SCIENCE CHINA
Information Sciences

• **Supplementary File** •

An Online Value Iteration Method for Linear-Quadratic Mean Field Social Control with Unknown Dynamics

Bing-Chang Wang[1], Shumei Li[2*] & Ying Cao[3]

School of Control Science and Engineering, Shandong University, Jinan 250000*, China*

Notation: The following notation will be used throughout this paper. $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\leqslant t\leqslant T}, \mathbb{P})$ is a complete probability space. I denotes the identity matrix; $\mathbb{R}$ denotes the set of real numbers; S^n denotes the normed space of all n-by-n real symmetric matrices; X^T denotes the transpose of a vector or matrix X; $\langle\cdot,\cdot\rangle_F$ denotes the Frobenius inner product; $\|\cdot\|$ denotes the Euclidean vector norm or Frobenius matrix norm and $\langle\cdot,\cdot\rangle$ is the standard Euclidean inner product. For a vector z and a matrix Q, $\|z\|_Q^2 = z^TQz$, $Q>0(Q\geqslant 0)$ means that Q is positive definite (positive semidefinite). For a matrix $M\in\mathbb{R}^{n\times m}$, $\mathrm{vec}(M) = [M_1^T M_2^T \cdots M_m^T]^T$, where $M_i\in\mathbb{R}^n$ is the ith column of M. For any $M\in S^n$, let $\mathrm{vech}(M) = [M_{11}M_{12}\cdots M_{1n}M_{22}M_{23}\cdots M_{(n-1)n}M_{nn}]^T$, where $M_{ij}\in\mathbb{R}$ is the (i,j)th element of matrix M. Denote $u=\{u_1,\cdots,u_N\}$.

Appendix A Proof of proposition 1

First we denote

$$\tilde{x}_i(t) = x_i(t) - x^{(N)}(t), \qquad \tilde{u}_i(t) = u_i(t) - u^{(N)}(t),$$
$$\tilde{W}_i(t) = W_i(t) - W^{(N)}(t), \quad \Xi \triangleq \Gamma^TQ + Q\Gamma - \Gamma^TQ\Gamma,$$
$$u^{(N)}(t) \triangleq \frac{1}{N}\sum_{i=1}^N u_i(t), \qquad W^{(N)}(t) \triangleq \frac{1}{N}\sum_{i=1}^N W_i(t).$$

Under (**A1**)-(**A2**), the AREs (3) and (4) admit the unique symmetric postive definite solutions [1]. By (1) , we have

$$dx^{(N)}(t) = \big((A+G)x^{(N)}(t) + Bu^{(N)}(t)\big)dt + DdW^{(N)}(t). \tag{A1}$$

Subtracting (1) from (A1), we have

$$d\tilde{x}_i(t) = (A\tilde{x}_i(t) + B\tilde{u}_i(t))\,dt + Dd\tilde{W}_i(t), \quad 1\leqslant i\leqslant N. \tag{A2}$$

Then by Itô's formula, for any $i=1,2,\cdots N$, we obtain

$$\begin{aligned}\tilde{x}_i(T)^TP\tilde{x}_i(T) - \tilde{x}_i(0)^TP\tilde{x}_i(0) = &\int_0^T\Big\{\langle(A^TP+PA)\tilde{x}_i(t),\tilde{x}_i(t)\rangle + 2\langle PB\tilde{u}_i(t),\tilde{x}_i(t)\rangle + \frac{N-1}{N}D^TPD\Big\}dt\\ &+\int_0^T 2\tilde{x}_i(t)^TPDd\tilde{W}_i(t),\end{aligned} \tag{A3}$$

and

$$\begin{aligned}&x^{(N)}(T)^T\Pi x^{(N)}(T) - x^{(N)}(0)^T\Pi x^{(N)}(0)\\ &=\int_0^T\Big\{\langle[(A+G)^T\Pi + \Pi(A+G)]x^{(N)}(t),x^{(N)}(t)\rangle + 2\langle\Pi Bu^{(N)}(t),x^{(N)}(t)\rangle + \frac{1}{N}D^T\Pi D\Big\}dt\\ &+\int_0^T 2x^{(N)}(t))^T\Pi DdW^{(N)}(t).\end{aligned} \tag{A4}$$

Let

$$J_{\mathrm{soc}}^{(N)}(T,u) \triangleq \sum_{i=1}^N\int_0^T\{\|x_i(t)-\Gamma x^{(N)}(t)\|_Q^2 + \|u_i(t)\|_R^2\}dt.$$

Then by (3) , (4) , (A3), (A4) and direct calculations,

$$\begin{aligned}J_{\mathrm{soc}}^{(N)}(T,u) =&\sum_{i=1}^N\int_0^T\big[\|x_i(t)\|_Q^2 - \|x^{(N)}(t)\|_\Xi^2 + \|u_i(t)\|_R^2\big]dt\\ =&\sum_{i=1}^N\int_0^T\big[\|x_i(t)-x^{(N)}(t)\|_Q^2 + \|x^{(N)}(t)\|_{Q-\Xi}^2 + \|u_i(t)-u^{(N)}(t)\|_R^2 + \|u^{(N)}(t)\|_R^2\big]dt\end{aligned}$$

* Corresponding author (email: shumeili@mail.sdu.edu.cn.)

$$
\begin{aligned}
&= \sum_{i=1}^{N} \int_0^T \left[\|u_i(t) - u^{(N)}(t) + R^{-1}B^T P(x_i(t) - x^{(N)}(t))\|_R^2 + \|u^{(N)}(t) + R^{-1}B^T \Pi x^{(N)}(t)\|_R^2 \right] dt \\
&\quad + \frac{N-1}{N} \int_0^T D^T P D dt + \frac{1}{N} \int_0^T D^T \Pi D dt + \tilde{x}_i(0)^T P \tilde{x}_i(0) - \tilde{x}_i(T)^T P \tilde{x}_i(T) \\
&\quad + x^{(N)}(0)^T \Pi x^{(N)}(0) - x^{(N)}(T)^T \Pi x^{(N)}(T) + 2\int_0^T (x^{(N)}(t))^T \Pi D dW^{(N)}(t) + 2\int_0^T \tilde{x}_i(t)^T P D d\tilde{W}_i(t).
\end{aligned} \tag{A5}
$$

Recalling $u \in \mathcal{U}_d$ and Lemma 12.3 of [2], for $\forall \epsilon > 0$, we have

$$
\int_0^T (x^{(N)}(t))^T \Pi D dW^{(N)}(t) = O(T^{\frac{1}{2}+\epsilon}), \tag{A6}
$$

$$
\int_0^T \tilde{x}_i(t)^T P D d\tilde{W}_i(t) = O(T^{\frac{1}{2}+\epsilon}). \tag{A7}
$$

Then the follwing holds:

$$
\begin{aligned}
J_{\rm soc}^{(N)}(u) &= \limsup_{T\to\infty} \frac{1}{T} J_{\rm soc}^{(N)}(T,u) \\
&= \limsup_{T\to\infty} \frac{1}{T} \Bigg[\sum_{i=1}^{N} (\|\tilde{x}_i(0)\|_P^2 - \|\tilde{x}_i(T)\|_P^2 + \|x^{(N)}(0)\|_\Pi^2 - \|x^{(N)}(T)\|_\Pi^2) + (N-1)\int_0^T \|D\|_P^2 dt \\
&\quad + \int_0^T \|D\|_\Pi^2 dt + \sum_{i=1}^{N} \int_0^T \{ \|u_i(t) - u^{(N)}(t) + R^{-1}B^T P(x_i(t) - x^{(N)}(t))\|_R^2 + \|u^{(N)}(t) + R^{-1}B^T \Pi x^{(N)}(t)\|_R^2 \} dt \Bigg] \\
&\geqslant \limsup_{T\to\infty} \frac{1}{T} \Bigg[\sum_{i=1}^{N} (\|\tilde{x}_i(0)\|_P^2 - \|\tilde{x}_i(T)\|_P^2 + \|x^{(N)}(0)\|_\Pi^2 - \|x^{(N)}(T)\|_\Pi^2) + (N-1)\int_0^T \|D\|_P^2 dt + \int_0^T \|D\|_\Pi^2 dt \Bigg].
\end{aligned} \tag{A8}
$$

Thus we can obtain the optimal control law (5), and the corresponding social cost is given by

$$
J_{\rm soc}^{(N)}(\hat{u}) = \limsup_{T\to\infty} \frac{1}{T} \Bigg[\sum_{i=1}^{N} (\|\tilde{x}_i(0)\|_P^2 - \|\tilde{x}_i(T)\|_P^2 + \|x^{(N)}(0)\|_\Pi^2 - \|x^{(N)}(T)\|_\Pi^2) + (N-1)\int_0^T \|D\|_P^2 dt + \int_0^T \|D\|_\Pi^2 dt \Bigg].
$$

■

Appendix B The robust ADP algorithm and more details

Due to the appearance of random disturbance, we consider the following variant

$$
A^T K_\triangle + K_\triangle A - K_\triangle B R^{-1} B^T K_\triangle + Q + \triangle = 0 \tag{B1}
$$

where $\triangle$ represents a stochastic noise. We can solve it by the following robust ADP algorithm [4].

Algorithm B1 An offline robust VI algorithm

---

1. Choose $K_0 = K_0^T \geqslant 0$. $k, q \leftarrow 0$.
2. **Loop** $K_{k+1/2} \leftarrow K_k + h_k\big(A^\top K_k + K_k A - K_k B R^{-1} B^T K_k + Q + \triangle_k\big)$
3. **If** $|K_{k+1/2}| > q$ or $K_{k+1/2} \ngeqslant 0$ **then**
4. $K_{k+1} \leftarrow K_0$. $q \leftarrow q+1$.
5. **else**
6. $K_{k+1} \leftarrow K_{k+1/2}$
7. $k \leftarrow k+1$

In the above, where $\triangle_k$ is a stochastic process defined on a complete probability space, and $\sum_{k=0}^{\infty} h_k \triangle_k$ converges with probability 1. $\{h_k\}_{k=0}^{\infty}$ is a real sequence satisfying $h_k > 0, \lim_{k\to\infty} h_k = 0, \sum_{k=0}^{\infty} h_k < \infty$.

Applying Itô's formula, we have

$$
\begin{aligned}
d(\tilde{x}_i^T P \tilde{x}_i) &= 2\tilde{x}_i^T P(A\tilde{x}_i + B\tilde{u}_i)dt + \frac{N-1}{N} D^T P D dt + 2\tilde{x}_i^T P D d\tilde{W}_i \\
&= (\psi^i(z^i))^T \theta(P) dt - r_1(z) dt + 2\tilde{x}_i^T P D d\tilde{W}_i,
\end{aligned} \tag{B2}
$$

$$d\left((x^{(N)})^T\Pi x^{(N)}\right)=2(x^{(N)})^T\Pi\left((A+G)x^{(N)}+Bu^{(N)}\right)dt+\frac{1}{N}D^T\Pi Ddt+2(x^{(N)})^T\Pi DdW^{(N)}$$
$$=\phi^T(y)\alpha(\Pi)dt-r_2(y)dt+2(x^{(N)})^T\Pi DdW^{(N)}, \tag{B3}$$

where $z^i=[\tilde{x}_i^T,\tilde{u}_i^T,1]^T$, $y=[(x^{(N)})^T,(u^{(N)})^T,1]^T$, $\phi(y)=[y_1^2,2y_1y_2,\cdots,2y_1y_{n+m+1},y_2^2,\cdots,y_{n+m+1}^2]^T$, $\psi^i(z^i)=[(z_1^i)^2,2z_1^iz_2^i,\cdots,2z_1^iz_{n+m+1}^i,(z_2^i)^2,\cdots,(z_{n+m+1}^i)^2]^T$, and

$$\theta(P)=\text{vech}\left(\begin{bmatrix}PA+A^TP+Q & PB & 0\\ B^TP & R & 0\\ 0 & 0 & \frac{N-1}{N}D^TPD\end{bmatrix}\right),$$

$$\alpha(\Pi)=\text{vech}\left(\begin{bmatrix}\Pi(A+G)+(A+G)^T\Pi+Q-\Xi & \Pi B & 0\\ B^T\Pi & R & 0\\ 0 & 0 & \frac{1}{N}D^TPD\end{bmatrix}\right).$$

Multiplying both sides of (B2) by $\psi^i(z)$, we can get

$$\frac{1}{T}\int_0^T\psi^i(\psi^i)^Tdt\theta(P)=\frac{1}{T}\int_0^T\psi^id\left(\tilde{x}_i^TP\tilde{x}_i\right)+\frac{1}{T}\int_0^T\psi^ir_1dt-\frac{2}{T}\int_0^T\psi^i\tilde{x}_i^TPDd\tilde{W}_i. \tag{B4}$$

Multiplying both sides of (B3) by $\phi(y)$, we can obtain

$$\frac{1}{T}\int_0^T\phi\phi^Tdt\alpha(\Pi)=\frac{1}{T}\int_0^T\phi d\left((x^{(N)})^T\Pi x^{(N)}\right)+\frac{1}{T}\int_0^T\phi r_2dt-\frac{2}{T}\int_0^T\phi(x^{(N)})^T\Pi DdW^{(N)}. \tag{B5}$$

By [3, Theorem 6.1], we know

$$\lim_{t_k\to\infty}\frac{1}{t_k^2}\mathbb{E}\left[\left\|\int_0^{t_k}\psi^i\tilde{x}_i^TPDd\tilde{W}_i\right\|^2\right]=\lim_{t_k\to\infty}\frac{N-1}{N}\frac{1}{t_k^2}\mathbb{E}\left[\int_0^{t_k}\left\|\psi^i\tilde{x}_i^TPD\right\|^2dt\right]=0, \tag{B6}$$

and

$$\lim_{t_k\to\infty}\frac{1}{t_k^2}\mathbb{E}\left[\left\|\int_0^{t_k}\phi(x^{(N)})^T\Pi DdW^{(N)}\right\|^2\right]=\lim_{t_k\to\infty}\frac{1}{N}\frac{1}{t_k^2}\mathbb{E}\left[\int_0^{t_k}\left\|\phi(x^{(N)})^T\Pi D\right\|^2dt\right]=0. \tag{B7}$$

Then under (**A3**) and (B6), (B7), we denote

$$\hat{\theta}^i(P,t_k)=\Big(\int_0^{t_k}\psi^i(\psi^i)^Tds\Big)^{-1}\Big(\int_0^{t_k}\psi^id\left(\tilde{x}_i^TP_k\tilde{x}_i\right)+\int_0^{t_k}\psi^ir_1ds\Big), \tag{B8}$$

and

$$\hat{\alpha}(\Pi,t_k)=\Big(\int_0^{t_k}\phi\phi^Tds\Big)^{-1}\Big(\int_0^{t_k}\phi d\left((x^{(N)})^T\Pi_kx^{(N)}\right)+\int_0^{t_k}\phi r_2ds\Big), \tag{B9}$$

where the time sequence $\{t_0,t_1,...,t_k,...\}$ is increasing. $\hat{\theta}^i(P,t_k)$ and $\hat{\alpha}(\Pi,t_k)$ represent the values of $\theta(P)$ and $\alpha(\Pi)$ at time t_k, respectively.

Appendix C Lemma 1 and proof of Theorem 1

First we denote

$$\Delta_k(P)=\hat{\theta}^i(P,t_k)-\theta(P),\qquad \Delta_k(\Pi)=\hat{\alpha}(\Pi,t_k)-\alpha(\Pi).$$

Let's take $\Delta_k(\Pi)$ as an example to start the following discussion, and $\Delta_k(P)$ is similar.

Lemma C1. $\lim_{k\to\infty}\Delta_k(\Pi)=0,\quad a.s.$

Proof. Under (**A3**), we have

$$\Delta_k(\Pi)=2\Big(\int_0^{t_k}\phi\phi^Tds\Big)^{-1}\int_0^{t_k}\phi(x^{(N)})^T\Pi DdW^{(N)}$$
$$\leqslant 2(t_k\bar{\beta}I)^{-1}\int_0^{t_k}\phi(x^{(N)})^T\Pi DdW^{(N)}.$$

Recalling $u\in\mathcal{U}_{ad}$ and Lemma 12.3 of [2], for $\forall\epsilon>0$, we have

$$\int_0^{t_k}\phi(x^{(N)})^T\Pi DdW^{(N)}=O(t_k^{\frac{1}{2}+\epsilon}),$$

so $\Delta_k(\Pi)\to 0,\ a.s.\ k\to\infty$.

Now the updating equation for Π_k in Algorithm 1 is equivalent to

$$\Pi_{k+1} \leftarrow \Pi_k + h_k\left(\mathscr{R}(\Pi_k) + \Delta_k(\Pi_k)\right),$$

where $\mathscr{R}(\Pi) = (A+G)^T\Pi + \Pi(A+G) - \Pi B R^{-1} B^T \Pi + Q - \Xi$.

Proof of Theorem 1 Below we only prove $\lim_{k\to\infty} \Pi_k = \Pi^*$. First we construct a differential matrix Riccati equation

$$\dot{\Pi} = (A+G)^T\Pi + \Pi(A+G) - \Pi B R^{-1} B^T \Pi + Q - \Xi,$$

by [4, Proposition 3.6], we know Π is exponentially stable at Π^*, and have a smooth Lyapunov function [4, Lemma 3.3] $\mathcal{V}: R_A \to \mathcal{R}$, where $R_A \subset S^n$ represents the attraction of Π^*, such that for $\Pi \in R_A$

$$\langle \partial_x \mathcal{V}(\Pi), \mathscr{R}(\Pi)\rangle_F < 0,\ \left\langle \partial_x \mathcal{V}\left(\Pi^*\right),\ \mathscr{R}\left(\Pi^*\right)\right\rangle_F = 0, \mathcal{V}(\Pi^*) = 0.$$

We can find a sufficiently small constant $\eta > 0$, such that for all $\xi \in S^n$ satisfying $\|\xi\| < \eta$,

$$\langle \partial_x \mathcal{V}(\Pi), \mathscr{R}(\Pi) + \xi\rangle_F = -\iota, \tag{C1}$$

where $\iota > 0$. At the same time, $\{\Pi : \mathcal{V}(\Pi) < C\}$ is compact and a subset of R_A for all $C > 0$. Therefore, there exist $C_0 > 0$ and $C_1 > 0$ such that

$$C_0 < \mathcal{V}(\Pi_0) < C_1.$$

By contradiction, suppose $\{\Pi_k\}_{k=0}^{\infty}$ is unbounded. Then there would exist an uncrossing interval $[C_2, C_3]$ such that $\{\Pi_k\}_{k=0}^{\infty}$ crosses this interval infinitely many times, and $\mathcal{V}(\Pi_0) < C_2 < C_3 < C_1$. Obviously, there exist two subsequences $\left\{\Pi_{k_j}\right\}$, $\left\{\Pi_{k_j'}\right\}$, such that

$$\mathcal{V}(\Pi_{k_j}) < C_2 < \mathcal{V}(\Pi_m) < C_3 < \mathcal{V}(\Pi_{k_j'}), \quad k_j \leqslant m < k_j'.$$

Choosing a sufficiently small $\varepsilon > 0$, such that for any $\Pi \in \left\{\Pi_{k_j}\right\}$, we have

$$\varepsilon < \|\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\| = \left\|\sum_{i=k_j}^{L_\varepsilon(j)-1} h_i\Big(\mathscr{R}(\Pi_i) + \Delta_i(\Pi_i)\Big)\right\| \leqslant \gamma \sum_{i=k_j}^{L_\epsilon(j)-1} h_i, \tag{C2}$$

where $\gamma > 0$ is a constant and $L_\varepsilon(j) = inf\{i \geqslant k_j : \|\Pi_i - \Pi_{k_j}\| > \varepsilon\}$. Then we have

$$\begin{aligned}
\mathcal{V}\left(\Pi_{L_\varepsilon(j)}\right) - \mathcal{V}\left(\Pi_{k_j}\right) &= \int_0^1 \left\langle \partial_x \mathcal{V}\left(\Pi_{k_j} + t\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)\right), \left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)\right\rangle_F dt \\
&= \left\langle \partial_x \mathcal{V}\left(\Pi_{k_j}\right), \left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)\right\rangle_F \\
&\quad + \int_0^1\int_0^1 \left\langle \frac{d}{ds}\partial_x \mathcal{V}\left(\Pi_{k_j} + st\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)\right), t\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)^2\right\rangle_F dsdt \\
&= \sum_{i=k_j}^{L_\varepsilon(j)-1} h_i \left\langle \partial_x \mathcal{V}\left(\Pi_{k_j}\right), \mathscr{R}\left(\Pi_{k_j}\right) + \mathscr{R}_{\Delta_i} + \Delta_i(\Pi_i)\right\rangle_F \\
&\quad + \int_0^1\int_0^1 \left\langle \frac{d}{ds}\partial_x \mathcal{V}\left(\Pi_{k_j} + st\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)\right), t\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)^2\right\rangle_F dsdt,
\end{aligned}$$

where $\mathscr{R}_{\Delta_i} = \mathscr{R}(\Pi_i) - \mathscr{R}(\Pi_{k_j})$. Note that

$$\lim_{j\to\infty} \|\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\| = \varepsilon,$$

because $\lim_{k\to\infty} h_k = 0$, then we have

$$\lim_{j\to\infty}\left|\int_0^1\int_0^1 \left\langle \frac{d}{ds}\partial_x \mathcal{V}\left(\Pi_{k_j} + st\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)\right), t\left(\Pi_{L_\varepsilon(j)} - \Pi_{k_j}\right)^2\right\rangle_F dsdt\right| = O(\varepsilon^2).$$

By Lemma 1 and $\lim_{k\to\infty} h_k = 0$, we have $\lim_{j\to\infty}\sum_{i=k_j}^{L_\varepsilon(j)-1} h_i\Delta_i(\Pi_i) = 0$ a.s., there exists a sufficiently large $\bar{j}$, such that for all $j > \bar{j}$, by choosing sufficiently small ε, it follows that

$$\begin{aligned}
\mathcal{V}\left(\Pi_{L_\varepsilon(j)}\right) - \mathcal{V}\left(\Pi_{k_j}\right) &= \sum_{i=k_j}^{L_\varepsilon(j)-1} h_i \left\langle \partial_x \mathcal{V}\left(\Pi_{k_j}\right), \mathscr{R}\left(\Pi_{k_j}\right) + \mathscr{R}_{\Delta_i} + \Delta_i(\Pi_i)\right\rangle_F + O(\varepsilon^2) \\
&< \sum_{i=k_j}^{L_\varepsilon(j)-1} h_i \left\langle \partial_x \mathcal{V}(\Pi_{k_j}), \Delta_i(\Pi_i)\right\rangle_F - \frac{\iota\varepsilon}{\gamma} + O(\varepsilon^2) \\
&< 0.
\end{aligned}$$

which shows that $\{\Pi_k\}_{k=0}^{\infty}$ is bounded with probalility 1. ■

Appendix D An offline robust learning algorithm for the unseparable case

(**A4**) There exists $t_0 > 0$, $c > 0$, such that for all $1 \leqslant i \leqslant N$, $t > t_0$, the inequality

$$\frac{1}{t}\int_0^t \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} \begin{bmatrix} (\psi^i)^T & \phi^T \end{bmatrix} ds \geqslant cI$$

with probability 1.

Similarly, we first apply the Itô's formula

$$\begin{aligned} d\left(\tilde{x}_i^T P \tilde{x}_i + (x^{(N)})^T \Pi x^{(N)}\right) &= 2\tilde{x}_i^T P(A\tilde{x}_i + B\tilde{u}_i)dt + \frac{N-1}{N} D^T P D dt + 2\tilde{x}_i^T P D d\tilde{W}_i \\ &\quad + 2(x^{(N)})^T \Pi \left((A+G)x^{(N)} + Bu^{(N)}\right) dt + \frac{1}{N} D^T \Pi D dt + 2(x^{(N)})^T \Pi D dW^{(N)} \\ &= (\psi^i)^T (z^i)\theta(P)dt + \phi^T(y)\alpha(\Pi)dt - r(z,y)dt + 2\tilde{x}_i^T P D d\tilde{W}_i + 2(x^{(N)})^T \Pi D dW^{(N)} \\ &= \begin{bmatrix} (\psi^i)^T(z^i) & \phi^T(y) \end{bmatrix} \begin{bmatrix} \theta(P) \\ \alpha(\Pi) \end{bmatrix} dt - r(z,y)dt + 2\tilde{x}_i^T P D d\tilde{W}_i + 2(x^{(N)})^T \Pi D dW^{(N)}. \end{aligned} \tag{D1}$$

Multiplying both sides of (D1) by $\begin{bmatrix} \psi^i(z^i) \\ \phi(y) \end{bmatrix}$, we have

$$\begin{aligned} \frac{1}{T}\int_0^T \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} \begin{bmatrix} (\psi^i)^T & \phi^T \end{bmatrix} dt \begin{bmatrix} \theta(P) \\ \alpha(\Pi) \end{bmatrix} &= \frac{1}{T}\int_0^T \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} d\left(\tilde{x}_i^T P \tilde{x}_i + (x^{(N)})^T \Pi x^{(N)}\right) + \frac{1}{T}\int_0^T \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} r dt \\ &\quad - \frac{2}{T}\int_0^T \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} \left(\tilde{x}_i^T P D d\tilde{W}_i + (x^{(N)})^T \Pi D dW^{(N)}\right). \end{aligned}$$

Denote

$$\hat{\beta}_{\theta,\alpha}(P,\Pi,t_k) = \left(\int_0^{t_k} \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} \begin{bmatrix} (\psi^i)^T & \phi^T \end{bmatrix} ds\right)^{-1} \left(\int_0^{t_k} \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} d\left(\tilde{x}_i^T P_k \tilde{x}_i + (x^{(N)})^T \Pi_k x^{(N)}\right) + \int_0^{t_k} \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} r ds\right),$$

and

$$\mathcal{H}\left(\hat{\beta}_k\right) = \begin{bmatrix} \mathcal{T}\left(\hat{\theta}_k^i\right) \\ \Lambda\left(\hat{\alpha}_k\right) \end{bmatrix},$$

then we have the following Algorithm D1.

Appendix E Numerical example

This section shows the effectiveness of the online robust VI algorithm described in Algorithm 1. Consider a stochastic two-dimensional system with four agents. Take the parameter matrices in (1) as

$$A = \begin{bmatrix} -2 & -2 \\ 4 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}, \quad G = \begin{bmatrix} 0.2 & 0.3 \\ 0.1 & 0.2 \end{bmatrix}, \quad D = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

The parameter matrices in the cost function (2) are chosen as $Q = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$, $R = 1.25$, and $\Gamma = \begin{bmatrix} 0.2 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$. Then we can get $\Xi = \begin{bmatrix} 0.64 & 2.88 \\ 2.88 & 0.16 \end{bmatrix}$ in (4) . Set the initial state values of four agents to be $x_1(0) = [1,-1]^T$, $x_2(0) = [2,-4]^T$, $x_3(0) = [-3,2]^T$, $x_4(0) = [4,3]^T$, respectively. Set the initial value of P^i and Π to be

$$P_0^i = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \Pi_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

P_k^i and Π_k are updated in real time when we run the Algorithm 1. By (B8) and (B9), we can obtain 4 sequences $\left\{P_k^i\right\}_{k=0}^{\infty}$, $i = 1,2,3,4$ and 1 sequence $\{\Pi_k\}_{k=0}^{\infty}$, which eventually convergence to P^* and Π^* respectively. The converge matrices P^* and Π^* are shown below:

$$P^* = \begin{bmatrix} 2.2283 & -4.0928 \\ -4.0928 & 5.4287 \end{bmatrix}, \qquad \Pi^* = \begin{bmatrix} 2.0362 & -4.3194 \\ -4.3194 & 5.8537 \end{bmatrix}.$$

The trajectories of $\left\{P_k^i\right\}_{k=0}^{\infty}$ and $\{\Pi_k\}_{k=0}^{\infty}$ are given in Figure 1 and Figure 2 respectively.

Figure 3 shows the evolution of the first state components of agents $1, 2, 3$, and 4. Figure 4 shows the evolution of the second state components of agents $1, 2, 3$, and 4.

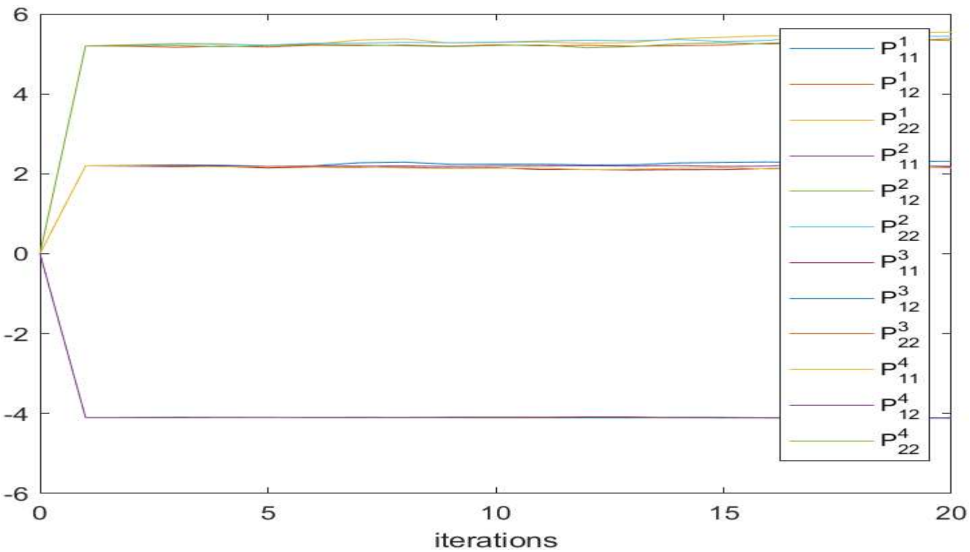

Figure E1 The iteration trajectories of $\{P_k\}_{k=0}^{\infty}$, $i = 1, 2, 3, 4$.

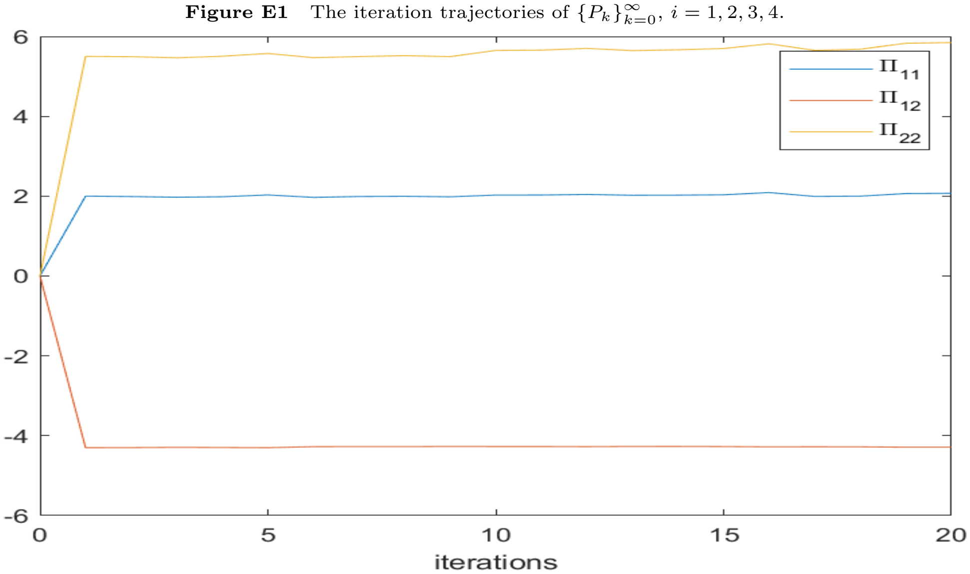

Figure E2 The iteration trajectories of $\{\Pi_k\}_{k=0}^{\infty}$

Algorithm D1 An online robust learning algorithm for mean field LQ social control: unseparable case

1. Initialize $P_0^i = (P_0^i)^T \geqslant 0$, $\Pi_0 = \Pi_0^T \geqslant 0$ $k, q \leftarrow 0$.
2. **Loop**

$$\hat{\beta}_k^i \leftarrow \left(\int_0^{t_k} \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} \begin{bmatrix} (\psi^i)^T & \phi^T \end{bmatrix} ds \right)^{-1} \left(\int_0^{t_k} \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} d\left(\tilde{x}_i^T P_k^i \tilde{x}_i + (x^{(N)})^T \Pi_k x^{(N)} \right) + \int_0^{t_k} \begin{bmatrix} \psi^i \\ \phi \end{bmatrix} r ds \right)$$

$$\begin{bmatrix} P_{k+1/2}^i \\ \Pi_{k+1/2} \end{bmatrix} \leftarrow \begin{bmatrix} P_k^i \\ \Pi_k \end{bmatrix} + h_k \mathcal{H}\left(\hat{\beta}_k^i\right)$$

3. **if** $\begin{bmatrix} P_{k+1/2}^i \\ \Pi_{k+1/2} \end{bmatrix} > 0$ and $(|P_{k+1/2}^i - P_k^i| + |\Pi_{k+1/2} - \Pi_k|)/h_k < \bar{\varepsilon}$ **then**

return $\begin{bmatrix} P_k^i \\ \Pi_k \end{bmatrix}$ as an approximation to $\begin{bmatrix} P^* \\ \Pi^* \end{bmatrix}$

else if $\left\| \begin{bmatrix} P_{k+1/2}^i \\ \Pi_{k+1/2} \end{bmatrix} \right\| > q$ or $\begin{bmatrix} P_{k+1/2}^i \\ \Pi_{k+1/2} \end{bmatrix} \leqslant 0$ **then**

$\begin{bmatrix} P_{k+1}^i \\ \Pi_{k+1} \end{bmatrix} \leftarrow \begin{bmatrix} P_0^i \\ \Pi_0 \end{bmatrix}$, $q \leftarrow q+1$.

else $\begin{bmatrix} P_{k+1}^i \\ \Pi_{k+1} \end{bmatrix} \leftarrow \begin{bmatrix} P_{k+1/2}^i \\ \Pi_{k+1/2} \end{bmatrix}$, $k \leftarrow k+1$

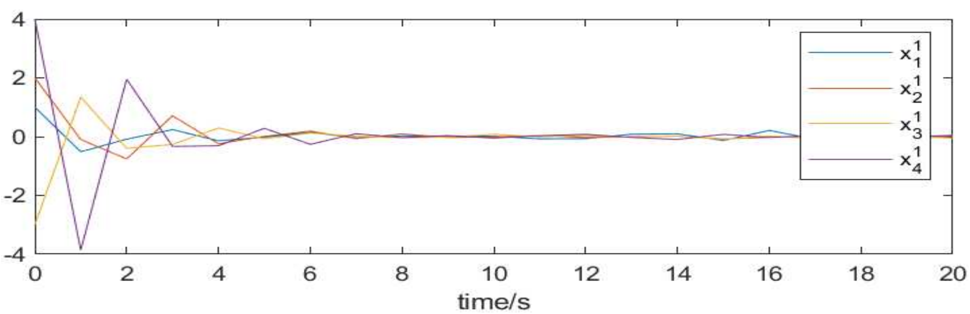

Figure E3 Evolution of the first state components of agents 1, 2, 3, and 4

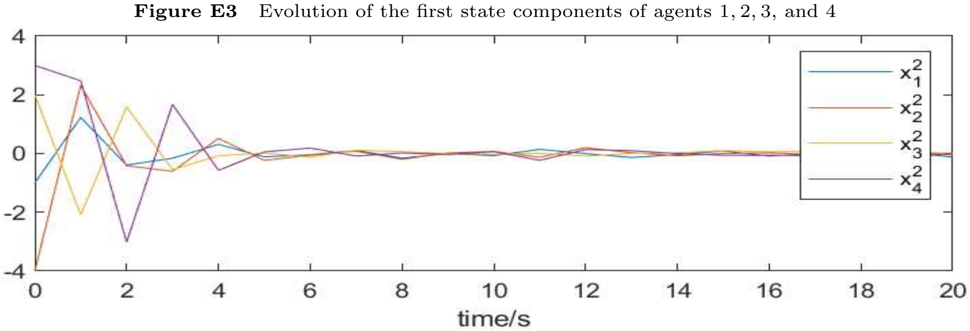

Figure E4 Evolution of the second state components of agents 1, 2, 3, and 4

References

1 Anderson B D O, Moore J B, Molinari B P. Linear optimal control. Englewood Cliffs, NJ : Prentice-Hall, 1971.
2 Chen H F, Guo L. Identification and stochastic adaptive control. Boston, MA: Birkhäuser, 1991.
3 Steele J M. Stochastic calculus and financial applications. New York: Springer, 2001.
4 Bian T, Jiang Z P. Continuous-time robust dynamic programming. SIAM J Control Optim, 2019, 57: 4150-4174.