Spiking Adaptive Dynamic Programming Based on Poisson Process for Discrete-Time Nonlinear Systems

Qinglai Wei, *Senior Member, IEEE*, Liyuan Han, and Tielin Zhang

***Abstract*— In this article, a new iterative spiking adaptive dynamic programming (SADP) method based on the Poisson process is developed to solve optimal impulsive control problems. For a fixed time interval, combining the Poisson process and the maximum likelihood estimation (MLE), the three-tuple of state, spiking interval, and probability of Poisson distribution can be computed, and then, the iterative value functions and iterative control laws can be obtained. A property analysis method is developed to show that the value functions converge to optimal performance index function as the iterative index increases from zero to infinity. Finally, two simulation examples are given to verify the effectiveness of the developed algorithm.**

***Index Terms*— Maximum likelihood estimation (MLE), nonlinear systems, optimal control, Poisson process, spike train, spiking adaptive dynamic programming (SADP).**

I. Introduction

IMPULSIVE behaviors exist widely in many dynamic systems, such as mathematical biology, engineering control, and information science [1]–[5]. An impulse is a sudden jump at an instant during the dynamic process, usually as one of a series. Although the moment of the jump is extremely short, it makes a significant influence on the performance of dynamic systems. Therefore, the research of impulsive control system has drawn a lot of attention worldwide. In [6], the stability, robust stabilization, and controllability are analyzed for singular-impulsive systems via switching control. In [7], the global stability of switching Hopfield neural networks with state-dependent impulses is described with an equivalent method. The strategy of hybrid impulsive and switching control is proposed to establish some new criteria for global exponential stability of synchronization of nonlinear systems based on switched Lyapunov functions [8]–[10]. It should be mentioned that previous impulsive control methods focus on linear systems [11], [12]. However, for nonlinear systems [13]–[16], the hybrid Bellman equation is generally analytically unsolvable. Thus, nonanalytical and approximate solutions are required to obtain the optimal impulsive control law.

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 [17]–[21]. In [22] and [23], the greedy ADP is extended to the control problem with ϵ-error bound and successfully solved the finite-time optimal control problem in the situation where the terminal time is not fixed. In [24], a novel data-based ADP algorithm was developed to solve the optimal control law for the unknown discrete-time system with time delays. In [25], a novel steering controller for autonomous vehicles is proposed, which is implemented by a policy iteration ADP to obtain the optimal control law. Value and policy iterations, which are two branches of ADP, have attracted much attention due to the convenience for property analysis. Generally, value iteration [26] starts with a zero or positive semidefinite value function to obtain the optimal performance index function as the iterative index increases from zero to infinity. Policy iteration [17], [27] begins with an admissible control law to ensure that the iterative value function converges monotonically and nonincreasingly to the optimum.

However, traditional ADP methods [28]–[34] cannot solve the impulsive control problem since only the optimal control input is considered rather than the impulse interval and amplitude in traditional ADP methods. To overcome this shortcoming, in [35], an ADP scheme for optimal control problems within a fixed terminal time is developed by turning parameters of a function approximator. In [36], an event-driven method based on dual heuristic dynamic programming is presented to study the optimal regulation, and the appropriate event-triggering condition is given. In [37], a data-driven iterative adaptive critic strategy is constructed to address the nonlinear optimal feedback control and is applied to a typical wastewater treatment plant. In [38], a discrete-time impulsive ADP algorithm was proposed to obtain the optimum iteratively, while the impulsive interval is required to constrain

Manuscript received October 31, 2020; revised March 24, 2021; accepted May 22, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1702300 and in part by the National Natural Science Foundation of China under Grant 62073321 and Grant 61673054. *(Corresponding author: Qinglai Wei.)*

Qinglai Wei and Liyuan Han are with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China (e-mail: qinglai.wei@ia.ac.cn; hanliyuan2019@ia.ac.cn).

Tielin Zhang is with the Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: tielin.zhang@ia.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3085781.

Digital Object Identifier 10.1109/TNNLS.2021.3085781

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.

In this article, for the first time, a new iterative ADP method with spike train, which is called spiking ADP (SADP) method, is developed to solve optimal impulsive control problems for discrete-time nonlinear systems. The main contributions of the SADP method are summarized as follows.

1) Breaking the mold of traditional fixed interval set, the spike train from real biology is applied to solving optimal impulsive control problems. Moreover, the spike train has been modeled as a Poisson process, the parameters of which are estimated by maximum likelihood estimation (MLE) in a novel method of cumulative calculation.
2) Given a fixed time interval, a three-tuple consisting of state, spiking interval, and probability can be obtained, which aims to achieve the optimal spiking control law rather than use state only in traditional impulsive control methods.
3) A new convergent approach by using conditional expectation is developed to prove that value functions can reach the optimum iteratively.

The remainder of this article is organized as follows. Section II represents the problem statement, including discrete-time nonlinear spiking control systems and infinite-horizon performance index function. Section III introduces the Poisson process model for spike train and the corresponding parameter estimation by using the theory of MLE first. Then, it states the SADP algorithm in detail and gives a new method of property analysis. In Section IV, two simulation examples are given to verify the effectiveness of the present algorithm. Finally, this article ends with some conclusions in Section V.

II. Problem Statement

In this article, we consider the following discrete-time nonlinear spiking control systems:

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

where $x_k \in \mathbb{R}^n$ is the state variable and $u_k \in \mathbb{R}^m$ is the spiking control input (spiking control in brief). Let $F(\cdot)$ be the system function. A series of spike trains include the time of the spike, the electrode channel on which the spike occurred, the amplitude of the detected spike, and the spike detection threshold in force at the time of detection [39]–[42]. Traversing the data due to the index of it, a single channel of spike train can be extracted. If the amplitude is greater than the threshold value, then the detected spike is in the depolarization phase, which means that the time of this detected spike is a spiking instant. Otherwise, it is in the repolarization phase or the hyperpolarization phase, i.e., absolute refractory period or relative refractory period.

For convenience of analysis, the results of this article are based on the following assumption.

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 $u_k(x_k) = \mu(\pi_k(x_k), \nu_k(x_k)) = 0$ for $x_k = 0$.

We define that $\mathbb{R}_+$ and $\mathbb{Z}_+$ are the sets of all nonnegative real numbers and integers, respectively. Let $\mathcal{T} = \{t^s\}$ be the set of spiking instants, where $t^s \in \mathbb{R}_+, s = 1, 2, \ldots$ For $T \in \mathbb{R}_+$ and $k = 0, 1, 2, \ldots$, we let τ_k be the number of spiking instants in the admissible spiking interval $[kT, (k+1)T]$ and we let λ_k be the firing rate of spike train in $[0, (k+1)T]$. According to $\mathcal{T}$, spiking interval can be expressed as $\mathfrak{t}_s = t^s - t^{s-1}$, $s = 1, 2, 3, \ldots$, where $t^0 = 0$. Let $\Gamma = \{\mathcal{F}_k\}, \mathcal{F}_k \subseteq \mathcal{F}_{k+1} \subseteq \Gamma, k = 0, 1, 2, 3, \ldots$, 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 $\mathcal{T}_\theta = \{\theta^s\}, \theta^s \in \mathbb{Z}_+, s = 0, 1, 2, \ldots$, be the spiking instants, where θ^s can be defined as

$$\theta^s = \text{round}\left(t^{\sum_{i=0}^{s} \tau_i}\right) \tag{2}$$

and round$(\cdot)$ is a rounding function. For $k = 0, 1, \ldots$, the spiking control u_k is expressed as

$$u_k = \begin{cases} 0, & k \neq \theta^s \\ \nu_k, & k = \theta^s \end{cases} \tag{3}$$

where $\nu_k = \nu_k(x_k) \in \mathbb{R}^m$ denotes the spiking control law at the spiking instant $k = \theta^s$. Employing the spiking control law $\pi_k = \pi_k(x_k)$, where $\pi_k \in \mathcal{Z}$ and $\mathcal{Z} = \{0, 1\}$ for $k = 0, 1, \ldots$, the controlling spiking instant can be expressed as

$$\begin{cases} \pi_k = 1, & k = \theta^s \\ \pi_k = 0, & k \neq \theta^s. \end{cases} \tag{4}$$

For $k = 0, 1, \ldots$, according to (3) and (4), the spiking control law can be rewritten 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

$$u_k = \mu(\pi_k, \nu_k) = \begin{cases} 0, & \pi_k = 0 \\ \nu_k, & \pi_k = 1. \end{cases} \tag{5}$$

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

$$\begin{aligned} J_0(x_0, \underline{u}_0) &= \mathbb{E}\left(\sum_{k=0}^{\infty} U(x_k, u_k) | \mathcal{F}_0\right) \\ &= \mathbb{E}\left(\sum_{k=0}^{\infty} U(x_k, \mu(\pi_k, \nu_k)) | \mathcal{F}_0\right) \end{aligned} \tag{6}$$

where the utility function $U(x_k, \mu(\pi_k, \nu_k))$ is positive definite for x_k and $\mu(\cdot)$.

Remark 1: Let $N(t)$ be a counting process, representing the sum of spiking instants in $[0, t]$. Poisson process, as one of the famous counting processes, has been used to model the spike train. τ_k is computed as the number of spiking instants in $[kT, (k+1)T]$. Furthermore, $\tau_k = N((k+1)T) - N(kT) = N(T)$, which is a random variable. Besides, τ_0 is included in $\mathcal{F}_0$, which is needed for

computing the performance index function. Therefore, (6) in Section II is in the form of conditional expectation.

We would like to find an optimal spiking control law $u_k^*(x_k) = \mu(\pi_k^*(x_k), \nu_k^*(x_k))$ such that the performance index function is minimized from any initial state. For each three-tuple $(x_k, \tau_k, p_{\tau_k})$, the optimal performance index function is denoted as

$$J_k^*(x_k) = \min_{\underline{u}_k} J_k(x_k, \underline{u}_k). \tag{7}$$

Obviously, (7) satisfies the Bellman equation [43], which is expressed as

$$J_k^*(x_k) = \min_{u_k} \mathbb{E}\{U(x_k, u_k) + J_{k+1}^*(x_{k+1})|\mathcal{F}_k\}. \tag{8}$$

Thus, the optimal control law can be expressed as

$$u_k^*(x_k) = \arg\min_{u_k} \mathbb{E}\{U(x_k, u_k) + J_{k+1}^*(x_{k+1})|\mathcal{F}_k\}. \tag{9}$$

According to (5), there exist two different types of control variables in the spiking control law u_k, which cause all of the traditional iterative ADP algorithms to fail to obtain the optimal spiking control law in this article. Thus, a new SADP algorithm is developed to address this difficulty.

III. SADP Method Based on Poisson Process

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

A. Derivation of Three-Tuple Under Poisson Process

Define the exponential probability density function as

$$f(\mathrm{t}, \lambda) = \lambda \exp(-\lambda \mathrm{t}) \quad \forall \mathrm{t} > 0. \tag{10}$$

For $k = 0, 1, 2, \ldots$, letting $\hbar_k = \sum_{j=0}^{k} \tau_j$ be the sum of intervals in $[0, (k+1)T]$, the likelihood function L can be computed as

$$\begin{aligned} L(\mathrm{t}, \lambda_k) &= \prod_{i=1}^{\hbar_k} f(\mathrm{t}_i, \lambda_k) \\ &= (\lambda_k)^{\hbar_k} \exp\left(-\lambda_k \sum_{i=1}^{\hbar_k} \mathrm{t}_i\right). \end{aligned} \tag{11}$$

Taking the partial differential L with respect to λ_k, we have

$$\begin{aligned} \frac{\partial L}{\partial \lambda_k} &= \hbar_k (\lambda_k)^{\hbar_k - 1} \exp\left(-\lambda_k \sum_{i=1}^{\hbar_k} \mathrm{t}_i\right) \\ &\quad - (\lambda_k)^{\hbar_k} \sum_{i=1}^{\hbar_k} \mathrm{t}_i \exp\left(-\lambda_k \sum_{i=1}^{\hbar_k} \mathrm{t}_i\right). \end{aligned} \tag{12}$$

Letting $(\partial L / \partial \lambda_k) = 0$, we can get

$$\lambda_k = \frac{\hbar_k}{\sum_{i=1}^{\hbar_k} \mathrm{t}_i}. \tag{13}$$

Thus, we can obtain the set of spiking intervals $\Pi = \{\tau_k\}$ and the set of firing rates $\Lambda = \{\lambda_k\}$, $k = 0, 1, 2, \ldots$. Let $\bar{\lambda}$ represent the average of $\{\lambda_k\}$, $k = 0, 1, 2, \ldots$. For $\Re = 0, 1, \ldots$, the Poisson process [44]–[46] can be expressed as

$$P(N(t) = \Re) = \frac{(\lambda t)^{\Re}}{\Re!} \exp(-\lambda t). \tag{14}$$

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

$$p_{\tau_k} = \frac{(\bar{\lambda} T)^{\tau_k}}{\tau_k!} \exp(-\bar{\lambda} T). \tag{15}$$

Thus, for each state $x_k \in \Omega_x$, $k = 0, 1, 2, \ldots$, 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 $k = 0, 1, 2, 3, \ldots$

Remark 2: It is a well-known fact that the information transmission between neurons in the human brain mainly depends on the spike train. As we have learned from [39]–[42], it is natural that the point process represented by the Poisson process becomes the dominant model for the spike train. Also, the work of [42] and [46] can also be the basis for this model. Besides that, we have tried to model the spike train with other stochastic processes, such as the Gamma process, but unfortunately, it is difficult to determine the parameters of the Gamma process. On the contrary, the firing rate of the spike train is consistent with the parameter λ of the Poisson process. Also, the Poisson process is a special Gamma process.

B. Transformation of the Utility Function

In this section, the transformation of the utility function will be introduced, which is necessary to establish our iterative SADP algorithm based on the Poisson process. According to the 3-tuples $(x_k, \tau_k, p_{\tau_k})$ and $\tau_k \in \Pi$, $k = 0, 1, 2, \ldots$, we can obtain

$$\begin{aligned} x_{k+1} &= F(x_k, \mu(0, 0)) \\ &\vdots \\ x_{k+\tau_k} &= F(x_{k+\tau_k - 1}, \mu(0, 0)) \\ x_{k+\tau_k+1} &= F(x_{k+\tau_k}, \mu(\pi_{k+\tau_k}, \nu_{k+\tau_k})) \\ &= F(x_{k+\tau_k}, \nu_{k+\tau_k}). \end{aligned} \tag{16}$$

From (16), for state $x_k \in \Omega_x$, the next spiking instant is $k + \tau_k$. It means that system (1) is zero input for time instants $k, k+1, k+2, \ldots, k+\tau_k - 1$, and the spiking control law is $u_{k+\tau_k} = \mu(\pi_{k+\tau_k}, \nu_{k+\tau_k})$. Furthermore, it can be derived that, for each three-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, there exists a new utility function $\mathcal{U}_{\tau_k}$ (called spiking utility function) such that

$$\begin{aligned} &\mathcal{U}_{\tau_k}(x_k, \mu(\pi_{k+\tau_k}, \nu_{k+\tau_k})) \\ &\quad = \mathbb{E}\left(\sum_{j=0}^{\tau_k} U(x_{k+j}, u_{k+j})|\mathcal{F}_k\right) \\ &\quad = \frac{1 - p_{\tau_k}}{\tau_k} \sum_{j=0}^{\tau_k - 1} U(x_{k+j}, 0) \\ &\qquad + p_{\tau_k} U(x_{k+\tau_k}, \mu(\pi_{k+\tau_k}, \nu_{k+\tau_k})) \\ &\quad = \frac{1 - p_{\tau_k}}{\tau_k} \sum_{j=0}^{\tau_k - 1} U(x_{k+j}, 0) + p_{\tau_k} U(x_{k+\tau_k}, \nu_{k+\tau_k}) \end{aligned} \tag{17}$$

where we define $\sum_j^i(\cdot) = 0$ for $j > i$.

According to the 3-tuples $(x_k, \tau_k, p_{\tau_k})$ and (17), we can define the optimal spiking value function $V_k^*(x_k)$ as

$$V_k^*(x_k) = \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k, \tau_k) J_{k+\tau_k+1}^*(j) \right\}. \tag{18}$$

C. Iterative SADP Method Based on Poisson Process

In this section, the new iterative SADP method is derived based on the Poisson process. In the previous impulsive ADP method [38], the impulsive instant is strictly constrained within an interval where upper and lower boundaries of the interval are determined. In the new iterative SADP method, however, the spike train is from the real nerve impulse dataset, which is modeled by the Poisson process, and the interspike interval (ISI) obeys the exponential distribution of the same firing rate λ.

Since the 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, is a basis of the SADP method, it is necessary to establish the Algorithm 1 to describe it specifically.

Algorithm 1 Obtain the 3-Tuples $(x_k, \tau_k, p_{\tau_k})$

1: Get the dataset of spike trains.
2: Traversing the data due to the index of it, a single channel of spike train can be extracted.
3: Give a fixed time interval T.
4: Calculate the number of spiking instants in $[kT, (k+1)T]$, $k = 0, 1, 2, \ldots$, i.e., interspike interval τ_k, and the firing rate in $[0, (k+1)T]$, $k = 0, 1, 2, \ldots$, i.e., λ_k.
5: Get the average of the firing rate $\bar{\lambda} = \frac{1}{n}\sum_{k=1}^{n} \lambda_k$.
6: According to (15), get the probability p_{τ_k} in $[kT, (k+1)T]$, $k = 0, 1, 2, \ldots$
7: For $x_k \in \Omega_x$, $k = 0, 1, 2, \ldots$, return the 3-tuple $(x_k, \tau_k, p_{\tau_k})$.

According to the three-tuple, for any state $x_k \in \Omega_x$, $k = 0, 1, 2, \ldots$, we know the corresponding spiking interval and probability of Poisson distribution, which can be used to obtain the transformation of the utility function and implemented in the SADP method. For $i = 0, 1, \ldots$, we define $V_i(x_k)$ and $\nu_i(x_k)$ as the iterative spiking value function and the iterative control law, respectively. Then, the SADP algorithm based on the Poisson process can be derived in Algorithm 2.

D. Property Analysis of the SADP Algorithm Based on Poisson Process

In this section, the property analysis of the SADP algorithm based on the Poisson process will be established. It will be proved that $J_k^*(x_k)$ defined in (22) is the limit of $V_i(x_k)$ as $i \to \infty$. Before the convergence analysis, the following theorem is necessary.

Theorem 1: Let $J_k^*(x_k)$ and $V_k^*(x_k)$, $k = 0, 1, 2, \ldots$, be the optimal performance index function and optimal spiking value function that satisfy (7) and (18), respectively. Then, for every 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, we have

$$J_k^*(x_k) = V_k^*(x_k). \tag{21}$$

Proof: Based on the 3-tuples $(x_k, \tau_k, p_{\tau_k})$ 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 (7), we can derive the Bellman equation

$$\begin{aligned}
&J_k^*(x_k)\\
&= \min_{\underline{u}_k} \left\{ \mathbb{E}\left(\sum_{j=0}^{\infty} U(x_{k+j}, u_{k+j}) | \mathcal{F}_k \right) \right\}\\
&= \min_{\underline{u}_k} \left\{ \mathbb{E}\left(\sum_{j=0}^{\tau_k} U(x_{k+j}, u_{k+j}) + \sum_{j=\tau_k+1}^{\infty} U(x_{k+j}, u_{k+j}) | \mathcal{F}_k \right) \right\}\\
&= \min_{\underline{\pi}_k, \underline{\nu}_k} \left\{ \mathbb{E}\left(\sum_{j=0}^{\tau_k} U(x_{k+j}, u_{k+j}) | \mathcal{F}_k \right) \right.\\
&\qquad \left. + \mathbb{E}\left(\sum_{j=\tau_k+1}^{\infty} U(x_{k+j}, u_{k+j}) | \mathcal{F}_{k+\tau_k} | \mathcal{F}_k \right) \right\}
\end{aligned}$$

Algorithm 2 SADP Algorithm Based on the Poisson Process

Require:
Give an initial state x_0 randomly.
Give a computation precision ϵ.
Give an arbitrary positive semi-definite function $\Psi(x)$.
Ensure:
1: Let the iteration index $i = 0$, and the initial iterative value function $V_0(x_k) = \Psi(x_k)$, $k = 0, 1, 2 \ldots$.
2: Obtain the 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$ by Algorithm 1.
3: Iterative spiking control law $\nu_i(x_k)$ can be computed as

$$\nu_i(x_k) = \arg\min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k, \tau_k) V_i(j) \right\}. \tag{19}$$

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

$$V_{i+1}(x_k) = \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k, \tau_k) V_i(j) \right\}. \tag{20}$$

5: If $|V_{i+1}(x_k) - V_i(x_k)| \le \epsilon, \forall x_k \in \Omega_x$, 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.

$$
\begin{aligned}
&= \min_{v_{k+\tau_k}} \left\{ \frac{1-p_{\tau_k}}{\tau_k} \sum_{j=0}^{\tau_k-1} U(x_{k+j},0) + p_{\tau_k} U(x_{k+\tau_k}, u_{k+\tau_k}) \right. \\
&\qquad \left. + \mathbb{E}\left(\min_{\underline{u}_{k+\tau_k+1}} \left\{ \mathbb{E}\left(\sum_{j=\tau_k+1}^{\infty} U(x_{k+j}, u_{k+j}) | \mathcal{F}_{k+\tau_k} \right) \right\} | \mathcal{F}_k \right) \right\} \\
&= \min_{v_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, v_{k+\tau_k}) + \mathbb{E}\big(J^*_{k+\tau_k+1}(x_{k+\tau_k+1}) | \mathcal{F}_k \big) \right\} \\
&= \min_{v_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, v_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k,\tau_k) J^*_{k+\tau_k+1}(j) \right\} \\
&= V^*_k(x_k) \qquad (22)
\end{aligned}
$$

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

$$
p(j|x_k,\tau_k) = \begin{cases} \dfrac{1-p_{\tau_k} p_{\tau_{k+\tau_k}}}{N-1}, & j \in \Omega_x,\ j \neq x_{k+\tau_k} \\ p_{\tau_k} p_{\tau_{k+\tau_k}}, & j = x_{k+\tau_k} \end{cases} \qquad (23)
$$

and N represents the number of the states in Ω_x. Equation (23) 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 every 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, the Bellman equation (8) can be expressed as

$$
\begin{aligned}
J^*_k(x_k) = {} & \frac{1-p_{\tau_k}}{\tau_k} \sum_{j=0}^{\tau_k-1} U(x_{k+j},0) \\
& + \min_{v_{k+\tau_k}} \left\{ p_{\tau_k} U(x_{k+\tau_k}, u_{k+\tau_k}) \right. \\
& \qquad \left. + \sum_{j\in\Omega_x} p(j|x_k,\tau_k) J^*_{k+\tau_k+1}(j) \right\}. \qquad (24)
\end{aligned}
$$

The Bellman equation (24) can be called "three-tuple Bellman equation."

Remark 3: Different from the traditional Bellman equation (8), the three-tuple Bellman equation (24) has an advantage. Due to the three-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, which is obtained by real sequence of spike train, for any state $x_k \in \Omega_x$, $k = 0, 1, 2, \ldots$, the next spiking instant is $k + \tau_k$, i.e., $\pi_{k+\tau_k} = 1$ with the probability of Poisson distribution, which ensures that the three-Tuple Bellman equation (24) only requires to minimize the $v_{k+\tau_k}$ rather than minimize the π_k and v_k simultaneously.

For every 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, according to (24), we can define the control law $v^*_{k+\tau_k}(x_{k+\tau_k})$, which is expressed as

$$
\begin{aligned}
v^*_{k+\tau_k}(x_{k+\tau_k}) = \arg\min_{v_{k+\tau_k}} & \left\{ \mathcal{U}_{\tau_k}(x_k, v_{k+\tau_k}) \right. \\
& \left. + \sum_{j\in\Omega_x} p(j|x_k,\tau_k) J^*_{k+\tau_k+1}(j) \right\}. \qquad (25)
\end{aligned}
$$

Thus, the following theorems can be derived.

Theorem 2: Let $J^*_k(x_k)$ be the optimal performance index function in (7) and $V^*_k(x_k)$ be the optimal spiking value function in (18). For every 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, we can obtain that the $v^*_{k+\tau_k}(x_{k+\tau_k})$ defined in (25) is the corresponding optimal control law.

Proof: For every 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$, according to (9), we can derive that $u^*_j = \mu(\pi^*_j, v^*_j) = \mu(0,0) = 0$ for instants $j = k, k+1, \ldots, k+\tau_k-1$. Thus, for instant $k+\tau_k$, according to (24), the optimal control law can be derived as

$$
\begin{aligned}
u^*_{k+\tau_k}(x_k) = \arg\min_{u_{k+\tau_k}} & \mathbb{E}\left\{ \sum_{j=0}^{\tau_k-1} U(x_{k+j},0) + U(x_{k+\tau_k}, u_{k+\tau_k}) \right. \\
& \left. + J^*_{k+\tau_k+1}(x_{k+\tau_k+1}) | \mathcal{F}_k \right\} \\
= \arg\min_{v_{k+\tau_k}} & \left\{ \mathcal{U}_{\tau_k}(x_k, v_{k+\tau_k}) \right. \\
& \left. + \sum_{j\in\Omega_x} p(j|x_k,\tau_k) J^*_{k+\tau_k+1}(j) \right\} \\
= {} & v^*_{k+\tau_k}(x_{k+\tau_k}). \qquad (26)
\end{aligned}
$$

The proof is complete. ■

Lemma 1: Let $J^*_k(x_k)$ and $V_{i+1}(x_k)$, $i = 0, 1, \ldots$, be defined in (24) and (20), respectively. Then, $J^*_k(x_k)$ and $V_{i+1}(x_k)$, $i = 0, 1, \ldots$, are positive definite functions for x_k.

The conclusion is easily derived according to the definition of the utility function and assumption 1, and the proof is omitted here.

Theorem 3: For $i = 0, 1, 2, \ldots$, and any $(x_k, \tau_k, p_{\tau_k})$ $k = 0, 1, 2, \ldots$, let $V_{i+1}(x_k)$ and $v_i(x_k)$ be the iterative value function and the iterative control law updated, respectively. According to (19) and (20) in Algorithm 2. Then, $V_i(x_k)$ converges to the optimal performance index function $J^*_k(x_k)$ as $i \to \infty$, which is defined as (24), that is

$$
\lim_{i\to\infty} V_i(x_k) = J^*_k(x_k). \qquad (27)
$$

Proof: The proof can be divided into three steps.

Let $(x_t, \tau_t, p_{\tau_t})$ $t = 0, 1, 2, \ldots$, and $(x_r, \tau_r, p_{\tau_r})$ $r = 0, 1, 2, \ldots$, be two 3-tuples, where (τ_t, p_{τ_t}), $t = 0, 1, 2, \ldots$ and (τ_r, p_{τ_r}), $r = 0, 1, 2, \ldots$, are calculated from the same spike train and the same fixed time interval under the Poisson process. Then, we can implement the first step of the proof.

Step 1): For any state $\mathbf{x} \in \Omega_x$, the optimal performance index function satisfies

$$
J^*_t(\mathbf{x}) = J^*_r(\mathbf{x}). \qquad (28)
$$

According to (6), we can obtain

$$
J_t(x_t) = \mathbb{E}\left(\sum_{j=0}^{\infty} U(x_{t+j}, u_{t+j}) | \mathcal{F}_t \right) \qquad (29)
$$

and

$$
J_r(x_r) = \mathbb{E}\left(\sum_{j=0}^{\infty} U(x_{r+j}, u_{r+j}) | \mathcal{F}_r \right). \qquad (30)
$$

Let $\mu^1(\pi_t(x_t), \nu_t(x_t))$ be the optimal spiking control law, generating the optimal control sequence $u^1 = \{0, 0, 0, \ldots, u^1_{t+\tau_t}, \ldots\}$, which minimizes $J_t(x_t)$, that is

$$J_t^*(\mathbf{x}) = \frac{1-p_{\tau_t}}{\tau_t}\sum_{j=0}^{\tau_t-1} U(x_{t+j}, 0) + p_{\tau_t} U(x_{t+\tau_t}, u_{t+\tau_t}) + \sum_{j\in\Omega_x} p(j|x_t, \tau_t) J^*_{t+\tau_t+1}(j) \quad (31)$$

where $x_t = \mathbf{x}$. Since (τ_t, p_{τ_t}) and (τ_r, p_{τ_r}) are computed by the same data, they have the same admissible spiking intervals and probability via Poission distribution. Therefore, it is feasible to substitute u^1 into (30), that is, $u_{r+j} = u^1_{t+j}$, $j = 0, 1, \ldots$ Letting $x_r = \mathbf{x}$ and $u_r = u^1_t$, it yields $x_{r+1} = F(x_r, u^1_t) = x_{t+1}$. Thus, for any $j = 0, 1, \ldots$, $U(x_{t+j}, u^1_{t+j}) = U(x_{r+j}, u^1_{t+j})$ can be obtained with initial states $x_r = x_t = \mathbf{x}$ and $\mathcal{F}_r = \mathcal{F}_r = \mathcal{F}$, which means

$$\begin{aligned} J_t^*(\mathbf{x}) &= \mathbb{E}\left(\sum_{j=0}^{\infty} U\left(x_{t+j}, u^1_{t+j}\right)\big|\mathcal{F}\right) \\ &= \mathbb{E}\left(\sum_{j=0}^{\infty} U\left(x_{r+j}, u^1_{t+j}\right)\big|\mathcal{F}\right) = J_r(\mathbf{x}). \end{aligned} \quad (32)$$

Next, by the contradiction, we prove that $\mu^1(\pi_t(x_t), \nu_t(x_t))$ is also the optimal spiking control law for $J_r(x_r)$. Assume that there exists another optimal spiking control law $\mu^2 = (\pi_r(x_r), \nu_r(x_r))$ for (30), where $\mu^2(\cdot) \neq \mu^1(\cdot)$. Let $u^2 = \{0, 0, 0, \ldots, u^2_{r+\tau_r}, \ldots\}$ be the optimal control sequence generated by $\mu^2(\pi_r(x_r), \nu_r(x_r))$. According to (32) and (30), with the initial state $x_r = \mathbf{x}$, we have

$$\begin{aligned} J_t^*(\mathbf{x}) &= \mathbb{E}\left(\sum_{j=0}^{\infty} U\left(x_{t+j}, u^1_{t+j}\right)\big|\mathcal{F}\right) = J_r(\mathbf{x}) \\ &\geq J_r^*(\mathbf{x}) \\ &= \mathbb{E}\left(\sum_{j=0}^{\infty} U\left(x_{r+j}, u^2_{r+j}\right)\big|\mathcal{F}\right). \end{aligned} \quad (33)$$

Now, substituting u^2 into (29), with $x_t = \mathbf{x}$ and $u_t = u^2_r$, we have $x_{t+1} = F(x_t, u^2_r) = x_{r+1}$. Following this iteration, $U(x_{t+j}, u^2_{r+j}) = U(x_{r+j}, u^2_{r+j})$, $j = 0, 1, 2, \ldots$, can be derived easily. It shows $J_t(\mathbf{x}) = J_r^*(\mathbf{x})$ under $\mu^2(\pi_r(x_r), \nu_r(x_r))$. According to (32) and (33), we can derive that $J_t^*(\mathbf{x}) \geq J_t(\mathbf{x})$, which is a contradition for $J_t^*(\mathbf{x})$ as its definition. Thus, the assumption is false and the conclusion holds.

Letting $t = k$ and $r = k + \tau_k + 1$, we can have

$$J_k^*(\mathbf{x}) = J^*_{k+\tau_k+1}(\mathbf{x}) \quad (34)$$

with the any initial state $\mathbf{x}$.

Next, $J_k^*(x_k)$ and $V_{i+1}(x_k)$, $i = 0, 1, \ldots$, are positive definite functions for x_k, $k = 0, 1, 2, \ldots$ Then, inspired by article [38], [47], there exist two constants $\underline{\delta}$ and $\overline{\delta}$, $0 \leq \underline{\delta} \leq 1 \leq \overline{\delta} < \infty$, such that

$$\underline{\delta} J_k^*(x_k) \leq V_0(x_k) \leq \overline{\delta} J_k^*(x_k). \quad (35)$$

Moreover, due to the positive definiteness of the spiking utility function, we can derive that there must exist two constants $\underline{\eta}$ and $\overline{\eta}$, $0 \leq \underline{\eta} \leq \overline{\eta} < \infty$, such that

$$\begin{aligned} \underline{\eta}\, \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) &\leq \sum_{j\in\Omega_x} p(j|x_k, \tau_k) J_k^*(j) \\ &\leq \overline{\eta}\, \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) \end{aligned} \quad (36)$$

for every 3-tuple $(x_k, \tau_k, p_{\tau_k})$, $k = 0, 1, 2, \ldots$ Then, we can implement the second step.

Step 2): For all $x_k \in \Omega_x$, the iterative spiking value function $V_i(x_k)$, $i = 0, 1, 2, \ldots$, satisfies that

$$\begin{aligned} \left(1 + \overline{\eta}^i \frac{\underline{\delta}-1}{(1+\overline{\eta})^i}\right) J_k^*(x_k) &\leq V_i(x_k) \\ &\leq \left(1 + \underline{\eta}^i \frac{\overline{\delta}-1}{(1+\underline{\eta})^i}\right) J_k^*(x_k). \end{aligned} \quad (37)$$

Now, we prove the inequality (37) on the left by induction. For $i = 0$, the conclusion can be easily obtained by (35).

For $i = 1$, according to (20) and (34), we have

$$\begin{aligned} V_1(x_k) &= \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k, \tau_k) V_0(j) \right\} \\ &\geq \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \underline{\delta} \sum_{j\in\Omega_x} p(j|x_k, \tau_k) J_k^*(j) \right\} \\ &\geq \min_{\nu_{k+\tau_k}} \left\{ \left(1 + \overline{\eta}\frac{\underline{\delta}-1}{1+\overline{\eta}}\right) \left(\mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k, \tau_k) J^*_{k+\tau_k+1}(j) \right) \right\} \\ &= \left(1 + \overline{\eta}\frac{\underline{\delta}-1}{1+\overline{\eta}}\right) J_k^*(x_k). \end{aligned} \quad (38)$$

Thus, the inequality (37) on the left holds for $i = 1$. Assuming that it holds for $i = n-1$, that is

$$\left(1 + \overline{\eta}^{n-1} \frac{\underline{\delta}-1}{(1+\overline{\eta})^{n-1}}\right) J_k^*(x_k) \leq V_{n-1}(x_k) \quad (39)$$

then we can obtain

$$\begin{aligned} V_n(x_k) &= \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k, \tau_k) V_{n-1}(j) \right\} \\ &\geq \min_{\nu_{k+\tau_k}} \left\{ \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \left(1 + \overline{\eta}^{n-1} \frac{\underline{\delta}-1}{(1+\overline{\eta})^{n-1}}\right) \times \sum_{j\in\Omega_x} p(j|x_k, \tau_k) J_k^*(j) \right\} \\ &\geq \min_{\nu_{k+\tau_k}} \left\{ \left(1 + \overline{\eta}^n \frac{\underline{\delta}-1}{(1+\overline{\eta})^n}\right) \mathcal{U}_{\tau_k}(x_k, \nu_{k+\tau_k}) + \left(1 + \underline{\delta}^{n-1} \frac{\underline{\delta}-1}{(1+\overline{\eta})^{n-1}} - \overline{\eta}^{n-1} \frac{\underline{\delta}-1}{(1+\overline{\eta})^n}\right) \right. \end{aligned}$$

$$\times \sum_{j\in\Omega_x} p(j|x_k,\tau_k)J_k^*(j)\Bigg\}$$
$$= \min_{\nu_{k+\tau_k}} \Bigg\{\left(1+\overline{\eta}^n\frac{\underline{\delta}-1}{(1+\overline{\eta})^n}\right)\Bigg(\mathcal{U}_{\tau_k}(x_k,\nu_{k+\tau_k}) + \sum_{j\in\Omega_x} p(j|x_k,\tau_k)J_{k+\tau_k+1}^*(j)\Bigg)\Bigg\}$$
$$= \left(1+\overline{\eta}^n\frac{\underline{\delta}-1}{(1+\overline{\eta})^n}\right)J_k^*(x_k). \tag{40}$$

Thus, the left half of (37) holds for $i=n$, which means that it holds for all $i=0,1,2,\ldots$ On the other hand, the proof of the right half is similar to (38)–(40), so it is omitted here. Hence, (37) holds for $i=0,1,\ldots$ The mathematical induction is complete.

Step 3): Prove that $V_i(x_k)$ converges to the optimal performance index function.

According to (37), for $k=0,1,2,\ldots$, as $i\to\infty$, we derive that

$$\lim_{i\to\infty}\left(1+\overline{\eta}^i\frac{\underline{\delta}-1}{(1+\overline{\eta})^i}\right)J_k^*(x_k) = \lim_{i\to\infty}\left(1+\underline{\eta}^i\frac{\overline{\delta}-1}{(1+\underline{\eta})^i}\right)J_k^*(x_k) = J_k^*(x_k) \tag{41}$$

which shows that the conclusion (27) is true. The proof is complete. ■

IV. Simulation Examples

In this section, two simulation examples are provided to show the effectiveness of the developed SADP method.

Example 1: We consider the torsional pendulum system to evaluate the performance of our developed algorithm. The dynamic system is expressed as

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = w$$
$$J\frac{\mathrm{d}w}{\mathrm{d}t} = u - Mgl\sin\theta - f_d\frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{42}$$

where J, M, g, l, and f_d are the rotary inertia, the mass, the gravity, the length of the pendulum bar, and the frictional factor, respectively. The parameters can be seen in Table I.

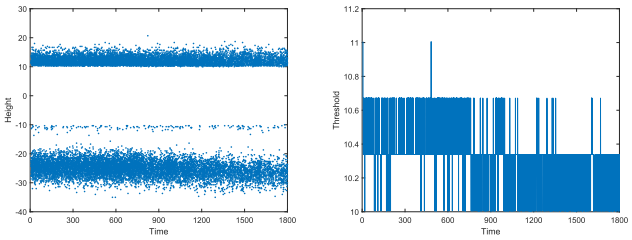
Discretizing the system using the Euler method with the sampling interval $\Delta t=0.01$ s, we can obtain

$$\begin{bmatrix}x_{1,k+1}\\x_{2,k+1}\end{bmatrix} = \begin{bmatrix}x_{1k}+\Delta t x_{2k}\\ -\frac{\Delta t Mgl}{J}\sin(x_{1k})+\left(1-\frac{\Delta t f_d}{J}\right)x_{2k}\end{bmatrix} + \begin{bmatrix}0\\\Delta t\end{bmatrix}u_k \tag{43}$$

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

TABLE I
Parameters of the Pendulum System

Parameters	Value
M	1/3 kg
g	9.8 m/s^2
l	3/2 m
J	4/3 ml^2
f_d	0.2
Δt	0.01 s



(a) (b)

Fig. 1. Height and the threshold of the spike train. (a) Height time. (b) Threshold time.

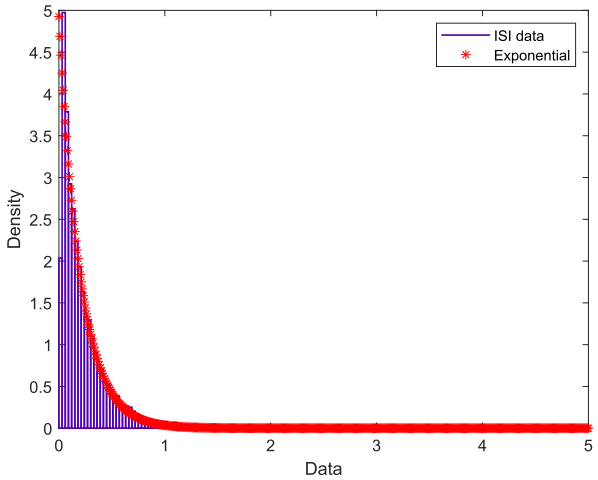


Fig. 2. Distribution of the ISI.

In this example, we use the dataset shared by Potter Lab [48], [49] to establish the 3-tuples. We choose the set of spike train, "6–3–34.spike," where "6–3" and "34" represent the serial number of nerve cell and days of neuron culture *in vitro*. What is more, one-channel electrode is chosen. The fixed time is 0.3 s. Applying Algorithm 1, the height and threshold of the spike train are shown in Fig. 1(a) and (b). Fig. 1 shows that the height has two obvious boundaries 10 and -10, while all the thresholds are greater than 10. The density can judge whether the spike train bursts.

Then, the ISI obeys the exponential distribution, which shows in Fig. 2. From Fig. 2, the blue area is ISI data and the red curve is the probability density function of the exponential distribution, the parameter $\bar{\lambda}=4.9256$, which is obtained by Algorithm 1. This also verifies that the occurrence of spike train obeys the Poisson process.

The state spaces and control spaces are $\Omega_x = \{x_k|x_k=[x_{1k},x_{2k}]^{\mathrm{T}}\in[-1.45,1.45]\times[-1.45,1.45]\}$ and

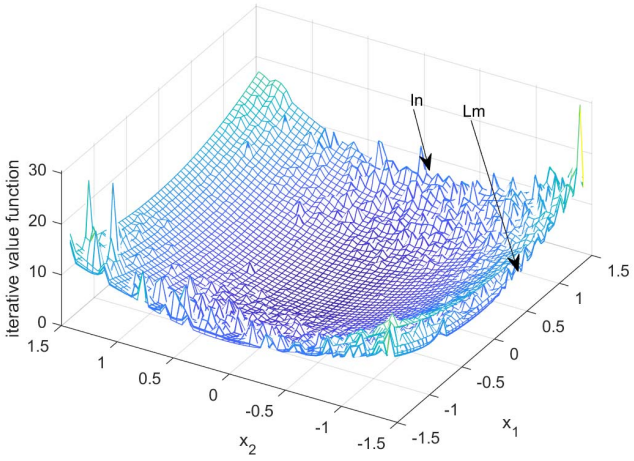


Fig. 3. Convergence plots of the iterative value functions for $\hat{\Omega}_x$.

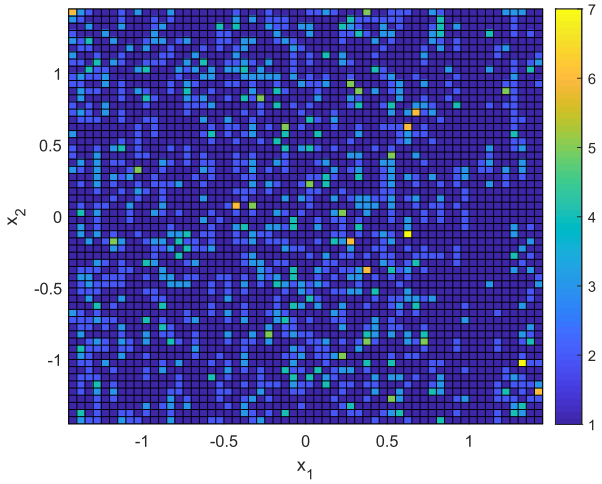


Fig. 4. Distribution of the optimal spiking intervals in the discretized state space $\hat{\Omega}_x$.

$\Omega_u = \{u_k | u_k \in [-2, 2]\}$, respectively. Choosing the discretization sizes $\sigma_x = 0.05$ and $\sigma_u = 0.01$, we have the value $\rho_x = 59 \times 59 = 3481$ and $\rho_u = 401$ and represent the number of discretized states and controls, respectively. Thus, the discretized state spaces $\hat{\Omega}_x = \{\hat{x}_k^{\varsigma} | \hat{x}_k^{\varsigma} = [\hat{x}_{1k}^{\varsigma}, \hat{x}_{2k}^{\varsigma}]^{\mathsf{T}}$, and $\varsigma = 1, 2, \ldots, \rho_x\}$ and control spaces $\hat{\Omega}_u = \{\hat{u}_k^{\varrho} | \ \varrho = 1, 2, \ldots, \rho_u\}$ can be obtained. We implement Algorithm 2 with $\hat{\Omega}_x$ for 20 iterations in order to urge the iterative value function to be convergent for all $\hat{x}_k^{\varsigma}, \varsigma = 1, 2, \ldots, \rho_x$. For the iterative value function $V_i(x_k)$ in the discretized state space $\hat{\Omega}_x$, the convergent plots are shown in Fig. 3.

In Fig. 3, "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. 4. From Fig. 4, seven kinds of intervals exist, which are from one to seven. For instance, if x_k was in the dark blue area, then the optimal spiking interval is 1, and if x_k was located in the cyan area, then the optimal spiking interval is 5. Thus, for all $x_k \in \hat{\Omega}_x$, the optimal spiking intervals can be obtained from Fig. 4.

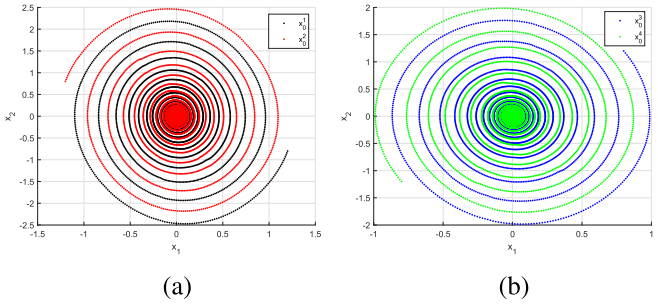


Fig. 5. Optimal trajectories of states. (a) Initialized with x_0^1 and x_0^2. (b) Initialized with x_0^3 and x_0^4.

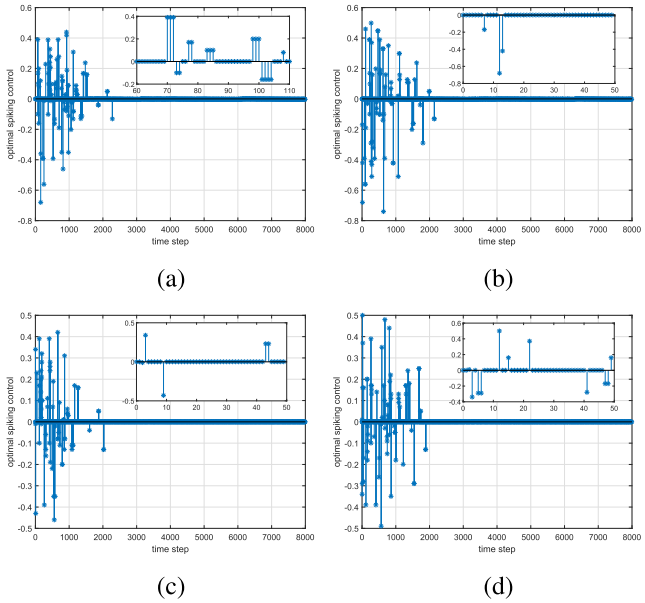


Fig. 6. Trajectroies of the optimal spiking controls with initial states x_0^i, $i = 1, 2, 3, 4$. (a) x_0^1. (b) x_0^2. (c) x_0^3. (d) x_0^4.

In this example, we choose four initial states, which are $x_0^1 = [1.2 \ \ -0.8]^{\mathsf{T}}$, $x_0^2 = [-1.2 \ 0.8]^{\mathsf{T}}$, $x_0^3 = [0.8 \ 1.2]^{\mathsf{T}}$, and $x_0^4 = [-0.8 \ \ -1.2]^{\mathsf{T}}$. Implementing the optimal spiking control law, obtained by Algorithm 2, to the system function (43) with the four initial states x_0^i, $i = 1, 2, 3, 4$ for 8000 time steps, we get the optimal state trajectories shown in Fig. 5 and the corresponding optimal spiking control is shown in Fig. 6(a)–(d), respectively.

From Figs. 5 and 6, the spiking control laws are obviously different for different system states at spiking instants. For example, the light blue area and cyan area in Fig. 4 indicate that the optimal spiking intervals are 3 and 5, respectively. On the other hand, from Fig. 6(b) and (c), the optimal spiking intervals for the initial state x_0^2 and x_0^3 are 5 and 3, respectively. This is a significant difference from those traditional ADP algorithms. Implementing our developed SADP algorithm based on the Poisson process, the optimal spiking control laws, including the optimal spiking instants and spiking control laws, can be obtained simultaneously, which verifies the effectiveness of the developed algorithm.

Example 2: We consider another discrete-time nonlinear system

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -x_1 + x_2 u$$

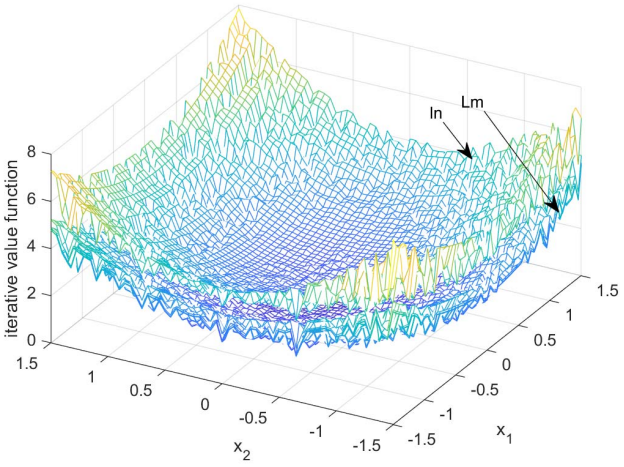


Fig. 7. Convergence plots of the iterative value functions for $\hat{\Omega}_x$.

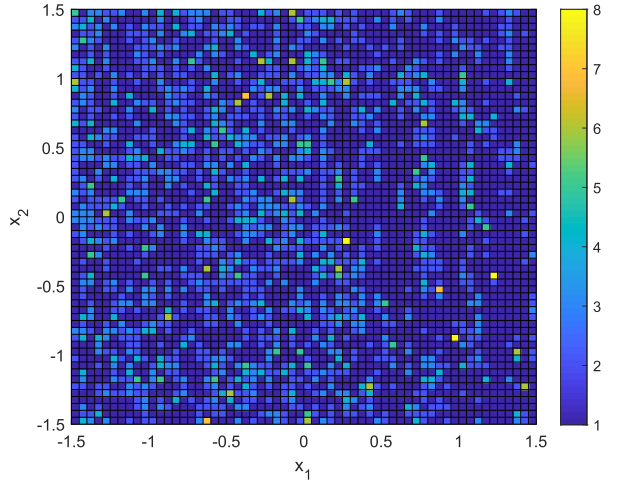


Fig. 8. Distribution of the optimal spiking intervals in the discretized state space $\hat{\Omega}_x$.

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_2 + (1+\cos^2(x_1))\sin(u). \tag{44}$$

Discretizing the system (44) with the sampling interval $\Delta t = 0.1$ s yields

$$\begin{aligned} x_{1,k+1} &= (1-\Delta t)x_{1k} + \Delta t x_{2k} u_k \\ x_{2,k+1} &= (1-\Delta t)x_{2k} + \Delta t u_k \\ &\quad + \Delta t[1+\cos^2(x_{1k})]\sin(u_k). \end{aligned} \tag{45}$$

The fixed time interval is 0.35, while the spike train, the utility function, and the initial value function are the same as before. The state spaces and control spaces are $\Omega_x = \{x_k | x_k = [x_{1k}, x_{2k}]^{\mathsf{T}} \in [-1.5, 1.5] \times [-1.5, 1.5]\}$ and $\Omega_u = \{u_k | u_k \in [-2, 2]\}$, respectively. With the discretization sizes $\sigma_x = 0.05$ and $\sigma_u = 0.01$, we have the discretized states $\rho_x = 61 \times 61 = 3721$ and discretized controls $\rho_u = 401$. Implementing Algorithm 2 for 20 iterations ensures the iterative value function to be convergent for all discretized states. The convergent plots and the distribution of the optimal spiking intervals are shown in Figs. 7 and 8, which are obviously different from Figs. 3 and 4. The iterative value functions in Fig. 7 are more rough than that in Fig. 3, while the spiking intervals in Fig. 8 have one more type than that in Fig. 4.

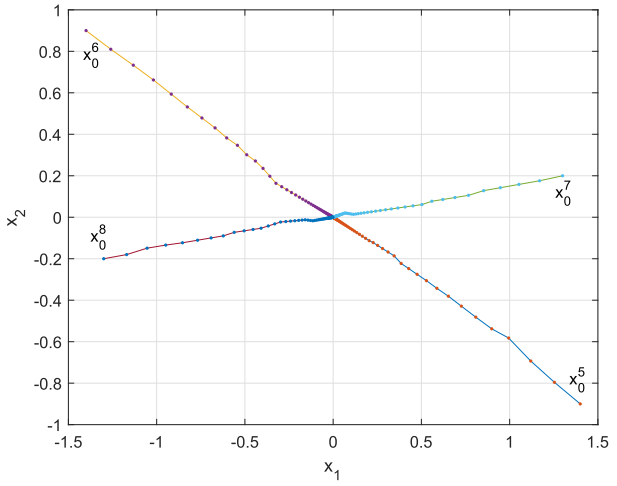


Fig. 9. Optimal trajectories of states initialized with x_0^5, x_0^6, x_0^7, and x_0^8.

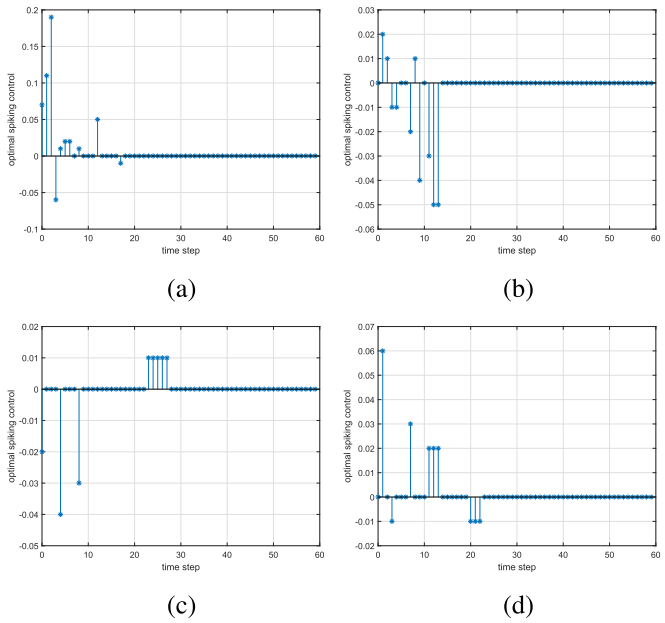


Fig. 10. Trajectories of the optimal spiking controls with initial states x_0^i, $i = 5, 6, 7, 8$. (a) x_0^5. (b) x_0^6. (c) x_0^7. (d) x_0^8.

Four initial states are chosen, which are $x_0^5 = [1.4 \ -0.9]^{\mathsf{T}}$, $x_0^6 = [-1.4 \ 0.9]^{\mathsf{T}}$, $x_0^7 = [1.3 \ 0.2]^{\mathsf{T}}$, and $x_0^8 = [-1.3 \ -0.2]^{\mathsf{T}}$. According to Algorithm 2, the optimal spiking control law can be obtained, which is applied to the system function (45) with four initial states x_0^i, $i = 5, 6, 7, 8$, for 60 time steps in order to get new optimal state trajectories. The figure of the optimal state trajectories and corresponding optimal spiking controls is shown in Figs. 9 and 10, respectively. From Figs. 9 and 10, it is obviously that different initial system states correspond to different spiking control laws at spiking instants. For instance, in Fig. 10(b), two different spiking control laws act on x_0^6 and x_1^6. Due to these actions of optimal spiking control laws in Fig. 10(a)–(d), initial states can converge on the equilibrium point, which verify the effectiveness of the present algorithm.

V. Conclusion

In this article, a new iterative SADP algorithm based on the Poisson process is presented to solve optimal control problems

for nonlinear systems. Before implementing this algorithm, by using the model of Poisson process and the method of MLE, we get the 3-tuples of states, spiking intervals, and probabilities of the Poisson distribution for a given fixed time interval in order to achieve the iterative value functions and iterative control laws. The property analysis is developed to guarantee that the value functions converge iteratively to the optimal performance index function as the iteration index increases from zero to infinity. Finally, two simulation examples are given to verify the effectiveness of the developed algorithm.

REFERENCES

[1] X.-H. Wang, J.-J. Yu, Y. Huang, H. Wang, and Z.-H. Miao, "Adaptive dynamic programming for linear impulse systems," *J. Zhejiang Univ. Sci. C*, vol. 15, no. 1, pp. 43–50, 2014.

[2] W. Li, L. Huang, Z. Guo, and J. Ji, "Global dynamic behavior of a plant disease model with ratio dependent impulsive control strategy," *Math. Comput. Simul.*, vol. 177, pp. 120–139, Nov. 2020, doi: 10.1016/j.matcom.2020.03.009.

[3] W. M. Haddad, V. Chellaboina, and N. A. Kablar, "Non-linear impulsive dynamical systems. Part II: Stability of feedback interconnections and optimality," *Int. J. Control*, vol. 74, no. 17, pp. 1659–1677, Jan. 2001.

[4] P. Sopasakis, P. Patrinos, H. Sarimveis, and A. Bemporad, "Model predictive control for linear impulsive systems," *IEEE Trans. Autom. Control*, vol. 60, no. 8, pp. 2277–2282, Aug. 2015.

[5] W.-H. Chen, S. Luo, and W. X. Zheng, "Generating globally stable periodic solutions of delayed neural networks with periodic coefficients via impulsive control," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1590–1603, Jul. 2017.

[6] J. Yao, Z.-H. Guan, G. Chen, and D. W. C. Ho, "Stability, robust stabilization and H_∞ control of singular-impulsive systems via switching control," *Syst. Control Lett.*, vol. 55, no. 11, pp. 879–886, Nov. 2006.

[7] X. Zhang, C. Li, and T. Huang, "Hybrid impulsive and switching hopfield neural networks with state-dependent impulses," *Neural Netw.*, vol. 93, pp. 176–184, Sep. 2017.

[8] Z. H. Guan, D. J. Hill, and J. Yao, "A hybrid impulsive and switching control strategy for synchronization of nonlinear systems and application to Chua's chaotic circuit," *Int. J. Bifurcation Chaos*, vol. 16, no. 1, pp. 229–238, 2006.

[9] W. Zhang, C. Li, S. Yang, and X. Yang, "Exponential synchronisation of complex networks with delays and perturbations via impulsive and adaptive control," *IET Control Theory Appl.*, vol. 13, no. 3, pp. 395–402, Feb. 2019.

[10] Y. Song, L. He, and Y. Wang, "Globally exponentially stable tracking control of self-restructuring nonlinear systems," *IEEE Trans. Cybern.*, early access, Nov. 19, 2020, doi: 10.1109/TCYB.2019.2951574.

[11] X. Li and S. Song, "Stabilization of delay systems: Delay-dependent impulsive control," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 406–411, Jan. 2017.

[12] Q. Zhang, L. Qiao, B. Zhu, and H. Zhang, "Dissipativity analysis and synthesis for a class of T–S fuzzy descriptor systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 8, pp. 1774–1784, Aug. 2017.

[13] B. Fan, Q. Yang, X. Tang, and Y. Sun, "Robust ADP design for continuous-time nonlinear systems with output constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2127–2138, Jun. 2018.

[14] X. Huang, H. K. Khalil, and Y. Song, "Regulation of nonminimum-phase nonlinear systems using slow integrators and high-gain feedback," *IEEE Trans. Autom. Control*, vol. 64, no. 2, pp. 640–653, Feb. 2019.

[15] Y.-D. Song, X. Huang, and Z.-J. Jia, "Dealing with the issues crucially related to the functionality and reliability of NN-associated control for nonlinear uncertain systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2614–2625, Nov. 2017.

[16] K. Zhao, Y. Song, and Z. Shen, "Neuroadaptive fault-tolerant control of nonlinear systems under output constraints and actuation faults," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 286–298, Feb. 2018.

[17] Q. Wei, H. Li, X. Yang, and H. He, "Continuous-time distributed policy iteration for multicontroller nonlinear systems," *IEEE Trans. Cybern.*, vol. 51, no. 5, pp. 2372–2383, May 2021, doi: 10.1109/TCYB.2020.2979614.

[18] Q. Wei, L. Zhu, R. Song, P. Zhang, D. Liu, and J. Xiao, "Model-free adaptive optimal control for unknown nonlinear multiplayer nonzero-sum game," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 27, 2020, doi: 10.1109/TNNLS.2020.3030127.

[19] Q. Yang and S. Jagannathan, "Reinforcement learning controller design for affine nonlinear discrete-time systems using online approximators," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 377–390, Apr. 2012.

[20] Q. Wei, L. Wang, Y. Liu, and M. M. Polycarpou, "Optimal elevator group control via deep asynchronous actor–critic learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5245–5256, Dec. 2020.

[21] C. Chen, H. Modares, K. Xie, F. L. Lewis, Y. Wan, and S. Xie, "Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4423–4438, Nov. 2019.

[22] X. Lin, N. Cao, and Y. Lin, "Optimal control for a class of nonlinear systems with state delay based on adaptive dynamic programming with ε-error bound," in *Proc. IEEE Symp. Adapt. Dyn. Program. Reinforcement Learn. (ADPRL)*, Singapore, Apr. 2013, pp. 177–182.

[23] F.-Y. Wang, N. Jin, D. Liu, and Q. Wei, "Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 24–36, Sep. 2010.

[24] H. Ren, H. Zhang, H. Su, and Y. Mu, "Data-based stable value iteration optimal control for unknown discrete-time systems with time delays," *Neurocomputing*, vol. 382, pp. 96–105, Mar. 2020.

[25] X. Lu, S. Tang, L. Zhang, P. Li, C. Li, and Y. Wang, "A novel steering control for real autonomous vehicles via PI adaptive dynamic programming," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Nanchang, China, Jun. 2019, pp. 926–930.

[26] H. Zhang, G. Xiao, Y. Liu, and L. Liu, "Value iteration-based H_∞ controller design for continuous-time nonlinear systems subject to input constraints," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 11, pp. 3986–3995, Nov. 2020, doi: 10.1109/TSMC.2018.2853091.

[27] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621–634, Mar. 2014.

[28] Q. Wei, Z. Liao, and G. Shi, "Generalized actor-critic learning optimal control in smart home energy management," *IEEE Trans. Ind. Informat.*, early access, Dec. 4, 2020, doi: 10.1109/TII.2020.3042631.

[29] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, Jan. 2021, Art. no. 142160.

[30] Q. Wei, L. Wang, J. Lu, and F.-Y. Wang, "Discrete-time self-learning parallel control," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Jun. 9, 2020, doi: 10.1109/TSMC.2020.2995646.

[31] Q. Wei, H. Li, and F.-Y. Wang, "Parallel control for continuous-time linear systems: A case study," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 919–926, Jul. 2020.

[32] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 2, pp. 168–176, Apr. 2017.

[33] X. Huang, Y. Song, and C. Wen, "Output feedback control for constrained pure-feedback systems: A non-recursive and transformational observer based approach," *Automatica*, vol. 113, Mar. 2020, Art. no. 108789.

[34] X. Huang, Y. Song, and J. Lai, "Neuro-adaptive control with given performance specifications for strict feedback systems under full-state constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 25–34, Jan. 2019.

[35] A. Heydari, "Optimal impulsive control using adaptive dynamic programming and its application in spacecraft rendezvous," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 17, 2020, doi: 10.1109/TNNLS.2020.3021037.

[36] D. Wang, M. Ha, and J. Qiao, "Self-learning optimal regulation for discrete-time nonlinear systems under event-driven formulation," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 1272–1279, Mar. 2020.

[37] D. Wang, M. Ha, and J. Qiao, "Data-driven iterative adaptive critic control toward an urban wastewater treatment plant," *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 7362–7369, Aug. 2021.

[38] Q. Wei, R. Song, Z. Liao, B. Li, and F. L. Lewis, "Discrete-time impulsive adaptive dynamic programming," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4293–4306, Oct. 2020.

[39] T. Zhang, Y. Zeng, D. Zhao, and M. Shi, "A plasticity-centric approach to train the non-differential spiking neural networks," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 620–627.
[40] E. M. Izhikevich, "Simple model of spiking neurons," *IEEE Trans. Neural Netw.*, vol. 14, no. 6, pp. 1569–1572, Nov. 2003.
[41] G. Sampath and S. Srinivasan, *Stochastic Models for Spike Trains of Single Neurons*. Berlin, Germany: Springer-Verlag, 2013.
[42] R. E. Kass *et al.*, "Computational neuroscience: Mathematical and statistical perspectives," *Annu. Rev. Statist. Appl.*, vol. 5, pp. 183–214, Mar. 2018.
[43] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 1994.
[44] M. Kordovan and S. Rotter, "Spike train cumulants for linear-nonlinear Poisson cascade models," 2020, *arXiv:2001.05057*. [Online]. Available: http://arxiv.org/abs/2001.05057
[45] C. E. R. Buxó and J. W. Pillow, "Poisson balanced spiking networks," *PLOS Comput. Biol.*, vol. 16, no. 11, Nov. 2020, Art. no. 836601.
[46] F. Gerhard, M. Deger, and W. Truccolo, "On the stability and dynamics of stochastic spiking neuron models: Nonlinear hawkes process and point process GLMs," *PLOS Comput. Biol.*, vol. 13, no. 2, Feb. 2017, Art. no. e1005390.
[47] M. Liang, D. Wang, and D. Liu, "Improved value iteration for neural-network-based stochastic optimal control design," *Neural Netw.*, vol. 124, pp. 280–295, Apr. 2020.
[48] J. P. Newman, M.-F. Fong, D. C. Millard, C. J. Whitmire, G. B. Stanley, and S. M. Potter, "Optogenetic feedback control of neural activity," *eLife*, vol. 4, Jul. 2015, Art. no. e07192.
[49] M.-F. Fong, J. P. Newman, S. M. Potter, and P. Wenner, "Upward synaptic scaling is dependent on neurotransmission rather than spiking," *Nature Commun.*, vol. 6, no. 1, pp. 1–11, May 2015.

Qinglai Wei (Senior Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2002 and 2009, respectively.

From 2009 to 2011, he was a Post-Doctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently a Professor of the Institute of Automation, Beijing, and the Associate Director of the State Key Laboratory of Management and Control for Complex Systems, Beijing. He has authored and coauthored four books and published more than 80 international journal articles. His research interests include adaptive dynamic programming, neural-network-based control, optimal control, nonlinear systems, and their industrial applications.

Dr. Wei is a Board of Governors (BOG) member of the International Neural Network Society (INNS) and a council member of CAA. He is the Secretary of the IEEE Computational Intelligence Society (CIS) Beijing Chapter since 2015. He was a Guest Editor for several international journals. He was a recipient of the IEEE/CAA JOURNAL OF AUTOMATICA SINICA Best Paper Award, the IEEE System, Man, and Cybernetics Society Andrew P. Sage Best Transactions Paper Award, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS Outstanding Paper Award, the Outstanding Paper Award of Acta Automatica Sinica, the IEEE 6th Data Driven Control and Learning Systems Conference (DDCLS2017) Best Paper Award, the Zhang Siying Outstanding Paper Award of Chinese Control and Decision Conference (CCDC), the Shuang-Chuang Talents in Jiangsu Province, China, Young Researcher Award of Asia Pacific Neural Network Society (APNNS), Young Scientist Award, and the Yang Jiachi Tech Award of Chinese Association of Automation (CAA).

Liyuan Han received the bachelor's degree in information and computing science, and electrical engineering and automation from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2019. He is currently pursuing the Ph.D. degree in control science and control engineering with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, and the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing.

His research interests include optimal control, adaptive dynamic programming, reinforcement learning, spiking neural network and their industrial applications.

Tielin Zhang received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2016.

He is an Associate Professor with the Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences. His current interests include theoretical research on neural dynamics and spiking neural networks.