Spiking Adaptive Dynamic Programming with Poisson Process

Qinglai Wei[1,2(✉)], Liyuan Han[1,3], and Tielin Zhang[4]

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

[2] Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China

[3] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

[4] Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract. A new iterative spiking adaptive dynamic programming (SADP) algorithm based on the Poisson process for optimal impulsive control problems is investigated with convergence discussion of the iterative process. For a fixed time interval, a 3-tuple can be computed, and then the iterative value functions and control laws can be obtained. Finally, a simulation example verifies the effectiveness of the developed algorithm.

Keywords: Spiking dynamic programming · Poission process · Nonlinear systems

1 Introduction

Impulsive behaviours exist widely in many dynamic systems, such as mathematical biology, engineering control, and information science [1–4]. An impulse is a sudden jump at an instant during the dynamic process. The research of impulsive control system has drawn a lot of attention worldwide. In [5], the stability, robust stabilization and controllability are analyzed for singular-impulsive systems via switching control. In [6], the global stability of switching Hopfield neural networks with state-dependent impulses is described with an equivalent method. It should be mentioned that previous impulsive control methods focus on linear systems [7,8]. However, for nonlinear systems, the hybrid Bellman equation is generally analytically unsolvable.

Adaptive dynamic programming (ADP), proposed by Werbos, is a method of solving optimal control problems, which combines the advantages of dynamic programming, reinforcement learning and function approximation [9–11]. ADP has two branches, value and policy iterations. However, traditional ADP methods [12–14] cannot solve impulsive control problem. To overcome this shortcoming,

Y. Tan and Y. Shi (Eds.): ICSI 2021, LNCS 12690, pp. 525–532, 2021.
https://doi.org/10.1007/978-3-030-78811-7_49

in [15], a new discrete-time impulsive ADP algorithm is proposed to obtain the optimum iteratively, while the impulsive interval is required to constrain in a fixed interval set. Furthermore, the interval set is generally difficult to determine. Until now, to the best of our knowledge, there are no discussions on optimal control problems with the spike train from real biology based on ADP algorithms, and this motivates our research.

2 Problem Statement

We consider the following discrete-time nonlinear control systems

$$x_{k+1} = F(x_k, u_k), k = 0, 1, ... \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state variable and $u_k \in \mathbb{R}^m$ is the spiking control input. Let $F(\cdot)$ be the system function.

Assumption 1. *The system (1) is controllable on a compact set $\Omega_x \subset \mathbb{R}^n$ containing the origin; the system state $x_k = 0$ is an equilibrium state of system (1) under the control $u_k = 0$, i.e., $F(0, 0) = 0$; the feedback control law satisfies $u_k(x_k) = \mu\left(\pi_k(x_k), \nu_k(x_k)\right) = 0$ for $x_k = 0$.*

Notations 1. *$\mathbb{R}_+$ and $\mathbb{Z}_+$ are the sets of all non-negative real numbers and integers, respectively. $\mathcal{T} = \{t^s\}$ is the set of spiking instants, where $t^s \in \mathbb{R}_+, s = 1, 2, ...$ τ_k is the number of spiking intants in interval $[kT, (k+1)T]$ and λ_k is the firing rate of spike train in $[0, (k+1)T]$, where $T \in \mathbb{R}_+$ and $k = 0, 1, 2, ...$ According to $\mathcal{T}$, spiking interval can be expressed as $\mathsf{t}_s = t^s - t^{s-1},\ s = 1, 2, 3...,$ where $t^0 = 0$. Let $\Gamma = \{\mathcal{F}_k\}, \mathcal{F}_k \subseteq \mathcal{F}_{k+1} \subseteq \Gamma, k = 0, 1, 2, 3...,$ where $\mathcal{F}_k$ includes the information for the computation, such as the state x_k and the number of spiking instants τ_k.*

Let $\mathcal{T}_\theta = \left\{\theta^s | \theta^s = \mathrm{round}(t^{\sum_{i=0}^{s} \tau_i}), \theta^s \in \mathbb{Z}_+, s = 0, 1, 2, ...\right\}$, be the spiking instants, where round$(\cdot)$ is a rounding function. Let $\nu_k = \nu_k(x_k) \in \mathbb{R}^m$ and $\pi_k = \pi_k(x_k) \in \mathcal{Z} = \{0, 1\}$ for $k = 0, 1, 2, ...$ When $k = \theta^s$, we have $u_k = \nu_k$ and $\pi_k = 1$. Thus, the spiking control law can be written as $u_k = \mu(\pi_k, \nu_k)$, $\mu(\cdot) : \mathcal{Z} \times \mathbb{R}^m \to \mathbb{R}^m$, where $\mu(\pi_k, \nu_k)$ can be defined as

$$u_k = \mu(\pi_k, \nu_k) = \begin{cases} 0, & \pi_k = 0 \\ \nu_k, & \pi_k = 1. \end{cases} \tag{2}$$

Let $\underline{u}_k = \{u_k, u_{k+1}, ...\}$, $\underline{\pi}_k = \{\pi_k, \pi_{k+1}, ...\}$ and $\underline{\nu}_k = \{\nu_k, \nu_{k+1}, ...\}$ $k = 0, 1, 2, ...$, respectively. The given infinite-horizon performace index function for initial state x_0 can be defined as

$$J_0(x_0, \underline{u}_0) = \mathbb{E}\left(\sum_{k=0}^{\infty} U(x_k, u_k) | \mathcal{F}_0\right) = \mathbb{E}\left(\sum_{k=0}^{\infty} U\left(x_k, \mu(\pi_k, \nu_k)\right) | \mathcal{F}_0\right) \tag{3}$$

where the utility function $U\left(x_k, \mu(\pi_k, \nu_k)\right) \geq 0$ for x_k and $\mu(\cdot)$. We desire to find an optimal spiking control law $u_k^*(x_k) = \mu\left(\pi_k^*(x_k), \nu_k^*(x_k)\right)$, such that the performace index function is minimum, i.e.,

$$J_k^*(x_k) = \min_{\underline{u}_k} J_k(x_k, \underline{u}_k), \tag{4}$$

satisfying Bellman Equation [16], which is expressed as

$$J_k^*(x_k) = \min_{u_k} \mathbb{E}\big\{U(x_k, u_k) + J_{k+1}^*(x_{k+1})|\mathcal{F}_k\big\}. \tag{5}$$

3 SADP Method Based on Poisson Process

In this section, the new iterative SADP algorithm based on Poisson process is described to obtain the optimal spiking control law for a discrete-time nonlinear system (1) with property analysis.

3.1 Transformation of the Utility Function

According to the MLE, the set $\Pi = \{\tau_k\}$ and the set $\Lambda = \{\lambda_k\}$, $k = 0, 1, 2, ...$ can be easily obtained. Let $\bar{\lambda}$ represent the average of $\{\lambda_k\}, k = 0, 1, 2, ...$ For $\mathfrak{K} = 0, 1, ...,$ Poisson process [17–19] can be expressed as

$$P\left(N(t) = \mathfrak{K}\right) = \frac{(\lambda t)^{\mathfrak{K}}}{\mathfrak{K}!} \exp(-\lambda t). \tag{6}$$

Due to the fixed time interval T, the probability of Poisson distribution in $[kT, (k+1)T], k = 0, 1, 2, ...$ can be calculated as

$$p_{\tau_k} = \frac{(\bar{\lambda} T)^{\tau_k}}{\tau_k!} \exp(-\bar{\lambda} T). \tag{7}$$

Thus, for each state $x_k \in \Omega_x$, $k = 0, 1, 2, ...$, we can get a 3-tuple $(x_k, \tau_k, p_{\tau_k})$. Also, the probability p_{τ_k} is added to $\mathcal{F}_k$ for $k = 0, 1, 2, 3...$ Thus, we can obtain a new utility function $\mathcal{U}_{\tau_k}$ expressed as

$$\mathcal{U}_{\tau_k}\left(x_k, \mu(\pi_{k+\tau_k}, \nu_{k+\tau_k})\right) = \frac{1 - p_{\tau_k}}{\tau_k} \sum_{j=0}^{\tau_k - 1} U(x_{k+j}, 0) + p_{\tau_k} U(x_{k+\tau_k}, \nu_{k+\tau_k}). \tag{8}$$

Thus, the optimal spiking value function $V_k^*(x_k)$ can be defined as

$$V_k^*(x_k) = \min_{\nu_{k+\tau_k}} \left\{\mathcal{U}_{\tau_k}\left(x_k, \nu_{k+\tau_k}\right) + \sum_{j \in \Omega_x} p\left(j|x_k, \tau_k\right) J_{k+\tau_k+1}^*(j)\right\}. \tag{9}$$

3.2 Iterative SADP Method Based on Poisson Process

Then, the SADP algorithm based on Poisson process can be derived in Algorithm 1.

Algorithm 1. SADP Algorithm based on Poisson Process

Require:

Give an initial state x_0 randomly, a computation precision ϵ and an arbitrary positive semi-definite function $\Psi(x)$.

Ensure:

1: Let the iteration index $i = 0$, and the initial iterative value function $V_0(x_k) = \Psi(x_k),\ k = 0, 1, 2....$

2: Obtain the 3-tuple $(x_k, \tau_k, p_{\tau_k}),\ k = 0, 1, 2, ...$

3: Iterative spiking control law $\nu_i(x_k)$ can be computed as

$$\nu_i(x_k) = \arg \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j \in \Omega_x} p\,(j|x_k, \tau_k)\, V_i(j) \right\}. \tag{10}$$

4: Iterative spiking value function $V_{i+1}(x_k)$ can be computed as

$$V_{i+1}(x_k) = \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j \in \Omega_x} p\,(j|x_k, \tau_k)\, V_i(j) \right\}. \tag{11}$$

5: If $|V_{i+1}(x_k) - V_i(x_k)| \leq \epsilon, \forall x_k \in \Omega_x$, then the optimal performance index function and optimal spiking control law can be obtained. Goto step 6. Otherwise, let $i = i + 1$, and goto step 2.

6: end.

3.3 Property Analysis of the SADP Algorithm Based on Poisson Process

In this section, the property analysis of the SADP algorithm based on Poisson process will be estabilished.

Theorem 1. *Let $J_k^*(x_k)$ and $V_k^*(x_k),\ k = 0, 1, 2, ...,$ be the optimal performance index function and optimal spiking value function which satisfy (4) and (9), respectively. Then, for each 3-tuple $(x_k, \tau_k, p_{\tau_k}),\ k = 0, 1, 2, ...,$ we have*

$$J_k^*(x_k) = V_k^*(x_k). \tag{12}$$

Proof. Based on the 3-tuple $(x_k, \tau_k, p_{\tau_k})$ obtained by the real sequence of spike train, for any state $x_k \in \Omega_x$, we can derive that $k + \tau_k$ is a spiking instant, i.e., $\pi_{k+\tau_k} = 1$, with the Poisson probability p_{τ_k}. Thus, according to (4), we can derive the following Bellman equation (13)

$$\begin{aligned} J_k^*(x_k) &= \min_{\underline{u}_k} \left\{ \mathbb{E}\left(\sum_{j=0}^{\infty} U(x_{k+j}, u_{k+j}) | \mathcal{F}_k \right) \right\} \\ &= \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j \in \Omega_x} p\left(j|x_k, \tau_k \right) J^*_{k+\tau_k+1}(j) \right\} \\ &= V_k^*(x_k), \end{aligned} \tag{13}$$

where $p(j|x_k, \tau_k)$ can be expressed as

$$p\left(j|x_k, \tau_k\right) = \begin{cases} \dfrac{1 - p_{\tau_k} p_{\tau_{k+\tau_k}}}{N-1}, & j \in \Omega_x, j \neq x_{k+\tau_k} \\ p_{\tau_k} p_{\tau_{k+\tau_k}}, & j = x_{k+\tau_k}. \end{cases} \tag{14}$$

and N represents the number of the states in Ω_x. The Eq. (14) shows that, for state $x_{k+\tau_k}$, the probability is the product of p_{τ_k} and $p_{\tau_{k+\tau_k}}$, while the probability is the same for other states, i.e., $(1 - p_{\tau_k} p_{\tau_{k+\tau_k}})/(N-1)$.

The proof is complete.

According to Theorem 1, for each 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, ...$, the Bellman equation (5) can be expressed as

$$\begin{aligned} J_k^*(x_k) = \frac{1 - p_{\tau_k}}{\tau_k} \sum_{j=0}^{\tau_k - 1} U(x_{k+j}, 0) + \min_{\nu_{k+\tau_k}} \Bigg\{ & p_{\tau_k} U(x_{k+\tau_k}, u_{k+\tau_k}) \\ & + \sum_{j \in \Omega_x} p\Big(j|x_k, \tau_k\Big) J_{k+\tau_k+1}^*(j) \Bigg\}. \end{aligned} \tag{15}$$

The Bellman equation (15) can be called "3-tuple Bellman equation".

Lemma 1. *For $i = 0, 1, 2, ...$, and any $(x_k, \tau_k, p_{\tau_k})$ $k = 0, 1, 2, ...$, let $V_{i+1}(x_k)$ and $\nu_i(x_k)$ be the iterative value function and the iterative control law updated, respectively. According to (1)–(1) in Algorithm 1. Then, the $V_i(x_k)$ converges to the optimal performance index function $J_k^*(x_k)$ as $i \to \infty$, which is defined as Eq. (15), that is*

$$\lim_{i \to \infty} V_i(x_k) = J_k^*(x_k). \tag{16}$$

The conclusion is easily derived and the proof is omitted here.

4 Simulation

We consider the torsional pendulum system to evaluate the performance of our developed algorithm. The dynamic system is expressed as

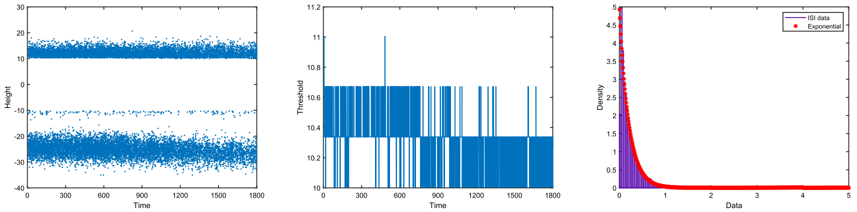
$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} x_{1k} + \Delta t x_{2k} \\ -\dfrac{\Delta t M g l}{J} \sin(x_{1k}) + \left(1 - \dfrac{\Delta t f_d}{J}\right) x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ \Delta t \end{bmatrix} u_k \tag{17}$$

where $J = 4/3\,\mathrm{ml}^2$, $M = 1/3\,\mathrm{kg}$, $g = 9.8\,\mathrm{m/s}^2$, $l = 3/2\,\mathrm{m}$ and $f_d = 0.2$ are the parameters of this system.

The utility function is chosen as $U(x_k, u_k) = x_k^{\mathsf{T}} P x_k + u_k^{\mathsf{T}} R u_k$, where $Q = I_1$ and $R = I_2$, denoting the identity matrices with suitable dimensions. Choose the initial value function with the form $\Psi(x_k) = x_k^{\mathsf{T}} P x_k$, where $P = [10\ 1; 1\ 2]$.

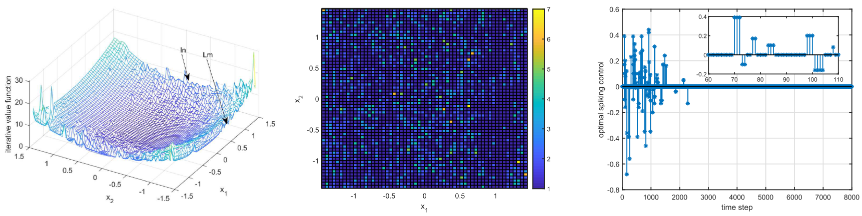
In this example, we use the data set shared by Potter Lab [20, 21] to establish the 3-tuple. The fixed time is 0.3 s. The spike train is shown in Fig. 1(a)–(c).

We implement Algorithm 1 with $\hat{\Omega}_x$ for 20 iterations in order to urge the iterative value function to be convergent, as shown in Fig. 2(a). where "In" and "Lm" represent first iteration and last iteration, respectively. We can also see that the iterative value function is not smooth in the discretized state space due to the effect of spike train. Thus, the optimal spiking instants may vary with the states. The distribution of the optimal spiking intervals in the discretized state space $\hat{\Omega}_x$ can be seen in Fig. 2(b), existing seven kinds of intervals, from one to seven. In this example, we choose an initial state $x_0^1 = [1.2 \ \ -0.8]^{\mathsf{T}}$ and we get the corresponding optimal spiking control as shown in Fig. 2(c), respectively.



(a) (b) (c)

Fig. 1. The spike train. (a) Height-time. (b) Threshold-time. (c) Interspike interval.



(a) (b) (c)

Fig. 2. (a) Convergence plots of the iterative value functions. (b) The distribution of the optimal spiking intervals. (c) Optimal spiking control with initial states x_0^1.

5 Conclusion

A new iterative SADP algorithm based on Poisson process is presented to solve optimal control problems for nonlinear systems. By using the model of Poisson process and the method of MLE, we get the 3-tuple. The property analysis is developed to guarantee that the value functions converge iteratively to optimal performance index function. Finally, a simulation example is given to verify the effectiveness of the developed algorithm.

References

1. Wang, X., Yu, J., Huang, Y., Wang, H., Miao, Z.: Adaptive dynamic programming for linear impulse systems. J. Zhejiang Univ. Sci. C **15**(1), 43–50 (2014). https://doi.org/10.1631/jzus.C1300145
2. Li, W., Huang, L., Guo, Z., Ji, J.: Global dynamic behavior of a plant disease model with ratio dependent impulsive control strategy. Math. Comput. Simul. **177**, 120–139 (2020)
3. Haddad, W.M., Chellaboina, V., Kablar, N.A.: Non-linear impulsive dynamical systems. Part II: stability of feedback interconnections and optimality. Int. J. Control **74**, 1659–1677 (2001)
4. Chen, W.-H., Luo, S., Zheng, W.X.: Generating globally stable periodic solutions of delayed neural networks with periodic coefficients via impulsive control. IEEE Trans. Cybern. **47**, 1590–1603 (2016)
5. Yao, J., Guan, Z.-H., Chen, G., et al.: Stability, robust stabilization and H? Control of singular-impulsive systems via switching control. Syst. Control Lett. **55**, 879–886 (2006)
6. Zhang, X., Li, C., Huang, T.: Hybrid impulsive and switching Hopfield neural networks with state-dependent impulses. Neural Netw. **93**, 176–184 (2017)
7. Li, X., Song, S.: Stabilization of delay systems: delay-dependent impulsive control. IEEE Trans. Autom. Control **62**, 406–411 (2016)
8. Zhang, Q., Qiao, L., Zhu, B., et al.: Dissipativity analysis and synthesis for a class of T-S fuzzy descriptor systems. IEEE Trans. Syst. Man Cybern. Syst. **47**, 1774–1784 (2016)
9. Woźniak, S., Pantazi, A., Bohnstingl, T., et al.: Deep learning incorporating biologically inspired neural dynamics and in-memory computing. Nat. Mach. Intell. **2**, 325–336 (2020)
10. Kiumarsi, B., Vamvoudakis, K.G., Modares, H., Lewis, F.L.: Optimal and autonomous control using reinforcement learning: a survey. IEEE Trans. Neural Netw. Learn. Syst. **29**, 2042–2062 (2017)
11. Jiang, Y., Jiang, Z.-P.: Robust Adaptive Dynamic Programming. Wiley, Hoboken (2017)
12. Wen, Y., Si, J., Gao, X., et al.: A new powered lower limb prosthesis control framework based on adaptive dynamic programming. IEEE Trans. Neural Netw. Learn. Syst. **28**, 2215–2220 (2016)
13. Liu, D., Wei, Q., Wang, D., Yang, X., Li, H.: Adaptive Dynamic Programming with Applications in Optimal Control. AIC. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50815-3
14. Liu, D., Xu, Y., Wei, Q., et al.: Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming. IEEE/CAA J. Automatica Sinica **5**, 36–46 (2017)
15. Wei, Q., Song, R., Liao, Z., et al.: Discrete-time impulsive adaptive dynamic programming. IEEE Trans. Cybern. **50**, 4293–4306 (2019)
16. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, Hoboken (2014)
17. Kordovan, M., Rotter, S.: Spike train cumulants for linear-nonlinear Poisson cascade models. arXiv preprint arXiv:2001.05057 (2020)
18. Bux, C.E.R., Pillow, J.W.: Poisson balanced spiking networks. bioRxiv **836601** (2019)

19. Gerhard, F., Deger, M., Truccolo, W.: On the stability and dynamics of stochastic spiking neuron models: nonlinear Hawkes process and point process GLMs. PLoS Comput. Biol. **13**, e1005390 (2017)
20. Newman, J.P., Fong, M.-f., Millard, D.C., et al.: Optogenetic feedback control of neural activity. Elife **4**, e07192 (2015)
21. Fong, M.-F., Newman, J.P., Potter, S.M., et al.: Upward synaptic scaling is dependent on neurotransmission rather than spiking. Nat. Commun. **6**, 1–11 (2015)