiltonian is written as

$$\begin{aligned} &\hat{H}(x, \hat{u}^*(x), \nabla\hat{J}^*(x)) \\ &= U(x, \hat{u}^*(x)) + \hat{\omega}_c^{\mathrm{T}}\nabla\sigma_c(x)[f(x) + g(x)\hat{u}^*(x)]. \end{aligned} \tag{6}$$

Owing to the fact that $H(x, u^*, \nabla J^*(x)) = 0$, we acquire $e_c = \hat{H}(x, \hat{u}^*(x), \nabla\hat{J}^*(x))$. Clearly, we have

$$\frac{\partial e_c}{\partial\hat{\omega}_c} = \nabla\sigma_c(x)[f(x) + g(x)\hat{u}^*(x)] \triangleq \phi, \tag{7}$$

where $\phi \in \mathbb{R}^{l_c}$ and the set composed of elements $\phi_1, \phi_2, \ldots, \phi_{l_c}$ is linearly independent.

We train the critic network to minimize the objective function $E_c = 0.5e_c^2$. Traditionally, based on (6) and (7), we can employ the normalized steepest descent algorithm

$$\dot{\hat{\omega}}_c = -\alpha_c\frac{1}{(1+\phi^{\mathrm{T}}\phi)^2}\left(\frac{\partial E_c}{\partial\hat{\omega}_c}\right) = -\alpha_c\frac{e_c}{(1+\phi^{\mathrm{T}}\phi)^2}\phi$$

to adjust the weight vector, where $\alpha_c > 0$ represents the learning rate to be designed, and $(1+\phi^{\mathrm{T}}\phi)^2$ is implemented for normalization. Note that in this design technique, we should choose a specified weight vector to create an initial stabilizing control law and then start the training process. Therefore, in this highlight, we introduce an additional Lyapunov function $J_s(x)$ to improve the critic learning mechanism and adopt it to facilitate updating the critic weight in a novel manner.

Now, we observe the feedback control law (5) and derive a gradient descent operation as

$$\begin{aligned} &-\frac{\partial\left[(\nabla J_s(x))^{\mathrm{T}}(f(x) + g(x)\hat{u}^*(x))\right]}{\partial\hat{\omega}_c} \\ &= \frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathrm{T}}(x)\nabla J_s(x). \end{aligned}$$

The importance of this calculation is emphasized later. When applying the approximate optimal control (5) to the controlled plant, in order to exclude the case that the closed-loop system is unstable, i.e., $(\nabla J_s(x))^{\mathrm{T}}[f(x) + g(x)\hat{u}^*(x)] > 0$, we introduce an additional term to reinforce the training process by adjusting the time derivative of $J_s(x)$ in the negative gradient direction. Therefore, the improved critic learning rule of this highlight is developed by an additive structure,

$$\begin{aligned}\dot{\hat{\omega}}_c =& -\alpha_c \frac{\phi}{(1+\phi^{\mathrm{T}}\phi)^2} e_c \\ &+ \frac{1}{2}\alpha_s \nabla\sigma_c(x) g(x) R^{-1} g^{\mathrm{T}}(x) \nabla J_s(x),\end{aligned} \tag{8}$$

where $\alpha_s > 0$ is the designed constant with respect to the additional stabilizing term. This parameter affects the extent of criterion improvement and can be determined by control practitioners. With two adjustable learning rates α_c and α_s, the designers can conduct more practical control tasks in light of their engineering experience and intuition. Note that as a special case of adaptive control, for the type of adaptive critic design, the persistence of excitation assumption is still required because we intend to identify the parameter of the critic network to approximate the optimal cost function.

The learning rule given in (8) stands for an efficient improvement to the traditional criterion used in [2,6] and the updated criterion proposed in [4]. The primary property lies in that it reduces the need of an originally stabilizing control law. Consequently, the weight vector of the critic network can be initialized as zero when running the control algorithm. Additionally, it can be used to improve the learning criterion for event-based adaptive critic control and distributed control designs [8,9]. As Werbos pointed out, adaptive dynamic programming may be the only approach to be able to achieve truly brain-like intelligence [1]. The novel strategy developed in this highlight is beneficial to promote the development of adaptive critic control methods, particularly with an optimality [3,5] and robustness guarantee [6,7,9] and the construction of higher level learning and intelligent systems [1,3,10]. In future, we intend to obtain solutions to dynamic programming with a manageable amount of computation and communication as well as an inclusive guarantee of adaptivity, optimality, and robustness.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61233001, 61273140, 61304018, 61304086, 61533017, U1501251), Beijing Natural Science Foundation (Grant No. 4162065), Tianjin Natural Science Foundation (Grant No. 14JCQNJC05400), Early Career Development Award of SKLMCCS, and Research Fund of Tianjin Key Laboratory of Process Measurement and Control (Grant No. TKLPMC-201612).

References

1 Werbos P J. Approximate dynamic programming for real-time control and neural modeling. In: Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches. New York: Van Nostrand Reinhold, 1992. 15: 493–525

2 Vamvoudakis K G, Lewis F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica, 2010, 46: 878–888

3 Lewis F L, Vrabie D, Vamvoudakis K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. IEEE Control Syst Mag, 2012, 32: 76–105

4 Zhang H, Cui L, Luo Y. Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. IEEE Trans Cyber, 2013, 43: 206–216

5 Mu C, Sun C, Song A, et al. Iterative GDHP-based approximate optimal tracking control for a class of discrete-time nonlinear systems. Neurocomputing, 2016, 214: 775–784

6 Wang D, Liu D, Zhang Q, et al. Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics. IEEE Trans Syst Man Cyber Syst, 2016, 46: 1544–1555

7 Gao W, Jiang Y, Jiang Z P, et al. Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. Automatica, 2016, 72: 37–45

8 Sahoo A, Xu H, Jagannathan S. Approximate optimal control of affine nonlinear continuous-time systems using event-sampled neurodynamic programming. IEEE Trans Neural Netw Learn Syst, 2017, 28: 639–652

9 Narayanan V, Jagannathan S. Distributed adaptive optimal regulation of uncertain large-scale interconnected systems using hybrid Q-learning approach. IET Control Theory Appl, 2016, 10: 1448–1457

10 Fu Q, Gu P P, Wu J R. Iterative learning control for one-dimensional fourth order distributed parameter systems. Sci China Inf Sci, 2017, 60: 012204