

# Robust control for Markov jump linear systems with unknown transition probabilities – an online temporal differences approach

Transactions of the Institute of  
Measurement and Control  
1–9

© The Author(s) 2020

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0142331220940208

[journals.sagepub.com/home/tim](https://journals.sagepub.com/home/tim)



Yaogang Chen, Jiwei Wen<sup>✉</sup>, Xiaoli Luan<sup>✉</sup> and Fei Liu

## Abstract

In this paper, an online temporal differences (TD) learning approach is proposed to solve the robust control problem for discrete-time Markov jump linear systems (MJLS) subject to completely unknown transition probabilities (TP). The TD learning algorithm consists of two parts: policy evaluation and policy improvement. In the first part, by observing the mode jumping trajectories instead of solving a set of coupled algebraic Riccati equations, value functions are updated and approximate the TP related matrices. In the second part, new robust controllers can be obtained until value functions converge in the previous part. Moreover, the convergence of the value functions is proved by initializing a feasible control policy. Finally, two examples are presented to illustrate the effectiveness of the proposed approach by comparing with existing results.

## Keywords

Temporal differences, Markov jump systems, robust control, unknown transition probabilities

## Introduction

The past decades have witnessed great advances of Markov jump linear systems (MJLS) with an ever-increasing complexity. The systems can effectively model stochastic dynamical processes governed by a Markov chain. To name a few, fault-tolerant control systems with abrupt random faults and networked control systems where network-induced imperfections vary in a random way (see, for example, Costa et al., 2006; do Valle Costa et al., 2012; Shi and Li, 2015).

As an essential factor, transition probabilities (TP) depict the stochastic jumping and also determine the system behaviour. Under the assumption that TPs are precisely known, partially known, or involve uncertainties, many issues on MJLS have been investigated. For example, stability and stabilization have been studied in Bolzern et al. (2006); Hou and Ma (2015) and Xiao et al. (2010); Zhang and Boukas (2009c). Linear quadratic optimal control problem has been solved in Costa and Fragoso (1995) and Chizeck et al. (1986). High-order moment stabilization and filtering problems for MJLS have been addressed in Luan et al. (2018); Wan et al. (2018) and Zhou et al. (2020). The purpose of robust control is to maintain some performance of the system when the system contains unknown or uncertain information or is subject to external interference. One of the common methods is  $H_\infty$  control, which has received extensive attention. For instance, with known TPs, robust stability condition for MJLS with parameter uncertainty has been studied in de Souza (2006). The robust  $H_\infty$  control problem has been addressed in Chen et al. (2019) by adding a state observer. Asynchronous controllers

have been designed for Markov jump time-delay systems and two-dimensional MJLS in Shen et al. (2019) and Wu et al. (2018), respectively. The approach in Shi and Li (2015) utilized quantized signals to design a sliding mode observer, which can be applied to state estimation field. With partially unknown TPs, robust finite-time  $H_\infty$  state feedback control and  $H_\infty$  dynamic output control have been investigated in Zong et al. (2013) and Zhang and Boukas (2009a), respectively. The methods in Zhang and Boukas (2009b) solved  $H_\infty$  filtering design problem. In some cases, some uncertainties or time-varying properties may exist in TPs of the jump mechanism. As a result, there are many literatures on MJLS with uncertain TPs. For example, robust stability and output feedback control problems have been addressed in Xiong et al. (2005) and Fioravanti et al. (2013), respectively. Based on Zhang and Boukas (2009b), the partly known TPs model is proved to be a particular case of polytopic uncertain TPs and corresponding filter has been designed in Gonçalves et al. (2011). An element-wise method has been proposed in Xiong and Lam (2006), where the system matrix and the transition rate matrix are assumed to be bounded. Luan et al. (2012)

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, Wuxi, China

## Corresponding author:

Xiaoli Luan, Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, 1800 Lihu Avenue, Wuxi, 214122, China.

Email: [xiaoli\\_luan@126.com](mailto:xiaoli_luan@126.com)

assumed that TP uncertainties can be depicted by a Gaussian density function. Robust stability conditions for continuous-time MJLS are developed in Faraji-Niri et al. (2014, 2017), where the time-varying information of transition rates can be described by a piecewise constant.

In the above literatures, it is always assumed that some information of TPs matrix must be known. Then, the stability condition or optimal control objective is achieved by solving coupled algebraic Riccati equations (CARE) or transformed into linear matrix inequalities (LMI). However, in some practical applications, the transition between each jumping mode of MJLS is not clear, that is, the TPs of the Markov chain are unpredictable or unknown in advance. Therefore, it is difficult to directly obtain feasible solutions from CAREs or LMIs. As a result, there are many studies on the estimation for TPs, for example, online Bayesian estimation (Jilkov and Li, 2004) and maximum likelihood estimation (Beirigo et al., 2017; Orguner and Demirekler, 2008). In another way, MJLS with unknown TPs are regarded as arbitrary switched systems and the sufficient condition for  $H_\infty$  control is given in Zhang and Boukas (2009a). Recently, reinforcement learning (RL) (Sutton and Barto, 2018) and adaptive dynamic programming (ADP) have been widely used in the systems control field. For instance, Q-learning algorithms are developed to solve optimal control problems for discrete-time linear systems with unknown dynamics in Kiumarsi et al. (2014) and Jiang et al. (2019). The  $H_\infty$  optimal control problem for continuous-time nonlinear system, which is described as two-player zero-sum game, has been addressed by formulating an actor-critic neural network in Vamvoudakis and Lewis (2012) and a novel critic-only algorithm in Zhang et al. (2016). Based on ADP, time-triggered and event-triggered  $H_\infty$  constrained control schemes for strict-feedback nonlinear large-scale systems have been investigated in Tan (2018) and Tan (2019), respectively. A self-triggered control method has been proposed in Wan et al. (2020) to further reduce the resource occupation. Passive filter has been designed to estimate the neuron state of semi-Markov jump systems by the model of neural network in Shi et al. (2016). Costa and Aya (2002) developed an offline Monte Carlo TD method to solve the optimal control problem of MJLS with unknown TPs, where the historical data of Markov chains are observed to update the value function of the proposed iterative algorithm. Such an algorithm has been improved to the online scenario in Beirigo et al. (2018).

In this work, inspired by offline (Costa and Aya, 2002) and online (Beirigo et al., 2018) TD methods which deal with the jump linear quadratic problem, an improved TD approach is presented to address the robust control for MJLS, where the TP matrix including both deterministic and uncertain cases is completely unknown. Different from traditional methods to solve CAREs, TP related matrices are approximated by designed value functions, which are updated via observing modes jumpings in online algorithm. The main challenge is to ensure the convergence of the value function under the condition that the unknown TPs of MJLS are time-varying, which is proved in this paper. Moreover, the proposed algorithm is implemented online so that fast convergence can be achieved. We demonstrate the strength of our approach by a numerical example and Samuelson's

macroeconomic model with uncertain TPs by comparing with the existing results in Beirigo et al. (2018) and Luan et al. (2012), respectively.

The main contribution of this paper is to design an improved TD learning algorithm and prove that the value function converges to TP related matrices of  $H_\infty$  CAREs. The method in Beirigo et al. (2017) has weak disturbance attenuation capability and is only applicable in the constant TPs scenario. The estimation approaches, such as Orguner and Demirekler (2008) and Beirigo et al. (2017), have limited capability to deal with time-varying or uncertain TPs. The exiting results of Zhang and Boukas (2009a) in terms of LMIs are very conservative because they totally ignore TPs knowledge in the case that TPs are unknown but do exist. By contrast, the method in this paper can greatly reduce the conservatism and obtain more effective controller by comparing with Zhang and Boukas (2009a), because it takes advantage of the information of mode jumping information. In addition, the proposed approach can not only guarantee a desired  $H_\infty$  attenuation level, but also handle the case that the unknown TP matrix is actually random varying.

The structure of the paper is as follows. In the following section, MJLS is introduced and the robust control problem is formulated. Next, an online TD learning algorithm is proposed and the proof of convergence is given. The simulation results are provided to show the effectiveness of proposed methods, and the paper ends with some concluding remarks.

**Notation:** The notations used throughout this paper are standard.  $\mathbb{R}^n$  denotes  $n$ -dimensional Euclidean space,  $\mathbb{H}^n$  denotes the set of positive semi-definite  $n$ -by- $n$  matrices,  $A^T$  and  $A^{-1}$  represent the transpose and the inverse of matrix, respectively,  $\|A\|$  is the Euclidean norm of matrix  $A$ ,  $E\{\cdot\}$  stands for the mathematics statistical expectation of the stochastic process,  $l_2[0, \infty)$  denotes the space of square summable infinite sequence and  $r_\sigma(\cdot)$  is the spectral radius of the linear operator.

## Problem formulations

Considering the following MJLS with uncertain TPs:

$$\begin{cases} x(k+1) = A_{\theta(k)}x(k) + B_{\theta(k)}u(k) + \underline{G_{\theta(k)}}\omega(k) \\ z(k) = C_{\theta(k)}x(k) + D_{\theta(k)}u(k), \end{cases} \quad (1)$$

where  $x(k) \in \mathbb{R}^n$  is the state vector,  $u(k) \in \mathbb{R}^m$  is the control input,  $\omega(k) \in \mathbb{R}^r$  is the disturbance,  $z(k) \in \mathbb{R}^p$  is the control output.  $\theta(k)$  represents the state of a Markov chain taking values in a given finite set  $\Theta = \{1, 2, \dots, N\}$  with TP matrix  $\Pi = \{\pi_{ij}(\xi_k)\}$ ,  $i, j \in \Theta$ , and the TP from the mode  $i$  at time  $k$  to mode  $j$  at time  $k+1$  are considered as

$$\Pr\{\theta(k+1) = j | \theta(k) = i, \xi_k\} = \pi_{ij}(\xi_k), \quad (2)$$

where  $\xi_k$  is a random variable which denotes the uncertain information of TPs, for  $\theta(k) = i \in \Theta$ , the system matrices are denoted by  $(A_i, B_i, C_i, D_i, G_i)$  with appropriate dimensions.

**Assumption 1.** The uncertainty of TPs can be represented by a distribution and there is an expectation of TPs  $\hat{\pi}_{ij} = E[\pi_{ij}(\xi_k)]$ , for  $i, j \in \Theta$ .

**Definition 1.** System (1) with  $\omega \equiv 0$  and any initial condition  $x(0) \in \mathbb{R}^n$  and  $\theta(0) \in \Theta$  is said to be mean square stable (MSS) if

$$\lim_{k \rightarrow \infty} E\{\|x(k)\|^2\} = 0. \quad (3)$$

**Definition 2.** Given a scalar  $\delta > 0$ , system (1) is said to be MSS and has an  $H_\infty$  attenuation level  $\delta$  if

$$\sum_{k=0}^{\infty} E\{z^T(k)z(k)\} < \delta^2 \sum_{k=0}^{\infty} \omega^T(k)\omega(k), \quad (4)$$

holds for all nonzero  $\omega(k) \in L_2[0, \infty)$  under zero initial condition.

The purpose of robust control is to design state feedback controllers  $u(k) = F_{\theta(k)}x(k)$  for system (1) with unknown TPs such that the closed-loop system is MSS and has a prescribed  $H_\infty$  attenuation level. Then, the operators  $\varepsilon_i(\cdot)$  and  $\Gamma_j(\cdot)$  are defined as, respectively,

$$\varepsilon_i(X) \triangleq \sum_{j=1}^N \hat{\pi}_{ij} X_j, \quad i \in \Theta, \quad (5)$$

$$\Gamma_j(X) \triangleq \sum_{i=1}^N \hat{\pi}_{ij} K_i X_i K_i^T, \quad i \in \Theta, \quad (6)$$

where  $X = (X_1, \dots, X_N) \in \mathbb{R}^n$ ,  $K = (K_1, \dots, K_N) \in \mathbb{R}^n$  with  $\varepsilon(X) = (\varepsilon_1(X), \dots, \varepsilon_N(X))$  and  $\Gamma(X) = (\Gamma_1(X), \dots, \Gamma_N(X))$ .

**Lemma 1.** Suppose that  $(C_i, A_i)$  is mean square detectable for all  $i \in \Theta$  and  $\delta > 0$  is a given scalar. Then there exists  $F = (F_1, \dots, F_N)$  such that the closed-loop system is mean square stable with an  $H_\infty$  index  $\delta$ , if and only if there exists  $P = (P_1, \dots, P_N) \in \mathbb{H}^n$  satisfying the following conditions:

$$\delta^2 I - G_i^T \varepsilon_i(P) G_i > \alpha^2 I, \quad (7)$$

$$P_i = C_i^T C_i + F_i^T F_i - U_i^T U_i + (A_i + B_i F_i + \frac{1}{\delta} G_i U_i)^T \varepsilon_i(P) (A_i + B_i F_i + \frac{1}{\delta} G_i U_i), \quad (8)$$

$$r_\sigma(\Gamma) < 1, \quad (9)$$

for all  $i \in \Theta$  and some  $\alpha > 0$ , where  $\Gamma$  is as defined in (6) with  $K_i = A_i + B_i F_i + \frac{1}{\delta} G_i U_i$ ,  $F_i$  and  $U_i$  are calculated by (10) and (11).

$$F_i = \mathcal{F}(\varepsilon_i(P)) = -(I + B_i^T \varepsilon_i(P) [I - \frac{1}{\delta^2} G_i G_i^T \varepsilon_i(P)]^{-1} B_i)^{-1} \times (B_i^T \varepsilon_i(P) [I - \frac{1}{\delta^2} G_i G_i^T \varepsilon_i(P)]^{-1} A_i), \quad (10)$$

$$U_i = \mathcal{U}(\varepsilon_i(P)) = \frac{1}{\delta} (I - \frac{1}{\delta^2} G_i^T \varepsilon_i(P) [I + B_i B_i^T \varepsilon_i(P)]^{-1} G_i)^{-1} \times (G_i^T \varepsilon_i(P) [I + B_i B_i^T \varepsilon_i(P)]^{-1} A_i). \quad (11)$$

**Remark 1.** Lemma 1 is a sufficient and necessary condition proved in Costa et al. (2006) and can be also applied to the case where TPs are certain when  $\xi(k)$  is a constant. In order to obtain the solution of CARE (8) with known expectation of TPs, a standard method and a recursive algorithm are respectively proposed in Shi et al. (1999) and Costa et al. (2006). Then, if the solution satisfies (7) and (9), the robust controller of system (1) can be obtained by (10). However, the method of obtaining the TP related matrices  $\varepsilon(P)$  by solving the CAREs (8) will be invalid when TP matrix is completely unknown.

## Main results

In this paper, an TD-learning algorithm is proposed to approximate  $\varepsilon(P)$  online without any transition knowledge. By observing subsequent trajectories of Markov chains, the value function keeps updated and eventually converges to the TP related matrices  $\varepsilon(P)$ . Proof of convergence of the algorithm is also provided in this section.

### Online TD learning algorithm

The value function  $Y(t)$  in each mode of the proposed algorithm is given by

$$Y_i(t, k+1) = Y_i(t, k) + \gamma_i(t) e_i(t, k) d(t, k), \quad (12)$$

$$Y_i(t, 0) = Y_i(t), Y_i(t+1) = Y_i(t, N(t)), \quad (13)$$

where  $k$  stands for the time step,  $t$  is the episode of the algorithm, and stepsize  $\gamma_i(t)$  satisfies

$$\sum_{t=0}^{\infty} \gamma_i(t) = \infty, \quad \sum_{t=0}^{\infty} \gamma_i^2(t) < \infty, \quad (14)$$

the eligibility coefficients is given by

$$e_i(t, k) = \begin{cases} 0, & k < k_i(t) \\ \lambda^{k-k_i(t)}, & k \geq k_i(t), \end{cases} \quad (15)$$

for  $0 < \lambda < 1$ , where  $k_i(t) = \inf_k \{\theta(t, k) = i\}$ ,  $i \in \Theta$ , and the temporal difference  $d(t, k)$  is defined as

$$d(t, k) = R_{\theta(t, k+1)} + K_{\theta(t, k+1)}^T Y_{\theta(t, k+1)}(t, k) K_{\theta(t, k+1)} - Y_{\theta(t, k)}(t, k), \quad (16)$$

where the reward function  $R_i = C_i^T C_i + F_i^T F_i - U_i^T U_i$ ,  $K_i = A_i + B_i F_i + \frac{1}{\delta} G_i U_i$  for  $\theta(t, k) = i$ .

Note that the form of (12)–(16) is a standard online TD( $\lambda$ ) updating algorithm which is a branch of RL. As can be seen from Algorithm 1, the initial feasible policies  $F_i^0$  and  $U_i^0$  ensure the closed-loop stability of system (1). The above algorithm consists of two steps: policy evaluation and policy improvement. In the first step, the algorithm starts with an initial mode and eligibility coefficients with zero. Then, the value function  $Y$  is updated immediately when a mode jumping occurs. To ensure the convergence of  $Y$  at the  $t$  th Markov mode trajectory, for example,  $N(t)$  can be the length of  $t$  th trajectory or be defined as

**Algorithm 1.** TD learning

**Initialization.** Start with feasible stabilizing control policies  $F_i^0$  and  $U_i^0$ ,  $Y_i^0 = \mathbf{0}$ ,  $\forall i \in \Theta$ , the iteration number  $l = 0$

**Policy evaluation.**

```

1: for  $t = 1, \dots, T$  do
2:   Start with  $\theta(0)$ ,  $e_i(t, 0) = 0$ 
3:    $F_i = F_i^l$ ,  $U_i = U_i^l$ ,  $\forall i \in \Theta$ 
4:   for  $k = 1, \dots, N(t)$  do
5:     Observe the next mode  $\theta(k+1)$ 
6:     Update  $e_i(t, k)$  via (15)
7:     Update  $Y_i(t, k+1)$  via (12)
8:   end for
9: end for
10:  $Y_i^{l+1} = Y_i(T)$ 

```

**Policy improvement.**

```

11:  $F_i^{l+1} = F(Y_i^{l+1})$ 
12:  $U_{i+1} = \mathcal{U}(Y_i^{l+1})$ 

```

**Until**  $\|Y_i^l - Y_i^{l-1}\| < \varepsilon$ ,  $\forall i \in \Theta$  for a small positive value of  $\varepsilon$ , Otherwise set  $l = l + 1$  and go to step 1

$$N(t) = \inf_k \{k \geq 0, \|Y(t, k+1) - Y(t, k)\| < \alpha\}, \quad (17)$$

where  $\alpha > 0$  is precision parameter. In the second step, new policies  $F_i^{l+1}$  and  $U_i^{l+1}$  are obtained using  $Y_i$  learned in the first step. Until the value function  $Y_i^l$  converges, the robust controllers are obtained by step 11 in Algorithm 1.

**Remark 2.** In Algorithm 1, the value function  $Y_i(t)$  is updated quickly based on the real-time observations of the mode trajectories which contain the time-varying information of TPs. The convergence value of  $Y_i(t)$  is used to calculate a new control policy  $F_i^{l+1}$ . Then, the improved policy is applied to next policy evaluation step. The entire process repeats until the control policy converges and the desired robust controller is obtained. Compared with the results in Costa and Aya (2002) and Beirigo et al. (2018), the proposed approach can not only ensure the system (1) mean square stability with a desired  $\delta$  disturbance attenuation, but also deal with the case of uncertain or time-varying TPs. Compared with the method in Luan et al. (2012), where the uncertainty of TPs is assumed to be depicted by Gaussian distribution and then the expected TP matrix is calculated via the mean and variance known in advance, the approach in this paper does not require any information of TPs.

## ★ Proof of convergence

**Lemma 2.** If  $r(t)$  is a sequence generated by

$$r(t+1) = [1 - \gamma(t)]r(t) + \gamma(t)[H(r(t)) + W(t)] \quad (18)$$

and satisfies the following conditions:

- (a) The stepsize  $\gamma(t)$  satisfies (14).
- (b) For every  $t$ ,  $E\{W(t)\} = 0$  and there exist constants  $A$  and  $B$  such that

$$E\{W^2(t)\} \leq A + B\|r(t)\|^2. \quad (19)$$

- (c)  $H(\cdot)$  is a contraction mapping, and there exists a vector  $r$ , a scalar  $\beta \in [0, 1]$ , such that

$$\|H(r(t)) - r\| \leq \beta\|r(t) - r\| \quad (20)$$

Then,  $r(t)$  converges to  $r$  with probability 1.

**Theorem 1.** If policies  $F_i$  and  $U_i$  satisfies  $\|K_i\| < 1$ ,  $\forall i \in \Theta$ , and for some scalars  $\mu > 0$  and  $0 \leq \rho < 1$ ,  $P(N(t) = k) \leq \mu\rho^{-k}$ , then the value function  $Y(t)$  converges to  $\varepsilon(P)$  with probability 1.

**Proof.** Inspired by some simulation methods presented in Bertsekas and Tsitsiklis (1996), we develop complete proof for the convergence of the online TD method. First, the off-line form is defined as

$$Y_i^*(t+1) = Y_i^*(t) + \sum_{k=0}^{N(t)-1} \gamma_i(t) e_i(t, k) d^*(t, k), \quad (21)$$

where  $d^*(t, k) = R_{\theta(t, k+1)} + K_{\theta(t, k+1)}^T Y_{\theta(t, k+1)}^*(t) K_{\theta(t, k+1)} - Y_{\theta(t, k)}^*(t)$ . (21) also can be written as

$$Y_i^*(t+1) = (1 - \hat{\gamma}_i(t))Y_i^*(t) + \hat{\gamma}_i(t)H_i(Y^*(t)) + \hat{\gamma}_i(t)W_i(t), \quad (22)$$

where  $\hat{\gamma}_i(t) = \gamma_i(t)\Delta_i(t)$ ,  $\Delta_i(t) = E\left\{\sum_{k=k_i(t)}^{N(t)-1} e_i(t, k)\right\}$ ,  $H_i(Y^*(t)) = E\left\{\sum_{k=0}^{N(t)-1} e_i(t, k)d^*(t, k)\right\}/\Delta_i(t)$ .

Since  $1 < \Delta_i(t) < N(t)$ , it is easy to get  $\sum_{t=0}^{\infty} \hat{\gamma}_i(t) = \infty$  and  $\sum_{t=0}^{\infty} \hat{\gamma}_i^2(t) < \infty$ , satisfying the condition (a) of Lemma 2.  $W_i(t)$  is defined by

$$W_i(t) = \left\{\sum_{k=0}^{N(t)-1} e_i(t, k)d^*(t, k) - E\left[\sum_{k=0}^{N(t)-1} e_i(t, k)d^*(t, k)\right]\right\}/\Delta_i(t). \quad (23)$$

Obviously,  $E\{W_i(t)\} = 0$ . Note that  $\|K_i\| < 1$ , then,  $|d^*(t, k)| \leq 2\|Y^*(t)\| + R^*$ , where  $R^*$  is an upper bound for  $|R_i|$ . In this case,

$$\begin{aligned} E\{W_i^2(t)\} &\leq \left\{E\left[\sum_{k=0}^{N(t)-1} e_i(t, k)d_i^*(t, k)\right]/\Delta_i(t)\right\}^2 \\ &\leq N^2(t) \cdot (2\|Y^*(t)\| + R^*)^2 \\ &\leq N^2(t)(8\|Y^*(t)\|^2 + 2R^{*2}), \end{aligned} \quad (24)$$

which satisfies the condition (b) of Lemma 2.

Let  $P_i^* = \varepsilon_i(P) = \varepsilon_i(K^T P^* K + R)$ , then,

$$E\left\{R_{\theta(t, k+1)} + K_{\theta(t, k+1)}^T P_{\theta(t, k+1)}^* K_{\theta(t, k+1)} - P_{\theta(t, k)}^*\right\} = 0. \quad (25)$$

So, we obtain  $H(P^*) = P^*$ . Also, it can be proved that  $H(\cdot)$  is a contraction mapping in Bertsekas and Tsitsiklis (1996) (see 215-217). In this case,

$$\begin{aligned} \|H(Y^*(t) - P^*)\| &= \|H(Y^*(t)) - H(P^*)\| \\ &= \|H(Y^*(t) - P^*)\| \leq \beta \|Y^*(t) - P^*\|. \end{aligned} \quad (26)$$

The condition (c) of Lemma 2 is satisfied. Therefore, it is proved that  $Y^*(t)$  converges with probability 1, that is,  $\lim_{t \rightarrow \infty} Y^*(t) = \varepsilon_i(P)$ .

And (21) can be also written in incremental form as

$$Y_i^*(t, k+1) = Y_i^*(t, k) + \gamma_i(t) e_i(t, k) d^*(t, k), \quad (27)$$

$$Y_i^*(t+1) = Y_i^*(t, N(t)). \quad (28)$$

Note that  $e(t, k) \leq 1$ , we obtain that

$$\|Y_i^*(t, k) - Y_i^*(t)\| \leq k \gamma_i(t) (2\|Y^*(t)\| + R^*). \quad (29)$$

Then, we try to prove the following

$$\|Y_i(t, k) - Y_i^*(t, k)\| \leq V(k) \bar{\gamma}(t) \gamma_i(t) \quad (30)$$

by induction approach for some  $V(k)$ , where  $\bar{\gamma}(t) = \max_i \gamma_i(t)$ . Note that (28) makes sense for  $k=0$  with  $V(0)=0$ . For  $k \geq 1$ , (28) is assumed as an induction hypothesis. Then, we obtain that

$$\begin{aligned} \|Y_i(t, k) - Y_i^*(t)\| &= \|Y_i(t, k) - Y_i^*(t, k) + Y_i^*(t, k) - Y_i^*(t)\| \\ &\leq \|Y_i(t, k) - Y_i^*(t, k)\| + \|Y_i^*(t, k) - Y_i^*(t)\| \\ &\leq V(k) \bar{\gamma}(t) \gamma_i(t) + k \gamma_i(t) (2\|Y^*(t)\| + R^*). \end{aligned} \quad (31)$$

Using (29) and  $\|K_i\| < 1$ , we obtain

$$\|d(t, k) - d^*(t, k)\| \leq 2V(k) \bar{\gamma}^2(t) + 2\bar{\gamma}(t) N(t) (2\|Y^*(t)\| + R^*). \quad (32)$$

Then, using (28) and (30), one has

$$\begin{aligned} \|Y_i(t, k+1) - Y_i^*(t, k+1)\| &\leq [V(k)(1 + 2\bar{\gamma}(t)) + 2N(t)(2\|Y^*(t)\| + R^*)] \bar{\gamma}(t) \gamma_i(t). \end{aligned} \quad (33)$$

Hence, the proof of induction (28) is complete, with

$$V(k+1) = V(k)(1 + 2\bar{\gamma}(t)) + 2N(t)(2\|Y^*(t)\| + R^*). \quad (34)$$

Let  $k = N(t)$ , we have

$$V(N(t)) \leq \eta V^2(k)(1 + 2\bar{\gamma}(t))^{N(t)} (\|Y^*(t)\| + 1) \quad (35)$$

for some constant  $\eta$ , and

$$\|Y_i(t+1) - Y_i^*(t+1)\| \leq V(N(t)) \bar{\gamma}(t) \gamma_i(t). \quad (36)$$

Hence, using (33) and (34) yields

$$\begin{aligned} \left\| \frac{Y_i(t+1) - Y_i^*(t+1)}{\gamma_i(t)} \right\| &\leq V(N(t)) \bar{\gamma}(t) \\ &\leq \eta \bar{\gamma}(t) V^2(k)(1 + 2\bar{\gamma}(t))^{N(t)} (\|Y^*(t)\| + 1). \end{aligned} \quad (37)$$

Let  $q(T) = \bar{\gamma}(T) V^2(k)(1 + 2\bar{\gamma}(T))^{N(T)}$ , and  $T$  is large enough such that  $(1 + 2\bar{\gamma}(T))^2 \rho < 1$ . Note that  $P(N(t) = k) \leq \mu \rho^{-k}$ , we obtain

$$\begin{aligned} E\{q^2(T)\} &= E\{\bar{\gamma}^2(T) N^4(T)(1 + 2\bar{\gamma}(T))^{2N(T)}\} \\ &\leq \mu \bar{\gamma}^2(T) \sum_{k=1}^{\infty} k^4 (1 + 2\bar{\gamma}(T))^{2k} \rho^k \\ &\leq \Lambda \bar{\gamma}^2(T), \end{aligned} \quad (38)$$

for some constant  $\Lambda$ . Define  $\psi(T+1) = \sum_{k=0}^T q^2(k)$ , note that

$$E\{\psi(T+1)\} = \psi(T) + E\{q^2(T)\} \leq \psi(T) + \Lambda \bar{\gamma}^2(T). \quad (39)$$

Since  $\sum_{t=0}^{\infty} \bar{\gamma}^2(t) < \infty$ ,  $\psi(T)$  converges with probability 1. Thus,  $q(t)$  converges to zero,  $Y(t)$  and  $Y^*(t)$  will converge to the same value by (35) as  $t$  goes to infinity. The proof is complete.

**Remark 3.** Theorem 1 illustrates a prerequisite that the value function  $Y_i$  eventually converges to TP-related matrix  $\varepsilon_i(P)$  in Algorithm 1. The entire proof of convergence involves two steps: (i) we learn from Lemma 2 that the offline value function  $Y_i^*$  defined by (21) converges to  $\varepsilon_i(P)$ ; (ii) the difference between  $Y_i$  and  $Y_i^*$  tends to zero as  $t$  goes to infinity for every  $i \in \Theta$ .

## Illustrative examples

**Example 1.** To compare the TD learning method in this paper with previous approach Beirigo et al. (2018), we consider the MJLS (1) with following parameters:

$$\begin{aligned} A_1 &= \begin{bmatrix} 0 & 1 \\ -0.0176 & 0.9440 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 1 \\ -0.2315 & 1.2074 \end{bmatrix}, \\ A_3 &= \begin{bmatrix} 0 & 1 \\ 0.0254 & -0.4420 \end{bmatrix}, \end{aligned}$$

$$B_1 = B_2 = B_3 = [0 \quad 1]^T, C_1 = C_2 = C_3 = [0 \quad 1],$$

$$D_1 = D_2 = D_3 = 1,$$

$$G_1 = G_2 = G_3 = [1 \quad 0.8]^T.$$

And the TP matrix of the model is given by

$$\Pi = \begin{bmatrix} 0.67 & 0.17 & 0.16 \\ 0.30 & 0.47 & 0.23 \\ 0.26 & 0.10 & 0.64 \end{bmatrix}. \quad (40)$$

The TD learning algorithms in Beirigo et al. (2018) and in this paper for the above MJLS follow the same settings:  $N(t) = 15$ ,  $T = 500$ ,  $\lambda = 0.1$  and  $\gamma_i(t) = 1/t$ . In Beirigo et al. (2018), the jump linear quadratic optimal control problem was considered, where the controlled output index was required to be minimized. By previous online TD algorithm, the corresponding optimal controller can be obtained by



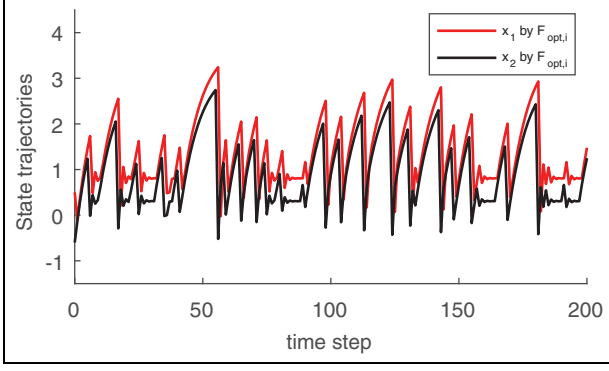


Figure 1. State trajectories of the closed-loop system by  $F_{opt,i}$ .

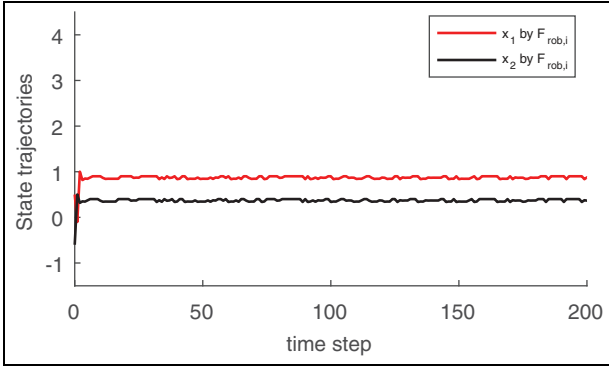


Figure 2. State trajectories of the closed-loop system by  $F_{rob,i}$ .

$$F_{opt,1} = [0.000895 \quad -0.0468], F_{opt,2} = [0.0159 \quad -0.0806], \\ F_{opt,3} = [-0.00450 \quad 0.0822].$$

Then, giving the  $H_\infty$  performance index  $\delta^* = 0.95$  and executing Algorithm 1, the robust controller is given by  $F_{rob,1} = [0.0209 \quad -1.1092]$ ,  $F_{rob,2} = [0.2675 \quad -1.3756]$ ,  $F_{rob,3} = [-0.0249 \quad 0.4423]$ .

Letting the initial condition  $x(0) = [0.5 \quad -0.6]^T$  and the external disturbance  $\omega(k) = 0.5(k)$ , the simulation results of the state response of closed-loop system (1) by optimal controller and robust controller are shown in Figure 1 and Figure 2, respectively. By observing the two figures, it can be found that both trajectories vibrate violently due to external disturbance in Figure 1, while the state curves are regulated approximately to equilibrium point in Figure 2. This implies the control policy obtained by proposed TD learning algorithm has better disturbance attenuation capability.

In Example 1, under the same conditions, our improved TD learning algorithm achieves better robustness by comparing with Beirigo et al. (2018). Next, we will provide another practical example to illustrate the convergence rapidity and accuracy of the proposed algorithm in the case that TPs are time-varying.

**Example 2.** In this part, the Samuelson's multiplier-accelerator model (Blair Jr and Sworder, 1975) which explains the periodic fluctuation phenomenon in the process of

economic growth is used to show the effectiveness of our result by comparing with the existing method in Luan et al. (2012). The state-space form which shows the effect of government expenditure on national income is as follows:

$$x(k+1) = \begin{bmatrix} 0 & 1 \\ -\alpha & 1-s+\alpha \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k), \quad (41)$$

where  $x(k)$  and  $u(k)$  denote national income and government expenditure, respectively,  $\alpha$  is the accelerator coefficients and  $s^{-1}$  is the multiplier. Based on the historical data from 1929 to 1971 by the US Department of Commerce, the model (40) can be classified by three modes according to different groups of  $s$  and  $\alpha$ : norm (both  $s$  and  $\alpha$  in mid-range), boom ( $s$  in low or  $\alpha$  in high range) and slump ( $s$  in high or  $\alpha$  in low range). And the corresponding parameters for the model with disturbance  $\omega(k)$  (net export capital) are as follows:

$$A_1 = \begin{bmatrix} 0 & 1 \\ -2.5 & 3.2 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 1 \\ -4.3 & 4.5 \end{bmatrix}, \\ A_3 = \begin{bmatrix} 0 & 1 \\ 5.3 & -5.2 \end{bmatrix}, B_1 = B_2 = B_3 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T, \\ C_1 = \begin{bmatrix} 1.5477 & -1.0976 \\ -1.0976 & 1.9145 \\ 0 & 0 \end{bmatrix}, \\ C_2 = \begin{bmatrix} 3.1212 & -0.5082 \\ -0.5082 & 2.7824 \\ 0 & 0 \end{bmatrix}, C_3 = \begin{bmatrix} 1.8385 & -1.2728 \\ -1.2728 & 1.6971 \\ 0 & 0 \end{bmatrix},$$

$$D_1 = D_2 = D_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T, G_1 = G_2 = G_3 = \begin{bmatrix} 0 & 0.1 \end{bmatrix}^T.$$

Here, different from Example 1, TPs of the above MJLS are time-varying and the uncertainty is described by a Gaussian distribution. The Gaussian transition probability density function (PDF) defined in Luan et al. (2012) is given by

$$N = \begin{bmatrix} n(0.67, 0.1) & n(0.17, 0.1) & n(0.16, 0.1) \\ n(0.30, 0.1) & n(0.47, 0.1) & n(0.23, 0.1) \\ n(0.26, 0.1) & n(0.10, 0.1) & n(0.64, 0.1) \end{bmatrix}, \quad (42)$$

where  $n(\mu_{ij}, \sigma_{ij})$  is the truncated Gaussian transition PDF of  $\pi_{ij}(\xi_k)$ ,  $\mu_{ij}$  and  $\sigma_{ij}$  are the mean and variance respectively.

Then by Algorithm 1,  $Y(0)$  is initialized with zeros, the length of the Markov chains  $N(t) = 20$ ,  $T = 500$ ,  $\lambda = 0.1$ , the  $H_\infty$  performance index  $\delta^* = 1$  and the stepsize  $\gamma_i(t) = 1/t$  for all  $i \in \Theta$ . The simulation results are illustrated in Figure 3, which show the evolutions of  $\|Y(t)\|$  of three modes with appropriate policies. In Figure 3, all three value functions converge rapidly in the policy evaluation step and the slight fluctuations are due to the single jump of the modes. To show the accuracy of our results, we define the error as follows:

$$\Delta_i = \frac{\|e_i(P) - Y_i\|}{\|e_i(P)\|}, \quad (43)$$

where  $Y_i$  is final output convergence value of  $Y_i(t)$  and  $e_i(P)$  is the standard solution of CARE (8). Then, we get the following results by  $\Delta_1 = 0.0872$ ,  $\Delta_2 = 0.0628$ ,  $\Delta_3 = 0.0480$ , which

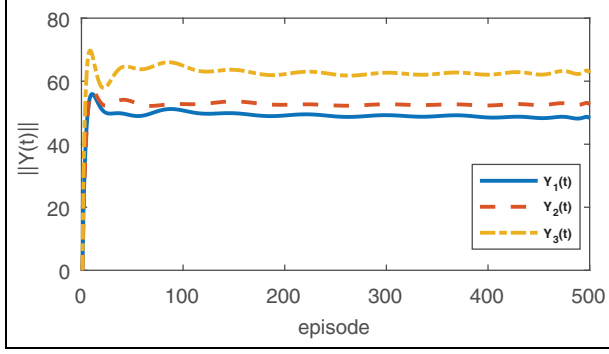


Figure 3. Evolution of  $\|Y(t)\|$  under 3 jumping modes.

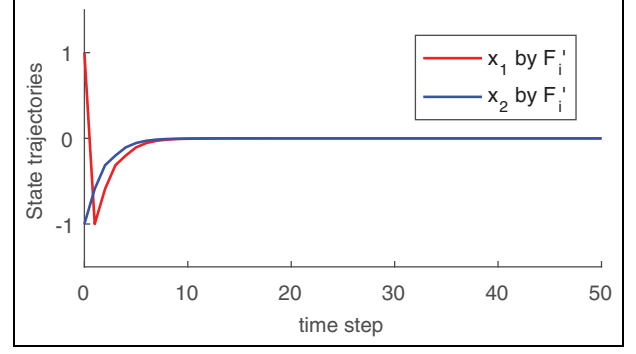


Figure 5. State trajectories of the closed-loop system by  $F'_i$ .

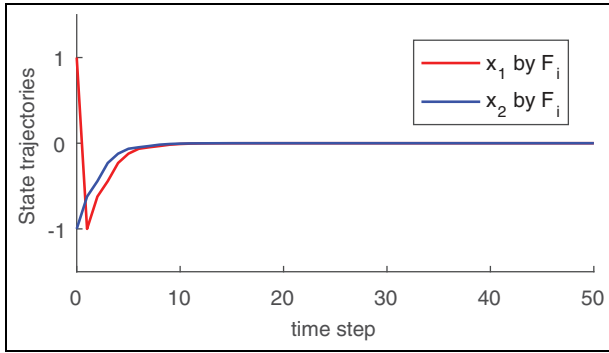


Figure 4. State trajectories of the closed-loop system by  $F_i$ .

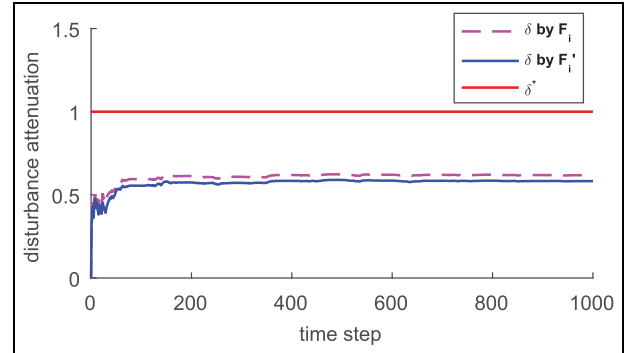


Figure 6. Comparisons between  $\delta^*$  and  $\delta(t)$ .

implies the precision of the proposed TD learning algorithm. The robust control gains can be obtained by  $F_i = \mathcal{F}_i(Y_i)$  as follows:  $F_1 = [2.4575 \quad -2.6197]$ ,  $F_2 = [4.2283 \quad -3.8338]$ ,  $F_3 = [-5.2250 \quad 5.7909]$ .

In Luan et al. (2012), the expectation of TPs  $\hat{\pi}_{ij}$  can be calculated from the truncated Gaussian transition PDF  $n(\mu_{ij}, \sigma_{ij})$ , which is assumed to be known in advance, by

$$\hat{\pi}_{ij} = \mu_{ij} + \frac{f\left(\frac{0-\mu_{ij}}{\sqrt{\sigma_{ij}}}\right) - f\left(\frac{1-\mu_{ij}}{\sqrt{\sigma_{ij}}}\right)}{F\left(\frac{0-\mu_{ij}}{\sqrt{\sigma_{ij}}}\right) - F\left(\frac{1-\mu_{ij}}{\sqrt{\sigma_{ij}}}\right)} \sqrt{\sigma_{ij}}, \quad (44)$$

where  $f(\cdot)$  is the PDF of the standard normal distribution and  $F(\cdot)$  is the cumulative distribution function of  $f(\cdot)$ . Based on the obtained  $\hat{\pi}_{ij}$ , sufficient conditions for existence of the desired controllers within LMIs are presented in Luan et al. (2012) and a feasible controller is given by:  $F'_1 = [2.4743 \quad -2.6352]$ ,  $F'_2 = [4.2587 \quad -3.8985]$ ,  $F'_3 = [-5.2558 \quad 5.7567]$ .

Based on the parameters above, some further simulations are performed to show the closed-loop stability and disturbance attenuation performance with initial condition  $x(0) = [1 \quad -1]^T$  and disturbance input  $\omega(k) = 0.2\sin(0.01\pi k + 0.1\pi)$ . Applying the obtained controllers  $F_i$  and  $F'_i$ , the trajectories of state are given in Figure 4 and Figure 5, where both state curves in the two figures are very similar and tend to zero, which means that the closed-loop system is mean square stable.

To show the disturbance attenuation performance, we define a  $H_\infty$  ratio as

$$\delta(k) = \sqrt{\frac{\sum_{m=0}^k z^T(k)z(k)}{\sum_{m=0}^k \omega^T(k)\omega(k)}}. \quad (45)$$

It can be seen in Figure 6 that two convergence values of the  $H_\infty$  ratio are very approximate ( $\delta$  by  $F_i$  is 0.6181,  $\delta$  by  $F'_i$  is 0.5824) and smaller than the upper bound  $\delta^* = 1$ , which implies that the proposed controller can achieve a prescribed  $H_\infty$  attenuation capability for the closed-loop system.

It requires 500 mode trajectories, each of which contains 20 jump modes, to execute Algorithm 1 to get the desired controller. In the Samuelson's macroeconomic system (41), the interaction between the multiplier analysis and the acceleration principle may lead to temporal business cycles. Based on the multiplier analysis, the national income is driven by the multiplier effect of investment and government expenditure. The acceleration principle states that changes in national income and consumption affect government expenditure in turn. This implies that the government can regulate the economic performance by changing fiscal and monetary policies to ensure the steady economic growth. Simulation results show that the obtained control signal  $F_i$  can meet the desired requirements.

The method in Luan et al. (2012) has to know the distribution, mean and variance of TPs in advance so as to calculate

the expected TP matrix and then obtain the controller by solving LMIs. However, the proposed approach is TP-free, that is, without any knowledge of TP matrix. By observing the Markov chains and updating the control policies, a desired robust controller can be obtained for the MJLS with unknown TP matrix, even if it involves randomness.

## Conclusion

In this work, a temporal differences learning algorithm has been presented to solve the robust control problem for Markov jump systems, where the unknown transition probabilities contain random uncertainty. In the policy evaluation step, the value function in the proposed algorithm is updated quickly by observing modes jumping trajectories. Then in the policy improvement step, a new robust controller is obtained. The above process repeats until the value function converges to the expected solution of Riccati equations. The convergence of value functions is proved and Samuelson's macroeconomic model with uncertain TPs is presented by comparing with existing methods, in which the mean and variance are assumed to be known in advance. As a result, an almost identical robust controller can be acquired, which implies the efficiency of the proposed TD algorithm. A possible direction for future work is to design reinforcement learning algorithm for MJLS control problems described by linear matrix inequalities.


## Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (Grant Nos. 61722306, 61833007 and 61991402).

## ORCID iDs

Jiwei Wen  <https://orcid.org/0000-0001-8780-4762>

Xiaoli Luan  <https://orcid.org/0000-0001-6576-8174>

## References

- Beirigo RL, Todorov MG and Barreto AdMS (2017) Count-based quadratic control of Markov jump linear systems with unknown transition probabilities. In: *IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, pp. 4315–4320.
- Beirigo RL, Todorov MG and Barreto AdMS (2018) Online TD ( $\lambda$ ) for discrete-time Markov jump systems. In: *IEEE 57th Annual Conference on Decision and Control (CDC)*, Miami Beach, FL, USA, 17–19 December 2018 pp. 2229–2234. IEEE.
- Bertsekas DP and Tsitsiklis JN (1996) *Neuro-dynamic Programming*, vol. 5. Belmont, MA: Athena Scientific/
- Blair W Jr and Sworder D (1975) Feedback control of a class of linear discrete systems with jump parameters and quadratic cost criteria. *International Journal of Control* 21(5): 833–841.
- Bolzern P, Colaneri P and De Nicolao G (2006) On almost sure stability of continuous-time Markov jump linear systems. *Automatica* 42(6): 983–988.
- Chen J, He T and Liu F (2019) Observer-based robust  $H_\infty$  control for uncertain Markovian jump systems via fuzzy lyapunov function. *Transactions of the Institute of Measurement and Control* 41(3): 657–667.
- Chizeck HJ, Willsky AS and Castanon D (1986) Discrete-time Markovian jump linear quadratic optimal control. *International Journal of Control* 43(1): 213–231.
- Costa OL and Aya JC (2002) Monte Carlo TD ( $\lambda$ )-methods for the optimal control of discrete-time Markovian jump linear systems. *Automatica* 38(2): 217–225.
- Costa OL and Fragoso MD (1995) Discrete-time LQ-optimal control problems for infinite Markov jump parameter systems. *IEEE Transactions on Automatic Control* 40(12): 2076–2088.
- Costa OLV, Fragoso MD and Marques RP (2006) *Discrete-time Markov jump linear systems*. Springer Science & Business Media.
- de Souza CE (2006) Robust stability and stabilization of uncertain discrete-time Markovian jump linear systems. *IEEE Transactions on Automatic Control* 51(5): 836–841.
- do Valle Costa OL, Fragoso MD and Todorov MG (2012) *Continuous-time Markov jump linear systems*. Springer Science & Business Media.
- Faraji-Niri M, Jahed-Motlagh MR and Barkhordari-Yazdi M (2014) Stochastic stabilization of uncertain Markov jump linear systems with time varying transition rates. In: *22nd Iranian Conference on Electrical Engineering (ICEE)*. IEEE, pp. 1186–1191.
- Faraji-Niri M, Jahed-Motlagh MR and Barkhordari-Yazdi M (2017) Stochastic stability and stabilization of a class of piecewise-homogeneous Markov jump linear systems with mixed uncertainties. *International Journal of Robust and Nonlinear Control* 27(6): 894–914.
- Fioravanti AR, Gonçalves AP and Geromel JC (2013) Discrete-time output feedback for Markov jump systems with uncertain transition probabilities. *International Journal of Robust and Nonlinear Control* 23(8): 894–902.
- Gonçalves AP, Fioravanti AR and Geromel JC (2011) Filtering of discrete-time Markov jump linear systems with uncertain transition probabilities. *International Journal of Robust and Nonlinear Control* 21(6): 613–624.
- Hou T and Ma H (2015) Exponential stability for discrete-time infinite Markov jump systems. *IEEE Transactions on Automatic Control* 61(12): 4241–4246.
- Jiang Y, Kiumarsi B, Fan J, Chai T, Li J and Lewis FL (2019) Optimal output regulation of linear discrete-time systems with unknown dynamics using reinforcement learning. *IEEE Transactions on Cybernetics* 50(7): 3147–3156.
- Jilkov VP and Li XR (2004) Online Bayesian estimation of transition probabilities for Markovian jump systems. *IEEE Transactions on Signal Processing* 52(6): 1620–1630.
- Kiumarsi B, Lewis FL, Modares H, Karimpour A and Naghibi-Sistani MB (2014) Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. *Automatica* 50(4): 1167–1175.
- Luan X, Huang B and Liu F (2018) Higher order moment stability region for Markov jump systems based on cumulant generating function. *Automatica* 93: 389–396.
- Luan X, Zhao S and Liu F (2012)  $H_\infty$  control for discrete-time Markov jump systems with uncertain transition probabilities. *IEEE Transactions on Automatic Control* 58(6): 1566–1572.



- Orguner U and Demirekler M (2008) Maximum likelihood estimation of transition probabilities of jump Markov linear systems. *IEEE Transactions on Signal Processing* 56(10): 5093–5108.
- Shen Y, Wu ZG, Shi P, Shu Z and Karimi HR (2019)  $H_\infty$  control of Markov jump time-delay systems under asynchronous controller and quantizer. *Automatica* 99: 352–360.
- Shi P and Li F (2015) A survey on Markovian jump systems: Modeling and design. *International Journal of Control, Automation and Systems* 13(1): 1–16.
- Shi P, Boukas EK and Agarwal RK (1999) Robust control for Markovian jumping discrete-time systems. *International Journal of Systems Science* 30(8): 787–797.
- Shi P, Li F, Wu L and Lim CC (2016) Neural network-based passive filtering for delayed neutral-type semi-Markovian jump systems. *IEEE Transactions on Neural Networks and Learning Systems* 28(9): 2101–2114.
- Shi P, Liu M and Zhang L (2015) Fault-tolerant sliding-mode-observer synthesis of Markovian jump systems using quantized measurements. *IEEE Transactions on Industrial Electronics* 62(9): 5910–5918.
- Sutton RS and Barto AG (2018) *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- Tan LN (2018) Distributed  $H_\infty$  optimal tracking control for strict-feedback nonlinear large-scale systems with disturbances and saturating actuators. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. Epub ahead of print. DOI: 10.1109/TSMC.2018.2861470.
- Tan LN (2019) Event-triggered distributed  $H_\infty$  constrained control of physically interconnected large-scale partially unknown strict-feedback systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. Epub ahead of print. DOI: 10.1109/TSMC.2019.2914160.
- Vamvoudakis KG and Lewis FL (2012) Online solution of nonlinear two-player zero-sum games using synchronous policy iteration. *International Journal of Robust and Nonlinear Control* 22(13): 1460–1483.
- Wan H, Luan X, Karimi HR and Liu F (2018) High-order moment filtering for Markov jump systems in finite frequency domain. *IEEE Transactions on Circuits and Systems II: Express Briefs* 66(7): 1217–1221.
- Wan H, Luan X, Karimi H and Liu F (2020) Dynamic self-triggered controller co-design for Markov jump systems. *IEEE Transactions on Automatic Control*. Epub ahead of print. DOI: 10.1109/TAC.2020.2992564.
- Wu ZG, Shen Y, Shi P, Shu Z and Su H (2018) Control for 2-D Markov jump systems in roesser model. *IEEE Transactions on Automatic Control* 64(1): 427–432.
- Xiao N, Xie L and Fu M (2010) Stabilization of Markov jump linear systems using quantized state feedback. *Automatica* 46(10): 1696–1702.
- Xiong J and Lam J (2006) Fixed-order robust  $H_\infty$  filter design for Markovian jump systems with uncertain switching probabilities. *IEEE Transactions on Signal Processing* 54(4): 1421–1430.
- Xiong J, Lam J, Gao H and Ho DW (2005) On robust stabilization of Markovian jump systems with uncertain switching probabilities. *Automatica* 41(5): 897–903.
- Zhang L and Boukas EK (2009a)  $H_\infty$  control for discrete-time Markovian jump linear systems with partly unknown transition probabilities. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal* 19(8): 868–883.
- Zhang L and Boukas EK (2009b) Mode-dependent  $H_\infty$  filtering for discrete-time Markovian jump linear systems with partly unknown transition probabilities. *Automatica* 45(6): 1462–1467.
- Zhang L and Boukas EK (2009c) Stability and stabilization of Markovian jump linear systems with partly unknown transition probabilities. *Automatica* 45(2): 463–468.
- Zhang Q, Zhao D and Zhu Y (2016) Event-triggered  $H_\infty$  control for continuous-time nonlinear system via concurrent learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47(7): 1071–1081.
- Zhou Z, Luan X and Liu F (2020) High-order moment stabilization for Markov jump systems with attenuation rate. *Journal of the Franklin Institute* 356: 9677–9688.
- Zong G, Yang D, Hou L and Wang Q (2013) Robust finite-time  $H_\infty$  control for Markovian jump systems with partially known transition probabilities. *Journal of the Franklin Institute* 350(6): 1562–1578.