https://doi.org/10.1016/j.automatica.2023.110917

A Distributed Stochastic Approximation Algorithm for Stochastic LQ Control with Unknown Uncertainty ⋆⋆

Zhaorong Zhang [a], Juanjuan Xu [a] and Xun Li [b]

[a] School of Control Science and Engineering, Shandong University, Jinan 250061, China

[b] Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong 999077, China

Abstract

This paper studies a discrete-time stochastic control problem with linear quadratic criteria over an infinite-time horizon. We focus on control systems whose system matrices are associated with random parameters involving unknown statistical properties. We design a distributed stochastic approximation algorithm to tackle the Riccati equation and derive the optimal controller stabilizing the system. The convergence analysis is provided.

Key words: Distributed stochastic approximation, stochastic control, multiplicative noise, unknown statistics.

1 Introduction

The past few decades have witnessed the rapid development of the studies of stochastic systems. In particular, stochastic linear-quadratic (LQ) control has become a research focus widely used in engineering systems. To name a few, [2] focused on output feedback regulation problems where the output signal is affected by additive noises, and they devised a robust linear control strategy. [3] solved the singular LQ problem for singular stochastic systems with additive noise by applying dynamic programming principle. However, the systems with multiplicative noises can better model the uncertainties of the system, such as packet dropouts, quantization errors, the constraints on signal-to-noise ratios and bandwidth limits. Therefore, stochastic LQ control with multiplicative noises has been extensively investigated.

In the literature (e.g., [4] [5]), system matrices and parameters should be known in advance, which is unrealistic in most applications. When the system dynamics and parameters are incomplete, reinforcement learning (RL) algorithms have been widely applied. Especially, the Q-learning algorithm is a kind of RL methods widely used in LQ control and regulation problems (e.g., [6] [7] [8]). Recently, [9] devised a Q-learning algorithm to solve the Bellman equation of discrete-time LQ control problems, where system parameters are associated with multiplicative noise with unknown Gaussian distribution.

Since the algorithm proposed by [9] is centralized, all data is transmitted in one packet, which creates an opportunity for attackers to steal all information. To cope with this, we propose a distributed stochastic approximation algorithm for stochastic control problems with random parameters involving unknown statistical information. Inspired by the work in [10], we design a novel distributed stochastic approximation algorithm. Under the distributed stochastic approximation scheme, we can approximate the zero point of a matrix equation, parameterize it to derive the solution of the Riccati equation, and design the optimal controller. The convergence of the proposed algorithm has been presented, and its correctness has been verified by numerical examples.

[⋆] Corresponding author: Juanjuan Xu.

[⋆⋆] This work is supported by the National Natural Science Foundation of China 61821004, 62250056, 61922051, 61873332 and U1806204, the Natural Science Foundation of Shandong Province ZR2021ZD14 and ZR2021JQ24, Science and Technology Project of Qingdao West Coast New Area 2019-32, 2020-20 and 2020-1-4, High-level Talent Team Project of Qingdao West Coast New Area RCTD-JC-2019-05, Key Research and Development Program of Shandong Province 2020CXGC01208 and the Research Grant Council of Hong Kong 15216720, 15221621 and 15226922.

Email addresses: zhaorong.zhang@uon.edu.au (Zhaorong Zhang), juanjuanxu@sdu.edu.cn (Juanjuan Xu), li.xun@polyu.edu.hk (Xun Li).

2 Problem Formulation and Preliminaries

2.1 Problem Formulation

We consider a discrete-time system described by:

$$x(k+1) = A(k)x(k) + B(k)u(k), \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the state, $u(k) \in \mathbb{R}^m$ is the control input, $A(k)$ and $B(k)$ are random matrices with compatible dimensions, which are written as: $A(k) = A+\bar{A}\omega(k)$, $B(k) = B + \bar{B}\omega(k)$. Here, A, $\bar{A}$, B and $\bar{B}$ are constant matrices, $\omega(k)$ is a multiplicative noise which follows the Gaussian distribution $N(\mu, \sigma^2)$. In particular, the statistics information μ and σ^2 are unknown when designing the controller. The cost functional is defined as

$$J(x,u) = \sum_{k=0}^{\infty} \left[x(k)^T \ u(k)^T\right] N \left[x(k)^T \ u(k)^T\right]^T, \tag{2}$$

where N is a diagonal matrix denoted as $N = \text{diag}\{Q, R\}$, where $Q{>}0$ and $R{>}0$ are constant matrices. Since the cost functional is given over an infinite horizon, the controller is chosen from the stabilizing ones. Namely, the admissible controller set is defined as $U = \{u(k) \Big| \mathbb{E} \sum_{k=0}^{\infty} \|u(k)\|^2{<}0, k \in \mathbb{N}\}$. Here, $u(k)$ is $\mathcal{F}(k-1)$-adapted, where $\mathcal{F}(k) = \sigma\{\omega(0), \cdots, \omega(k)\}$. Our objective is to minimize the expected value of the cost functional and obtain the optimal and stabilizing controller in admissible control set U.

Remark 1 *We emphasize that the LQ control studied in this paper is associated with multiplicative noise involving unknown statistical properties. Although the research on the LQ control problem of known system parameters has been quite mature, how to solve LQ control with unknown uncertainty in a distributed manner is still an open question.*

2.2 Preliminaries

The stabilizability of system (1) is essentially equivalent to determine whether the value function

$$V(k, x(k)) := \min_{u(s), s\geq k} \sum_{s=k}^{\infty} \mathbb{E}[x(s)^T Q x(s) + u(s)^T R u(s)]$$

is finite for all $x \in \mathbb{R}^n$. In addition, in the case where the system matrices are known, the generalized ARE can be expressed as:

$$\begin{aligned} P = {} & \mathbb{E}[Q + A(k)^T P A(k)] - \mathbb{E}[A(k)^T P B(k)] \\ & \times \mathbb{E}[B(k)^T P B(k) + R]^{-1} \mathbb{E}[B(k)^T P A(k)] \end{aligned} \tag{3}$$

Lemma 1 *Assume that equation (3) has a unique solution $P{>}0$, then the optimal controller is given by*

$$u^*(k) = -\left\{\mathbb{E}[B(k)^T P B(k) + R]\right\}^{-1} \mathbb{E}[B(k)^T P A(k)] x(k). \tag{4}$$

Also, the value function has the form of

$$V(k, x(k)) = \mathbb{E}[x(k)^T P x(k)]. \tag{5}$$

By simple calculation, one can observe that the optimal controller is closely related to the expectation and covariance of the random parameter. The authors of [9] have rewritten the generalized Riccati equation as:

$$P = \Pi\left(\mathbb{E}\begin{bmatrix} Q + A(k)^T P A(k) & A(k)^T P B(k) \\ B(k)^T P A(k) & B(k)^T P B(k) + R \end{bmatrix}\right), \tag{6}$$

where $\Pi(P) = P_{xx} - P_{xu} P_{uu}^{\dagger} P_{ux}$ is defined as a mapping for a matrix P according to the partition $P = \begin{bmatrix} P_{xx} & P_{xu} \\ P_{ux} & P_{uu} \end{bmatrix}$. Define $\Gamma(P) = -P_{uu}^{\dagger} P_{ux}$, then the optimal controller can be formulated as

$$u^*(k) = \Gamma\left(\mathbb{E}\begin{bmatrix} Q + A(k)^T P A(k) & A(k)^T P B(k) \\ B(k)^T P A(k) & B(k)^T P B(k) + R \end{bmatrix}\right) x(k).$$

Let

$$G = \mathbb{E}\begin{bmatrix} Q + A(k)^T P A(k) & A(k)^T P B(k) \\ B(k)^T P A(k) & B(k)^T P B(k) + R \end{bmatrix}, \tag{7}$$

the Riccati equation (6) can be equivalently written as

$$P = \Pi(G). \tag{8}$$

Substituting (8) into equation (7) yields

$$G = \mathbb{E}\begin{bmatrix} Q + A(k)^T \Pi(G) A(k) & A(k)^T \Pi(G) B(k) \\ B(k)^T \Pi(G) A(k) & B(k)^T \Pi(G) B(k) + R \end{bmatrix}. \tag{9}$$

In this scenario, solving the Riccati equation (6) can be converted into seeking for the zero point of equation (9). In [9], the authors applied an iterative algorithm:

$$G(k+1) = G(k) + \alpha(k) Y(G(k)), \tag{10}$$

where $Y(\cdot)$ is defined as

$$\begin{aligned} & Y(G(k)) \\ = & \begin{bmatrix} Q + A(k)^T \Pi(G(k)) A(k) & A(k)^T \Pi(G(k)) B(k) \\ B(k)^T \Pi(G(k)) A(k) & B(k)^T \Pi(G(k)) B(k) + R \end{bmatrix} \\ & - G(k) \end{aligned} \tag{11}$$

and $\alpha(k)$ is the learning rate sequence satisfying $\sum_{k=0}^{\infty} \alpha(k) = \infty$ *and* $\sum_{k=0}^{\infty} \alpha(k)^2 \leq \infty$.

Lemma 2 *Let $\{G(k)\}$ be the sequence constructed by stochastic approximation algorithm (10) for $k = 1, 2, \ldots$. Then, the following statements are equivalent:*

a. The LQ problem (1)-(2) is well-posed;
b. ARE (6) admits a solution $P>0$;
c. $\{G(k)\}$ is bounded with a positive probability;
d. $\{G(k)\}$ converges almostly surely (a.s.) to a deterministic matrix $G^* \in \mathbb{S}_+^{m+n}$.

Moreover, if either statement is valid, one has the following properties:

(1) The value function $V(0, x) = x^T P x$ for all $x \in \mathbb{R}^n$;
(2) The solution of ARE (6) is given by $P = \Pi(G^*)$;
(3) The optimal control is given by $u^*(k) = \Gamma(G^*) x(k)$;
(4) $G^* = \mathbb{E} \begin{bmatrix} Q + A(k)^T \Pi(G^*) A(k) & A(k)^T \Pi(G^*) B(k) \\ B(k)^T \Pi(G^*) A(k) & B(k)^T \Pi(G^*) B(k) + R \end{bmatrix}$.

Remark 2 *The algorithm proposed by [9] is centralized. That is, the algorithm requires the entire information of $G(k)$ and other system parameters to update $G(k+1)$ iteratively. However, this is not adaptive to problems where information security and privacy are emphasized.*

3 Main Results

3.1 A Distributed Stochastic Approximation Algorithm

Comparing to the centralized algorithm proposed by [9], our algorithm involves N sensors, each of which only has access to partial information. In particular, an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, which contains a vertex set $\mathcal{V}$ and an edge set $\mathcal{E}$, is formed by N sensors. Sensor j is said to be a neighbor of sensor i, if i and j are connected by an edge. The set of the neighbors of sensor i is denoted by $\mathcal{N}_i$. Each sensor $i = 1, \ldots, N$ is assigned to collect measurement data and carry out the estimates of G^*. The graph $\mathcal{G}$ is assumed to be connected throughout the paper. Select matrices L_i, $i = 1, \cdots, N$ such that $\sum_{i=1}^{N} L_i = NI$. And based on the graphical model above, each sensor iteratively computes

$$G_i(k+1) = G_i(k) + \sum_{j \in \mathcal{N}_i} (G_j(k) - G_i(k)) + \alpha(k) L_i Y(G_i(k)). \tag{12}$$

Remark 3 *In the proposed algorithm (12), sensor i only needs to use $G_i(k)$ and $G_j(k), j \in \mathcal{N}_i$ to update $G_i(k+1)$ instead of knowing the estimates of the whole network.*

3.2 Boundness of Distributed Algorithm (12)

Theorem 1 *Under the condition that ARE (3) has a positive-definite solution $P>0$, then $\{G_i(k), k \geq 0\}$ is bounded with a positive probability for each $i = 1, 2, \cdots, N$.*

Proof: Since $Q > 0$ and $R > 0$, it follows from (9) that $G \geq diag\{Q, R\} \geq \varepsilon I$, where $\varepsilon > 0$. In addition, from the non-decreasing property of Π in [9], it follows from (6) that we have $P \geq \Pi(diag\{Q, R\}) \geq \varepsilon I$. Let G^* be the solution to (9) and $P = \Pi(G^*)$. There exist invertible matrices T_1 and T_2 such that $T_1^T P T_1 = I, T_2^T G_{uu}^* T_2 = I$. Let $T = \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix}$ with $C = -{G_{uu}^*}^{-1} G_{ux}^*$. Also define $\tilde{\Upsilon}(k) = T_1^{-1} \Upsilon(k) T$, $\tilde{N} = T^T N T$ and $\Upsilon(k) = \begin{bmatrix} A(k) & B(k) \end{bmatrix}$. In view of the fact that $\Pi(T^T G T) = T_1^T \Pi(G) T_1$, we obtain

$$I = T_1^T P T_1 = T_1^T \Pi(G) T_1 = \Pi\Big(\mathbb{E}[\tilde{\Upsilon}^T(k) \tilde{\Upsilon}(k) + \tilde{N}]\Big). \tag{13}$$

We can infer that the solution to ARE (13) is I. Subsequently, we reformulate algorithm (12). By letting $\tilde{G}_i(k) = T^T G_i(k) T$, it follows from (12) that we have

$$\begin{aligned} \tilde{G}_i(k+1) = & \tilde{G}_i(k) + \sum_{j \in \mathcal{N}_i} (\tilde{G}_j(k) - \tilde{G}_i(k)) \\ & + \alpha(k) \Big(\tilde{\Upsilon}(k)^T \Pi(\tilde{G}_i(k)) \tilde{\Upsilon}(k) + \tilde{N} - \tilde{G}_i(k) \Big). \end{aligned}$$

To prove $\tilde{G}_i(k)$ is bounded a.s., we denote

$$\begin{aligned} \tilde{\Phi}_G(k) = \tilde{\Upsilon}(k)^T \Pi(G(k)) \tilde{\Upsilon}(k) + \tilde{N}, \tilde{\Phi}(G(k)) = \mathbb{E}\left[\tilde{\Phi}_G(k)\right], \\ \tilde{\Psi}_G(k) = \tilde{\Upsilon}(k)^T G_{xx}(k) \tilde{\Upsilon}(k) + \tilde{N}, \tilde{\Psi}(G(k)) = \mathbb{E}\left[\tilde{\Psi}_G(k)\right]. \end{aligned}$$

Now we prove that I is a fixed point of $\tilde{\Phi}(\cdot)$. Firstly, it can be verified that $G^* = \mathcal{T}^T \mathcal{T}$, where $\mathcal{T} = T^{-1}$. Hence, we have

$$\begin{aligned}\mathcal{T}^T\mathcal{T} =&\mathbb{E}\left[N + \Upsilon(k)^T\Pi(\mathcal{T}^T\mathcal{T})\Upsilon(k)\right]\\ =&\mathbb{E}\left[N + \Upsilon(k)^T T_1^{-T}T_1^{-1}\Upsilon(k)\right],\end{aligned}$$

which implies $\mathcal{T}^T\mathcal{T} = \mathbb{E}\left[N + \Upsilon(k)^T T_1^{-T}\Pi(I)T_1^{-1}\Upsilon(k)\right]$. Then we have $I = T^T\mathcal{T}^T\mathcal{T}T = \mathbb{E}\left[\tilde{N} + \tilde{\Upsilon}(k)^T\Pi(I)\tilde{\Upsilon}(k)\right]$ $= \tilde{\Phi}(I)$. Therefore, we can infer I is a fixed point of $\tilde{\Phi}(\cdot)$. We can further obtain $I = \tilde{\Psi}(I)$, that is, $\mathbb{E}\Big(\tilde{\Upsilon}(k)^T\tilde{\Upsilon}(k) + \tilde{N}\Big) = I$. In other words, I is a fixed point of $\tilde{\Psi}(\cdot)$. We are now ready to prove that $\tilde{\Psi}(\cdot)$ is a contraction mapping. Since $\mathbb{E}(N) > 0$, it follows that $\mathbb{E}(\tilde{N}) > 0$. Then there exists a positive number $\lambda < 1$ such that $\mathbb{E}\Big(\tilde{\Upsilon}(k)^T\tilde{\Upsilon}(k)\Big) = I - \tilde{N} \leq \lambda I$. Accordingly, for any M_1 and M_2, it follows that

$$\begin{aligned}&\Big\|\tilde{\Psi}(M_1) - \tilde{\Psi}(M_2)\Big\|_2\\ =&\Big\|\mathbb{E}\left[\tilde{\Upsilon}(k)^T M_{1,xx}\tilde{\Upsilon}(k) - \tilde{\Upsilon}(k)^T M_{2,xx}\tilde{\Upsilon}(k)\right]\Big\|_2\\ \leq&\lambda\|M_1 - M_2\|_2,\end{aligned}$$

which implies that $\tilde{\Psi}(M)$ is a contraction mapping with respective to M by using $\lambda < 1$. Define:

$$\begin{aligned}\mathcal{G}_i(k+1) =&\mathcal{G}_i(k) + \sum_{j\in\mathcal{N}_i}(\mathcal{G}_j(k) - \mathcal{G}_i(k)) + \alpha(k)\\ &\times[\tilde{\Psi}_{\mathcal{G}_i}(k) - \mathcal{G}_i(k)],\end{aligned} \tag{14}$$

with initial value $\mathcal{G}_i(0) = T^T G_i(0)T$. Since $\mathcal{G}_{ixx}(k) \geq \Pi(\mathcal{G}_i(k))$, we have that $\tilde{\Psi}_{\mathcal{G}_i}(k) \geq \tilde{\Phi}_{\mathcal{G}_i}(k)$. Thus, it holds that $\tilde{G}_i(k) \leq \mathcal{G}_i(k), k = 0, 1, 2, \cdots$, which gives that $\mathcal{G}_i(k)$ is an upper bound process of $\tilde{G}_i(k)$. Let $\hat{G}_i(k) = \mathcal{G}_i(k) - I$, it follows from (14) and $I = \mathbb{E}[\tilde{\Upsilon}(k)^T\tilde{\Upsilon}(k)] + \tilde{N}$ that,

$$\begin{aligned}&\hat{G}_i(k+1)\\ =&[1-\alpha(k)]\hat{G}_i(k) + \sum_{j\in\mathcal{N}_i}[\hat{G}_j(k) - \hat{G}_i(k)] + \alpha(k)\Theta_i(k),\end{aligned}$$

where

$$\begin{aligned}\Theta_i(k) =&\Big(\tilde{\Upsilon}(k)^T G_{ixx}(k)\tilde{\Upsilon}(k)\Big) - \Big(\tilde{\Upsilon}(k)^T G_{ixx}(k)\tilde{\Upsilon}(k)\Big)\\ &+ \Big(\tilde{\Upsilon}(k)^T G_{ixx}(k)\tilde{\Upsilon}(k)\Big) - [\tilde{\Upsilon}(k)^T\tilde{\Upsilon}(k)],\end{aligned}$$

together with $\mathbb{E}\Big(\tilde{\Upsilon}(k)^T\tilde{\Upsilon}(k)\Big) \leq \lambda I$, we have $\mathbb{E}[\Theta_i(k)|\mathcal{F}(k-1)] \leq \lambda\|\hat{G}_i(k)\|_2 I$. Moreover, it is easy to verify that $\mathbb{E}[\|\Theta_i(k)\|^2|\mathcal{F}(k-1)] \leq 36\mu + 30\mu\|\hat{G}_1(k)\|^2$, where μ satisfies $\mathbb{E}[\Upsilon(k)^T\Upsilon(k)] + N \leq \mu$. By applying similar discussions to Lemma 3.4 in [9], it yields that $\hat{G}_i(k)$ converges to 0 a.s., which implies that $\mathcal{G}_i(k)$ is bounded a.s.. As a consequence with $\tilde{G}_i(k) = T^T G_i(k)T$ and $\tilde{G}_i(k) \leq \mathcal{G}_i(k)$, it follows that $G_i(k)$ is bounded a.s..

3.3 Convergence Analysis

The convergence analysis consists of the following two parts:

$$\lim_{k\to\infty}\|G_i(k) - G_j(k)\| = 0, \forall i, j \in N_i, \ \ a.s., \tag{15}$$

$$\lim_{k\to\infty}\|G_i(k) - G^*\| = 0, \forall i \in N_i, \ \ a.s., \tag{16}$$

where (15) and (16) indicate (12) achieves consensus and the consensus value is the solution to (9), respectively.

3.3.1 Consensus Analysis

Theorem 2 *Suppose that equation (3) has the solution of $P{>}0$, then the proposed algorithm (12) achieves consensus.*

Proof:

Denote $F(k) = \mathbf{col}\{G_1(k), G_2(k), \ldots, G_N(k)\}$ and $\Phi(k) = \mathbf{col}\{L_1Y_1(k), L_2Y_2(k), \ldots, L_NY_N(k)\}$, it follows from (12) that $F(k+1) = \mathcal{A}F(k) + \alpha(k)\Phi(k)$, where $\mathcal{A} = I - L$ and L is the Laplacian matrix. Let $M = \dfrac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$ and $\delta(k) = (I - M)F(k)$, we have

$$\delta(k+1) = (\mathcal{A} - M)\delta(k) + \alpha(k)(I - M)\Phi(k), \tag{17}$$

where the facts $\mathcal{A}M = M\mathcal{A} = M^2 = M$ have been used in the derivation of the last equality. By applying iterative calculation to (17), it yields that

$$\begin{aligned}\delta(k+1) =&(\mathcal{A} - M)^{k+1}\delta(0) + \sum_{\tau=0}^{k}\alpha(\tau)(\mathcal{A} - M)^{k-\tau}\\ &\times(I - M)\Phi(\tau).\end{aligned}$$

To prove that the algorithm achieves consensus in the almost sure sense, the key point is to analyze the norm of $\delta(k)$. Specifically, we have:

$$\begin{aligned}&\|\delta(k)\|\\ =&\|(A - M)^k\delta(0) + \sum_{\tau=0}^{k-1}\alpha(\tau)(A - M)^{k-\tau-1}(I - M)\Phi(\tau)\|\\ \leq&c\rho^k\|\delta(0)\| + \sum_{\tau=0}^{k-1}\alpha(\tau)\|(A - M)^{k-\tau-1}\|\|I - M\|\|\Phi(\tau)\|.\end{aligned}$$

Since given a connected graph, it yields that $\|(\mathcal{A} - M)^k\| \leq c\rho^k$ where $c{>}0$ and $\rho \in (0, 1)$. According to

$\|I-M\|<\infty$, $\|\delta(0)\|<\infty$ and $\|\Phi(\tau)\|<\infty$, we have $\lim\limits_{k\to\infty} c\rho^k\|\delta(0)\| \to 0$ and $\lim\limits_{k\to\infty}\sum\limits_{\tau=0}^{k-1}\|(A-M)^{k-\tau-1}\|\|(I-M)\|\|\Phi(\tau)\| \to 0$. Thus, $\lim\limits_{k\to\infty}\|\delta(k)\|=0$, a.s. That is, equation (15) holds.

3.3.2 Convergence Analysis to the Solution of (12)

Theorem 3 *Under the assumption that ARE (3) has a solution $P>0$, then $G_i(k)$, $i=1,\ldots,N$ converge a.s. to $G(k)$.*

Proof: Define $\bar{G}(k+1)=\dfrac{1}{N}\sum\limits_{i=1}^{N}G_i(k+1)$. From (12), we have $\bar{G}(k+1)=\bar{G}(k)+\dfrac{\alpha(k)}{N}\sum\limits_{i=1}^{N}L_iY(G_i(k))$, which gives $\sum\limits_{i=1}^{N}L_iY_i(G^*)=0$. Recall algorithm (10) and let $\Delta(k)=\bar{G}(k)-G(k)$, the iteration equation of $\Delta(k)$ is:

$$\begin{aligned}\Delta(k+1)=&\Delta(k)+\frac{\alpha(k)}{N}\sum_{i=1}^{N}L_iY(G_i(k))-\alpha(k)Y(G(k))\\ =&[1-\alpha(k)]\Delta(k)+\alpha(k)\mathcal{W}(k)\\ &+\frac{\alpha(k)}{N}\Big[\sum_{i=1}^{N}L_i\Big(Y(G_i(k))-Y(\bar{G}(k))\Big)\Big].,\end{aligned}\tag{18}$$

where $\mathcal{W}(k)=\begin{bmatrix}A(k)^TWA(k) & A(k)^TWB(k)\\ B(k)^TWA(k) & B(k)^TWB(k)\end{bmatrix}$ with $W(k)=\Pi(\bar{G}(k))-\Pi(G(k)$. The derivation of (18) depends on the following fact: $Y(\bar{G}(k))-Y(G(k))=\mathcal{W}(k)-\Big(\bar{G}(k)-G(k)\Big)$. Denote $\Psi(k)=\mathcal{W}(k)+\dfrac{1}{N}\Big[\sum\limits_{i=1}^{N}L_i\Big(Y(G_i(k))-Y(\bar{G}(k))\Big)\Big]$. Then, equation (18) can be reformulated as $\Delta(k+1)=[1-\alpha(k)]\Delta(k)+\alpha(k)\Psi(k)$. Together with $\|G(k)\|<\infty$, it follows that, $\Big\|\mathcal{W}(k)\Big\|<\infty$. According to the above analysis, we have $\|\Psi(k)\|<\infty$. Applying the facts that $0<1-\alpha(k)<1$, $\|\Psi(k)\|<\infty$ and $\lim_{k\to\infty}\alpha(k)=0$, we obtain $\lim_{k\to\infty}\|\Delta(k+1)\|=0$. This gives the second condition (16).

Theorem 4 *Under the assumption that the ARE (3) has a positive-definite solution, the distributed algorithm (12) is able to converge a.s. to G^*.*

Proof: We have proved that the sensors can reach consensus and their consensus states will converge to

$$\begin{aligned}&\lim_{k\to\infty}\|G_i(k)-G_j(k)\|=0,\forall i,j\in N_i,\ a.s.,\\ &\lim_{k\to\infty}\|\bar{G}(k)-G(k)\|=0,\forall i,\ a.s.,\end{aligned}$$

which further imply that $\lim_{k\to\infty}\|G_i(k)-G(k)\|=0,\forall i$ a.s. Based on the convergence analysis in [9], we finally obtain $\lim_{k\to\infty}\|G_i(k)-G^*\|=0,\forall i$, a.s. Thus, the proposed algorithm converges a.s. to G^*.

4 Numerical Example

We implement algorithm (12) in a discrete-time model as follows: $A=\text{diag}\{0.2,0.6\}$, $\bar{A}=\text{diag}\{0.7,0.8\}$, $B=\begin{bmatrix}0.7 & 0.3\end{bmatrix}^T$, $\bar{B}=\begin{bmatrix}0.1 & 0.7\end{bmatrix}^T$, $Q=\text{diag}\{0.4,0.7\}$, $R=1$, $\alpha(k)=(\dfrac{1}{k+2})^{0.6}$. The random parameters follow the Gaussian distribution $N(\mu,\sigma^2)$, where $\mu=1$, $\sigma^2=0.1$. The system is associated with a networked system where sensors $i=1,\ldots,4$ are employed to calculate $G_i(k)$, which are the estimates of the Q-factor in the k-th iteration. The induced graph is shown by Fig. 1. Fig. 2 illustrates the 1-norm of $G_i(k)$ for $i=1,\ldots,4$ after running the proposed algorithm for 200 times. Fig. 3 reveals that for $i=1$, $G_1(k)$ converges to the correct solution G^*. Other sensors have similar convergence behaviors.

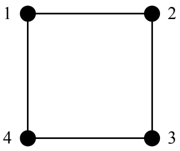

Fig. 1. A 4-sensor graph

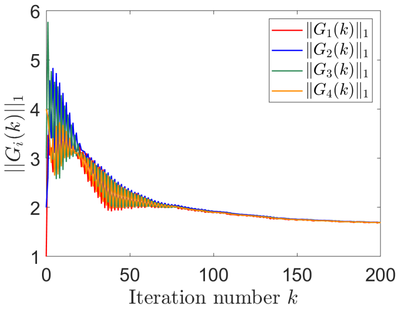

Fig. 2. Performance of the distributed algorithm

5 Conclusion

This paper presents a distributed stochastic approximation algorithm for stochastic LQ control involving random parameters with unknown statistical properties. We have proved that the correct solution to the Riccati equation and the optimal controller under the proposed distributed scheme can be derived. In the future, we are motivated to address stochastic LQ control problems associated with both unknown uncertainties and time delays by using distributed methods.

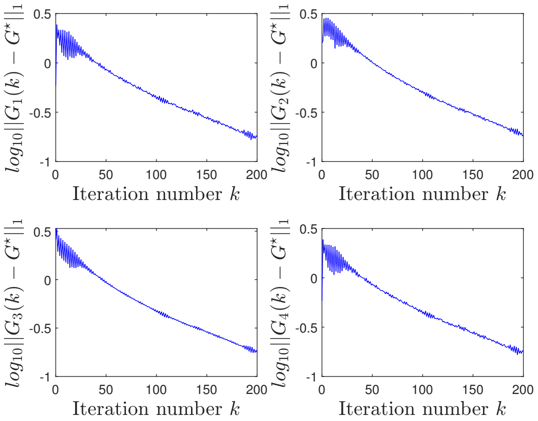

Fig. 3. Performance of the distributed algorithm

References

[1] W. M. Wonham, "Optimal stationary control of a linear system with state dependent noise," SIAM Journal on Control and Optimization, vol. 5, pp. 486–500, 1967.

[2] M. I. Taksar, A. S. Poznyak and A. Iparraguirre,"Robust output feedback control for linear stochastic systems in continuous time with time-varying parameters," IEEE Transactions on Automatic Control, vol. 43, no. 8, pp. 1133-1136, 1998.

[3] J. Feng, P. Cui and Z. Hou, "Singular linear quadratic optimal control for singular stochastic discrete-time systems. Optimal Control Applications and Methods, 34(5): 505–516, 2013.

[4] J. Willems and G. Blankenship,"Frequency domain stability criteria for stochastic systems," IEEE Transactions on Automatic Control, vol. 16, no. 4, pp. 292-299, 1971.

[5] H. Zhang, L. Li, J. Xu and M. Fu, "Linear quadratic regulation and stabilization of discrete-time systems with delay and multiplicative noise," IEEE Transactions on Automatic Control, vol. 60, no. 10, pp. 2599-2613, 2015.

[6] J. Li, T. Chai, F. L. Lewis, Z. Ding and Y. Jiang, "Off-Policy interleaved Q-learning: optimal control for affine nonlinear discrete-time systems," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 5, pp. 1308-1320, 2019.

[7] H. Xu and S. Jagannathan, "Stochastic optimal control of unknown linear networked control system using Q-learning methodology," Proceedings of the 2011 American Control Conference, 2011, pp. 2819-2824.

[8] L. Zhang, E.-K. Boukas, "Stability and stabilization of Markovian jump linear systems with partly unknown transition probabilities," Automatica, vol. 45, no. 2, pp. 463-468, 2009.

[9] K. Du, Q. Meng and F. Zhang, "A Q-learning algorithm for discrete-time linear-quadratic control with random parameters of unknown distribution: convergence and stabilization," SIAM Journal of Control and Optimization, vol. 60, no. 4, pp. 1991-2015, 2022.

[10] P. Bianchi, G. Fort and W. Hachem, "Performance of a distributed stochastic approximation Algorithm," IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7405-7418, 2013.