Optimal Control for Discrete-Time Nonlinear Systems with Actuator Saturation Based on Generalized Policy Iteration Adaptive Dynamic Programming Algorithm

Qiao Lin[1], Qinglai Wei[1⋆], and Bo Zhao[1]

1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
{linqiao2014,qinglai.wei,zhaobo}@ia.ac.cn,

Abstract. In this study, a nonquadratic performance function is introduced to overcome the saturation nonlinearity in actuators. Then, a novel generalized policy iteration Adaptive dynamic programming algorithm is developed to deal with the optimal control problem. Two neural networks are introduced to approximate the control law and performance index function and one simulation example is given to illustrate the convergence and feasibility of the developed algorithm.

Keywords: Adaptive dynamic programming, optimal control, saturating actuators

1 Introduction

In the control field, saturation nonlinearity of the actuators is universal phenomenon. So researchers have made a big effort to solve the control problem of systems with saturating actuators [1, 2]. However, these traditional methods were proposed without considering the optimal control problem. In order to overcome this shortcoming, Lewis et al. [3] used the adaptive dynamic programming (ADP) algorithm to deal with the above problem. The ADP algorithm [4–6], which is an effective brain-like method and can solve the Hamilton-Jacobi-Bellman (HJB) equation forward-in-time, is an important method to get the optimal control policy. Value iteration [7] and policy iteration [8] algorithms are primary tools in ADP algorithms. Considering the superiority of the ADP algorithm, more and more researchers chose the ADP algorithm to deal with the optimal control problem. Zhang et al. [9] used greedy ADP algorithm to design the infinite-time optimal tracking controller. Qiao et al. [10] studied the ADP algorithm to manage the Coordinated reactive power control of a large wind farm and a STATCOM. Liu et al. [11] designed an optimal controller for unknown discrete-time nonlinear systems with control constraints by DHP. In [12] the ADP algorithm was used to solve the optimal control problem for a class of time-delay systems with actuator

⋆ Corresponding author of this paper.

saturation. However, there's no research on how to solve the constrained optimal control problem via the generalized policy iteration ADP algorithm [13, 14] .

In this paper, we use the generalized policy iteration ADP algorithm to obtain the optimal controller for the discrete-time nonlinear systems with actuator saturation. The present algorithm has i-iteration and j-iteration. When j is equal to zero, the developed algorithm can be thought as a value iteration algorithm. When j approaches the infinity, the developed algorithm can be regarded as a policy iteration algorithm. First, in order to overcome the saturation nonlinearity in actuators, the nonquadratic performance function is introduced. Then, the process of the generalized policy iteration algorithm is given. Finally, we use a simulation example to verify the effectiveness of the developed method.

2 Problem statement

We will study the following discrete-time nonlinear systems:

$$\begin{aligned} x_{k+1} &= F(x_k, u_k) \\ &= f(x_k) + g(x_k)u_k \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state vector, $u_k \in \mathbb{R}^m$ is control vector, $f(x_k) \in \mathbb{R}^n$ and $g(x_k) \in \mathbb{R}^{n\times m}$ are system functions. We denote $\Omega_u = \{u_k | u_k = [u_{1k}, u_{2k}, ..., u_{mk}]^\mathsf{T} \in \mathbb{R}^m, |u_{ik}| \leq \overline{u}_i, i = 1, 2, ..., m\}$, where $\overline{u}_i$ can be regarded as the saturating bound. Let $\overline{U} = diag[\overline{u}_1, \overline{u}_2, ..., \overline{u}_m]$.

The generalized nonquadratic performance index function is $J(x_k, \underline{u}_k) = \sum\limits_{i=k}^{\infty} \left\{x_i^\mathsf{T} Q x_i + W(u_i)\right\}$, where $\underline{u}_k = \{u_k, u_{k+1}, u_{k+2}, ...\}$, the weight matrix Q and $W(u_i) \in \mathbb{R}$ are positive definite.

Inspired by the paper [3], we can introduced $W(u_i) = 2\int_0^{u_i} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds$, where R is positive definite, $s \in \mathbb{R}^m$, $\Lambda \in \mathbb{R}^m$, $\Lambda^{-\mathsf{T}}$ denotes $(\Lambda^{-1})^\mathsf{T}$, and $\Lambda(\cdot)$ can choose $\tanh(\cdot)$.

Then we can use $J^*(x_k) = \min\limits_{\underline{u}_k} J(x_k, \underline{u}_k)$ to stand for the optimal performance index function and use u_k^* to be the optimal control law. So from discrete-time Bellman's optimality principle, we can obtain the optimal performance index function as

$$J^*(x_k) = \min_{u_k} \left\{ x_k^\mathsf{T} Q x_k + 2\int_0^{u_k} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + J^*(x_{k+1}) \right\}. \tag{2}$$

And we can use the following equation to stand for the optimal control law:

$$u_k^* = \arg\min_{u_k} \left\{ x_k^\mathsf{T} Q x_k + 2\int_0^{u_k} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + J^*(x_{k+1}) \right\}. \tag{3}$$

The goal of this paper is to obtain the optimal performance index function $J^*(x_k)$ and the optimal control law u_k^*.

3 Derivation of the generalized policy iteration ADP algorithm

From [16], we can konw that the traditional ADP algorithm just have one iteration procedure. However, the generalized policy iteration ADP algorithm has i-iteration and j-iteration. Moreover, for i-iteration, the generalized policy iteration ADP algorithm don't need to obtain the solution of the HJB equation, which lead to the convergence rate of the developed ADP algorithm can be sped up.

According to [17], if a control law can not only stabilize the system (1) but also make the performance index function finite, we can say that the control law is said to be admissible.

Next, we will get that the control law and cost function of the developed generalized policy iteration ADP algorithm are updated by iterations. First, we can obtain the initial cost function $V_0(x_k)$ as follows:

$$V_0(x_k) = x_k^{\mathsf{T}} Q x_k + 2\int_0^{v_0(x_k)} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_0(F(x_k, v_0(x_k))), \quad (4)$$

where the $v_0(x_k)$ is an initial admissible control law. Then, for $i = 1$, we can compute the control law $v_1(x_k)$ by:

$$v_1(x_k) = \arg\min_{u_k} \left\{ x_k^{\mathsf{T}} Q x_k + 2\int_0^{u_k} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_0(F(x_k, u_k)) \right\}. \quad (5)$$

Then, we will introduced the second iteration procedure. Define an arbitrary non-negative integer sequence, that is $\{L_1, L_2, L_3, \ldots\}$. L_1 is the upper boundary of j_1. When j_1 increases from 0 to L_1, we can have the iterative cost function by

$$V_{1,j_1+1}(x_k) = x_k^{\mathsf{T}} Q x_k + 2\int_0^{v_1(x_k)} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_{1,j_1}(F(x_k, v_1(x_k))), \quad (6)$$

where

$$V_{1,0}(x_k) = x_k^{\mathsf{T}} Q x_k + 2\int_0^{v_1(x_k)} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_0(F(x_k, v_1(x_k))). \quad (7)$$

Define the iterative cost function as $V_1(x_k) = V_{1,L_1}(x_k)$. For $i = 2, 3, 4, \ldots$, the control law and cost function of the generalized policy iteration ADP algorithm can be updated by:

1) i-iteration

$$v_i(x_k) = \arg\min_{u_k} \left\{ x_k^{\mathsf{T}} Q x_k + 2\int_0^{u_k} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_{i-1}(F(x_k, u_k)) \right\}, \quad (8)$$

2) j-iteration

$$V_{i,j_i+1}(x_k) = x_k^{\mathsf{T}} Q x_k + 2\int_0^{v_i(x_k)} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_{i,j_i}(F(x_k, v_i(x_k))), \quad (9)$$

where $j_i = 0, 1, 2, \ldots, L_i$,

$$V_{i,0}(x_k) = x_k^{\mathsf{T}} Q x_k + 2\int_0^{v_i(k)} \Lambda^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds + V_{i-1}(F(x_k, v_i(x_k))) \tag{10}$$

and we can get the iterative cost function by

$$V_i(x_k) = V_{i,L_i}(x_k). \tag{11}$$

From (4)–(11), the iterative cost function $V_{i,j_i}(x_k)$ and the iterative control law $v_i(x_k)$ are used to approximate $J^*(x_k)$ and u_k^*, respectively. In the following, we use one simulation example to illustrate the convergence and feasibility of the developed ADP algorithm.

4 Simulation example

The following nonlinear system is mass-spring system:

$$x(k+1) = f(x(k)) + g(x(k))u(k), \tag{12}$$

where

$$x(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix},$$
$$f(x(k)) = \begin{bmatrix} x_1(k) + 0.05x_2(k) \\ -0.0005x_1(k) - 0.0335x_1^3(k) + x_2(k) \end{bmatrix},$$
$$g(x(k)) = \begin{bmatrix} 0 \\ 0.05 \end{bmatrix},$$

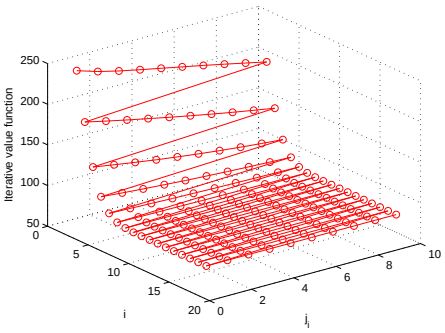
and the control constraint is set to $|u| \leq 0.6$. The cost function is defined as

$$J(x_k) = \sum_{i=k}^{\infty} \left\{ x_i^{\mathsf{T}} Q x_i + 2\int_0^{u_i} \tanh^{-\mathsf{T}}(\overline{U}^{-1}s)\overline{U}Rds \right\},$$

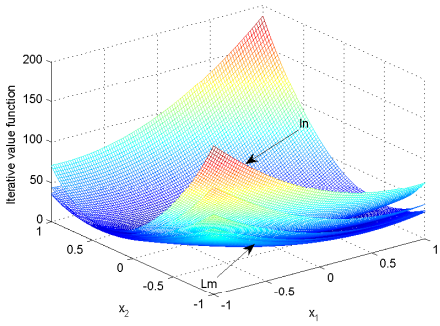
where $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $R = 0.5$, $\overline{U} = 0.6$.

We use NNs to implement the developed ADP algorithm. The structures of critic network and action network are 2–10–1, 2–10–1. The networks are trained for 17 iterations. For each iteration step, we train the networks for 4000 training steps so that the training error become minimum. The learning rate of the above two networks both are 0.01.

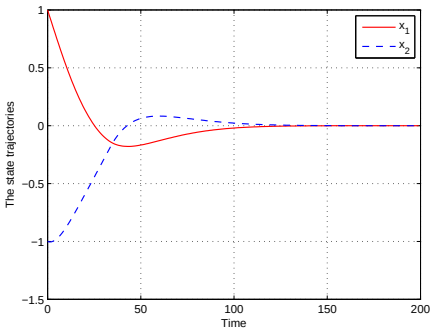
From Fig. 1(a) and 1(b), we can get the convergent process of the cost function $V_{i,j_i}(x(k))$ and the subsequence $V_i(x(k))$. Next, we apply the optimal control laws to system (12) with the initial state $x(0) = [1, -1]^{\mathsf{T}}$ for 200 time steps. The changing curves of the state x and the control u for the system (12) with actuator saturation are shown in Fig. 1(c) and 1(d). From the simulation results, we can get that the developed algorithm is effective in solving optimal control problem for discrete-time nonlinear systems with actuator saturation.



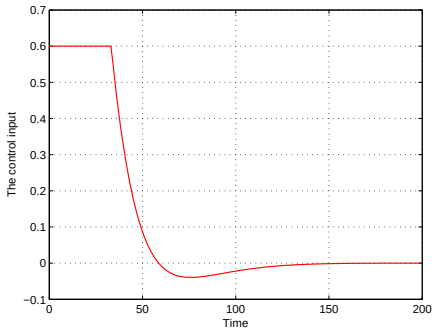
(a) V_{i,j_i} for $k = 0$



(b) V_i for the entire state space



(c) The state trajectory x



(d) The control input u

Fig. 1. The simulation trajectories

5 Conclusion

In this paper, we use a novel ADP algorithm to deal with the optimal control problem for discrete-time nonlinear systems with actuator saturation. One example demonstrates the convergence and feasibility of the generalized policy iteration ADP algorithm. Since the time-delay problem is another hot topic in the control field, it's significant to use the developed algorithm to handle the optimal control problem for time-delay systems in the future.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants 61233001, 61273140, 61304086, 61374105, 61374051, 61533017 and U1501251.

References

1. Saberi, A., Lin, Z., Teel, A.: Control of linear systems with saturating actuators. IEEE Transactions on Automatic Control, 41(3), 368–378 (1996)
2. Sussmann, H., Sontag, E., Yang, Y.: A general result on the stabilization of linear systems using bounded controls. IEEE Transactions on Automatic Control, 39(12), 2411–2425 (1994)
3. Abu-Khalaf, M., Lewis, F.: Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica, 41(5), 779–791 (2005)
4. Werbos, P.: Approximate dynamic programming for real-time control and neural modeling. in: D.A. White, D.A. Sofge (Ed.): Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches (1992)
5. Liu, D., Wang, D., Zhao, D., *et al.*: Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming. IEEE Transactions on Automation Science and Engineering, 9(3), 628–634 (2012)
6. Wei, Q., Song, R., and Yan, P.: Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP. IEEE Trans on Neural Networks and Learning Systems, 27(2), 444-458 (2016)
7. Wei, Q., Liu, D., Shi, G., *et al.*: Optimal multi-battery coordination control for home energy management systems via distributed iterative adaptive dynamic programming. IEEE Transactions on Industrial Electronics, 42(7), 4203–4214 (2015)
8. Bhasin, S., Kamalapurkar, R., Johnson, M., *et al.*: A novel actorcritic- identifier architecture for approximate optimal control of uncertain nonlinear systems. Automatica, 49(1), 82–92 (2013)
9. Zhang H., Wei Q., Luo Y.: A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm. IEEE Transaction on Systems, Man, and Cybernetics-Part B:Cybernetics, 38(4), 937-942 (2008)
10. Qiao W., Harley R.G., Venayagamoorthy G.K.: Coordinated reactive power control of a large wind farm and a STATCOM using heuristic dynamic programming. IEEE Transactions on Energy Conversion, 24(2), 493-503 (2009)

11. Liu, D., Wang, D., Yang, X.: An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs. Information Sciences, 220(1), 331–342 (2013)
12. Song, R., Zhang, H., Luo, Y., *et al.*: Optimal control laws for time-delay systems with saturating actuators based on heuristic dynamic programming. Neurocomputing, 73, 3020–3027 (2010)
13. Vrabie, D., Vamvoudakis, K., Lewis, F.: Adaptive optimal controllers based on generalized policy iteration in a continuous-time framework. In: 17th Mediterranean Conference on Control & Automation, Thessaloniki, Greece, 1402–1409 (2009)
14. Lin, Q., Wei, Q., Liu, D.: A novel optimal tracking control scheme for a class of discrete-time nonlinear systems using generalized policy iteration adaptive dynamic programming algorithm. International Journal of Systems Science, 48(3), 525–534 (2017)
15. Apostol, T.: Mathematical Analysis (2nd ed) (Addison-Wesley Press)
16. Wang, F., Zhang, H., Liu, D.: Adaptive Dynamic Programming: An Introduction. IEEE Computational Intelligence Magazine, 4(2), 39–47 (2009)
17. Liu, D., Wei, Q.: Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. IEEE Transactions on Neural Networks & Learning Systems, 25(3), 621–634 (2014)