

## Supplementary material: Exploiting inter-agent coupling information for efficient model-free reinforcement learning of cooperative LQR

### Appendix A. Proof of Lemma 3.1

**Proof** We prove a stronger version of the lemma that holds irrespective of the linear dynamics and quadratic cost assumption. For some  $i, j \in \mathcal{V}$ , let  $j \in \mathcal{I}_Q^i$ . For the sake of contradiction, assume that  $\exists$  a  $k \in \mathcal{R}_{SO}^j$  such that  $k \notin \mathcal{I}_Q^i$ . By the definition of  $\mathcal{I}_Q^i$ ,  $j \in \mathcal{I}_Q^i$  implies that for some some  $t' \geq t$ ,  $\exists$  a function (or composition of functions)  $f : \mathcal{S} \times \mathcal{U} \rightarrow \mathbb{R}$  such that

$$c_i(x_{\mathcal{I}_C^i}(t'), u_{\mathcal{I}_C^i}(t')) = f(x_j(t), u_j(t), \bigcup_{g \in \mathcal{I}_Q^i \setminus j} \{x_g(\cdot), u_g(\cdot)\}). \quad (14)$$

Recall that the control  $u_j(t) \in \mathcal{U}$  depends only on its partial observation  $o_j(t)$ , current state  $x_j(t)$ , and local policy  $\pi_j(\cdot)$ . Therefore,  $\exists$  a function  $g_j : \mathcal{Z}_j \rightarrow P(\mathcal{U}_j)$  such that

$$u_j(t) \sim g_j(o_j(t)) = g_j(\{x_m(t)\}_{m \in \mathcal{I}_O^j}) \quad (15)$$

Similarly, due to the Markovian assumption for each  $x_j(t)$ ,  $\exists$  a mapping  $h_j : \prod_{n \in \mathcal{I}_S^j} \mathcal{S}_n \times \prod_{n \in \mathcal{I}_S^j} \mathcal{U}_n \rightarrow P(\mathcal{S}_j)$  such that

$$x_j(t) \sim h_j(\{x_n(t-1)\}_{n \in \mathcal{I}_S^j}, \{u_n(t-1)\}_{n \in \mathcal{I}_S^j}). \quad (16)$$

Using (15) and (16), (14) can be rewritten as

$$c_i(x_{\mathcal{I}_C^i}(t'), u_{\mathcal{I}_C^i}(t')) = f(x_j(t), u_j(t), \bigcup_{g \in \mathcal{I}_Q^i \setminus j} x_g, u_g) \quad (17)$$

$$= f(h_j(\{x_n(t-1)\}_{n \in \mathcal{I}_S^j}, \{u_n(t-1)\}_{n \in \mathcal{I}_S^j}), g_j(\{x_m(t)\}_{m \in \mathcal{I}_O^j}), \bigcup_{g \in \mathcal{I}_Q^i \setminus j} \{x_g(\cdot), u_g(\cdot)\}) \quad (18)$$

$$= f(h_j(\{x_n(t-1), u_n(t-1)\}_{n \in \mathcal{I}_S^j}), g_j(\{\{x_l(t-1), u_l(t-1)\}_{l \in \mathcal{I}_S^m}\}_{m \in \mathcal{I}_O^j}), \bigcup_{g \in \mathcal{I}_Q^i \setminus j} \{x_g(\cdot), u_g(\cdot)\}). \quad (19)$$

On recursive expansion of (19), it is straightforward to verify that  $c_i(x_{\mathcal{I}_C^i}(t'), u_{\mathcal{I}_C^i}(t'))$  depends on  $\{x_s(t''), u_s(t'')\}_{s \in \mathcal{R}_{SO}^j}$ , for some  $t'' \leq t \leq t'$ . Thus,  $i \in \mathcal{I}_{GD}^s \forall s \in \mathcal{R}_{SO}^j$  which implies that  $s \in \mathcal{I}_Q^i \forall s \in \mathcal{R}_{SO}^j$ . But as  $k \in \mathcal{R}_{SO}^j$ ,  $k \in \mathcal{I}_Q^i$  which is a contradiction. Therefore, our assumption is false and hence if  $j \in \mathcal{I}_Q^i$ , then  $\forall k \in \mathcal{R}_{SO}^j$ ,  $k \in \mathcal{I}_Q^i$  as required. ■

### Appendix B. Proof of Theorem 3.1

**Proof** For the networked system, observe that the individual cost-to-go for each agent  $Q_i$  is dependent on the global state and control due to the long-term inter-agent dependencies between the

agents. Recall that

$$Q_i(x, u) = c_i(x_{\mathcal{I}_C^i}, u_{\mathcal{I}_C^i}) + \mathbb{E} \left[ \sum_{t=1}^T c_i(x_{\mathcal{I}_C^i}(t), u_{\mathcal{I}_C^i}(t)) \right]. \quad (20)$$

For LTI dynamics (1) and quadratic cost (2), (20) can be rewritten as

$$\begin{aligned} Q_i(x, u) &= \begin{bmatrix} x_{\mathcal{I}_C^i}(t) \\ u_{\mathcal{I}_C^i}(t) \end{bmatrix}^\top \begin{bmatrix} S_i & 0 \\ 0 & R_i \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_C^i}(t) \\ u_{\mathcal{I}_C^i}(t) \end{bmatrix} + \mathbb{E}_{w(t), \eta(t)} \left[ \begin{bmatrix} x_{\mathcal{I}_C^i}(t+1) \\ u_{\mathcal{I}_C^i}(t+1) \end{bmatrix}^\top \begin{bmatrix} S_i & 0 \\ 0 & R_i \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_C^i}(t+1) \\ u_{\mathcal{I}_C^i}(t+1) \end{bmatrix} \right. \\ &\quad \left. + \mathbb{E}_{w(t+1), \eta(t+1)} \left[ \begin{bmatrix} x_{\mathcal{I}_C^i}(t+2) \\ u_{\mathcal{I}_C^i}(t+2) \end{bmatrix}^\top \begin{bmatrix} S_i & 0 \\ 0 & R_i \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_C^i}(t+2) \\ u_{\mathcal{I}_C^i}(t+2) \end{bmatrix} + \mathbb{E}[\dots] \right] = \\ &\quad \sum_{j, k \in \mathcal{I}_C^i} \left[ (x_j(t))^\top S_{jk} x_k(t) + (u_j(t))^\top R_{jk} u_k(t) + [\sigma_w^2 \text{Tr}(S_i) + \sigma_\eta^2 \text{Tr}(R_i)]_{j \in \mathcal{I}_C^i} + \right. \\ &\quad \left. \left[ x_{\mathcal{I}_S^j}^\top(t) A_j^\top S_i A_j x_{\mathcal{I}_S^j}(t) + u_{\mathcal{I}_S^j}^\top(t) B_j^\top S_i B_j u_{\mathcal{I}_S^j}(t) + 2x_{\mathcal{I}_S^j}^\top(t) A_j^\top S_i B_j u_{\mathcal{I}_S^j}(t) + x_{\mathcal{I}_O^j}^\top(t) K_j^\top R_i K_j x_{\mathcal{I}_O^j}(t) \right]_{j \in \mathcal{I}_C^i} \right. \\ &\quad \left. + \sigma_\eta^2 \text{Tr} \left( B_j^\top S_i B_j \mathbb{I}_{n_u|\mathcal{I}_S^j|} \right) + 2\text{Tr} \left( A_j^\top S_i B_j w_k(t) \eta_l^\top(t) \right)_{\substack{k \in \mathcal{I}_S^j \\ l \in \mathcal{I}_O^j}} + \sigma_w^2 \text{Tr} \left( A_j^\top S_i A_j \mathbb{I}_{n_x|\mathcal{I}_S^j|} \right) + \dots \right] \end{aligned} \quad (21)$$

Therefore, from (21), it is clear that for time-invariant inter-agent couplings, the  $Q_i(\cdot)$  for each  $i \in \mathcal{V}$  depends on its neighbors in the cost graph which in turn depend on their neighbors in the state, and observation graphs and so on. In other words,  $\forall i \in \mathcal{V}$ ,  $Q_i(\cdot)$  depends on a subset of agents  $\mathcal{I}_Q^i := \{\mathcal{I}_C^i \cup \{\mathcal{R}_{SO}^k\}_{k \in \mathcal{I}_C^i}\} = \{\mathcal{R}_{SO}^k\}_{k \in \mathcal{I}_C^i}$ . By Lemma 3.1, we have that  $\mathcal{I}_Q^i$  is closed under  $\mathcal{R}_{SO}$  which implies that the information of agents in  $\mathcal{I}_Q^i$  is sufficient to exactly compute the the future costs of agent  $i$ . Thus, it follows that  $Q_i(x(t), u(t)) = Q_i(x_{\mathcal{I}_Q^i}(t), u_{\mathcal{I}_Q^i}(t))$  as required. ■

## Appendix C. Proof of Theorem 3.2

**Proof** Recall that

$$\begin{aligned} Q(x, u) &= \mathbb{E}_\pi \left[ \sum_{i=1}^N \sum_{t=0}^{\infty} c_i(x_{\mathcal{I}_C^i}(t), u_{\mathcal{I}_C^i}(t)) | x(0)=x, u(0)=u \right] \\ &= \mathbb{E}_\pi \left[ \sum_{j \in \mathcal{I}_{\text{GD}}^i} \sum_{t=0}^{\infty} c_j(x_{\mathcal{I}_C^j}(t), u_{\mathcal{I}_C^j}(t)) | x(0)=x, u(0)=u \right] \\ &\quad + \mathbb{E}_\pi \left[ \sum_{j \setminus \mathcal{I}_{\text{GD}}^i} \sum_{t=0}^{\infty} c_j(x_{\mathcal{I}_C^j}(t), u_{\mathcal{I}_C^j}(t)) | x(0)=x, u(0)=u \right] \\ &= \sum_{j \in \mathcal{I}_{\text{GD}}^i} Q_j(x_{\mathcal{I}_Q^j}, u_{\mathcal{I}_Q^j}) + \sum_{k \setminus \mathcal{I}_{\text{GD}}^i} Q_k(x_{\mathcal{I}_Q^k}, u_{\mathcal{I}_Q^k}) = \hat{Q}_i(x_{\mathcal{I}_Q^i}, u_{\mathcal{I}_Q^i}) + \bar{Q}_i(x_{\mathcal{I}_Q^i}, u_{\mathcal{I}_Q^i}), \end{aligned} \quad (22)$$

where  $\bar{Q}_i(x_{\mathcal{I}_Q^i}, u_{\mathcal{I}_Q^i}) = Q(x, u) - \hat{Q}_i(x_{\mathcal{I}_Q^j}, u_{\mathcal{I}_Q^j}) = \sum_{k \in \mathcal{I}_{\text{GD}}^i} Q_k(x_{\mathcal{I}_Q^k}, u_{\mathcal{I}_Q^k})$ . From Theorem 3.1, the reward of each agent  $i \in \mathcal{V}$  depends on  $x_j(t)$ ,  $u_j(t) \forall j \in \mathcal{I}_Q^i$  and  $\mathcal{E}_{\text{GD}} = \mathcal{E}_Q^{\text{T}}$  by definition of  $\mathcal{G}_{\text{GD}}$ . Therefore, if  $j \notin \mathcal{I}_{\text{GD}}^i$ , then  $i \notin \mathcal{I}_Q^j$ . Hence,  $\sum_{j \in \mathcal{I}_{\text{GD}}^i} c_j(x_{\mathcal{I}_C^j}(t), u_{\mathcal{I}_C^j}(t))$  is independent of  $u_i(t)$  and thus  $K_i$ . It then follows that  $Q_j(\cdot)$  is independent of  $K_i, \forall j \notin \mathcal{I}_{\text{GD}}^i$ , which implies

$$\begin{aligned} \nabla_{K_i} \bar{Q}_i &= \nabla_{K_i} \mathbb{E}_{\pi} \left[ \sum_{j \in \mathcal{I}_{\text{GD}}^i} \sum_{t=0}^{\infty} c_j(x_{\mathcal{I}_C^j}(t), u_{\mathcal{I}_C^j}(t)) | x(0)=x, u(0)=u \right] \\ &\stackrel{(a)}{=} \mathbb{E}_{\pi} \left[ \nabla_{K_i} \sum_{j \in \mathcal{I}_{\text{GD}}^i} \sum_{t=0}^{\infty} c_j(x_{\mathcal{I}_C^j}(t), u_{\mathcal{I}_C^j}(t)) | x(0)=x, u(0)=u \right] = 0, \end{aligned} \quad (23)$$

where (a) in (23) is obtained by interchanging the derivative and integral assuming that each  $Q_j(\cdot)$  is sufficiently smooth in state and control. Hence, the gradient of the global action value function with respect to  $K_i$  is given by  $\nabla_{K_i} Q(s, a) = \nabla_{K_i} [\hat{Q}_i + \bar{Q}_i] = \nabla_{K_i} \hat{Q}_i$ , as required.  $\blacksquare$

#### Appendix D. Proof of Proposition 4.1

**Proof** From (9), we have

$$\hat{Q}_i(x_{\mathcal{I}_Q^i}, u_{\mathcal{I}_Q^i}) = \begin{bmatrix} x_{\mathcal{I}_Q^i}(t) \\ u_{\mathcal{I}_Q^i}(t) \end{bmatrix} \begin{bmatrix} S_{\mathcal{I}_Q^i} & 0 \\ 0 & R_{\mathcal{I}_Q^i} \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_Q^i}(t) \\ u_{\mathcal{I}_Q^i}(t) \end{bmatrix} + \mathbb{E} \left[ \hat{Q}_i(x_{\mathcal{I}_Q^i}(t+1), u_{\mathcal{I}_Q^i}(t+1)) \right]. \quad (24)$$

Then, the expected future Q-value can be rewritten as

$$\begin{aligned} &\mathbb{E} \left[ \hat{Q}_i(x_{\mathcal{I}_Q^i}(t+1), u_{\mathcal{I}_Q^i}(t+1)) \right] \\ &= \mathbb{E} \left[ \begin{bmatrix} x_{\mathcal{I}_Q^i}(t+1) \\ u_{\mathcal{I}_Q^i}(t+1) \end{bmatrix} \begin{bmatrix} S_{\mathcal{I}_Q^i} & 0 \\ 0 & R_{\mathcal{I}_Q^i} \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_Q^i}(t+1) \\ u_{\mathcal{I}_Q^i}(t+1) \end{bmatrix} \right] + \mathbb{E} \left[ \mathbb{E} \left[ \hat{Q}_i(x_{\mathcal{I}_Q^i}(t+2), u_{\mathcal{I}_Q^i}(t+2)) \right] \right] \\ &= \mathbb{E} \left[ (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t) + w_{\mathcal{I}_Q^i}(t))^{\text{T}} S_{\mathcal{I}_Q^i} (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t) + w_{\mathcal{I}_Q^i}(t)) \right] + \\ &\mathbb{E} \left[ (K_{\mathcal{I}_Q^i} (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t) + w_{\mathcal{I}_Q^i}(t)))^{\text{T}} R_{\mathcal{I}_Q^i} (K_{\mathcal{I}_Q^i} (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t) + w_{\mathcal{I}_Q^i}(t))) \right] \\ &+ \mathbb{E} \left[ \mathbb{E} \left[ \hat{Q}_i(x_{\mathcal{I}_Q^i}(t+2), u_{\mathcal{I}_Q^i}(t+2)) \right] \right] \\ &= (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t))^{\text{T}} S_{\mathcal{I}_Q^i} (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t)) + \sigma_w^2 \text{Tr} \left( S_{\mathcal{I}_Q^i} + K_{\mathcal{I}_Q^i}^{\text{T}} R_{\mathcal{I}_Q^i} K_{\mathcal{I}_Q^i} \right) \\ &+ (K_{\mathcal{I}_Q^i} (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t)))^{\text{T}} R_{\mathcal{I}_Q^i} (K_{\mathcal{I}_Q^i} (A_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) + B_{\mathcal{I}_Q^i} u_{\mathcal{I}_Q^i}(t))) \\ &+ \mathbb{E} \left[ \mathbb{E} \left[ \hat{Q}_i(x_{\mathcal{I}_Q^i}(t+2), u_{\mathcal{I}_Q^i}(t+2)) \right] \right] \end{aligned} \quad (25)$$

$$\begin{aligned}
&= \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} \begin{bmatrix} A_{\mathcal{I}_{\hat{Q}}}^{\top} \\ B_{\mathcal{I}_{\hat{Q}}}^{\top} \end{bmatrix} (S_{\mathcal{I}_{\hat{Q}}}^i + K_{\mathcal{I}_{\hat{Q}}}^{\top} R_{\mathcal{I}_{\hat{Q}}} K_{\mathcal{I}_{\hat{Q}}}^i) \begin{bmatrix} A_{\mathcal{I}_{\hat{Q}}}^i & B_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} + \sigma_w^2 \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix}^{\top} \begin{bmatrix} S_{\mathcal{I}_{\hat{Q}}}^i & 0 \\ 0 & R_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix} \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix} \\
&+ \mathbb{E} \left[ \mathbb{E} \left[ \hat{Q}_i(x_{\mathcal{I}_{\hat{Q}}}^i(t+2), u_{\mathcal{I}_{\hat{Q}}}^i(t+2)) \right] \right]. \tag{26}
\end{aligned}$$

Recursive expansion of (26) yields

$$\hat{Q}_i(x_{\mathcal{I}_{\hat{Q}}}^i, u_{\mathcal{I}_{\hat{Q}}}^i) = \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} \hat{Q}_i \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} + \sigma_w^2 \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix}^{\top} \hat{Q}_i \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix}, \tag{27}$$

where with a slight abuse of notation

$$\hat{Q}_i = \begin{bmatrix} S_{\mathcal{I}_{\hat{Q}}}^i & 0 \\ 0 & R_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix} + \begin{bmatrix} A_{\mathcal{I}_{\hat{Q}}}^{\top} \\ B_{\mathcal{I}_{\hat{Q}}}^{\top} \end{bmatrix} \mathcal{L} \left( A_{\mathcal{I}_{\hat{Q}}}^i + B_{\mathcal{I}_{\hat{Q}}}^i K_{\mathcal{I}_{\hat{Q}}}^i, S_{\mathcal{I}_{\hat{Q}}}^i + K_{\mathcal{I}_{\hat{Q}}}^{\top} R_{\mathcal{I}_{\hat{Q}}} K_{\mathcal{I}_{\hat{Q}}}^i \right) \begin{bmatrix} A_{\mathcal{I}_{\hat{Q}}}^i & B_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix},$$

and  $\mathcal{L}(X, Y)$  is the analytical solution of the discrete time Lyapunov equation  $\mathcal{P} = X\mathcal{P}X^{\top} + Y$ . ■

## Appendix E. Proof of Lemma 5.1

**Proof** Define  $\mathcal{R}_{(SO)\tau}^i = \{j \in \mathcal{V} | j \xrightarrow{\mathcal{E}_{SO}^{\top}} i\}$ , and  $\mathcal{I}_{C\tau}^i = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}_O^{\top}\}$ .

- (a) **Necessary condition.** Assume that  $\mathcal{I}_{\hat{Q}}^i \subset \mathcal{V}$ . Then there exists a  $k \in \mathcal{V}$  such that  $k \notin \bigcup_{j \in \mathcal{I}_{\text{GD}}^i} \mathcal{I}_Q^j$  i.e.,  $k \notin \mathcal{I}_Q^j, \forall j \in \mathcal{I}_{\text{GD}}^i$ . This implies that  $\forall j \in \mathcal{I}_{\text{GD}}^i$ , we have  $k \notin \mathcal{I}_C^j$  and  $k \notin \{\mathcal{R}_{SO}^p\}_{p \in \mathcal{I}_C^j}$ . Similarly, as  $j \in \mathcal{I}_{\text{GD}}^i$  implies  $i \in \mathcal{I}_Q^j$ , we have that either  $i \in \mathcal{I}_C^j$  or  $i \in \{\mathcal{R}_{SO}^q\}_{q \in \mathcal{I}_C^j}$ .

Consider the case where  $i \in \{\mathcal{R}_{SO}^q\}_{q \in \mathcal{I}_C^j}$ . Suppose that there exists an  $r \in \mathcal{I}_C^j$  for which  $i \in \mathcal{R}_{SO}^r$ . Then as  $k \notin \mathcal{I}_C^j$  and  $k \notin \{\mathcal{R}_{SO}^q\}_{q \in \mathcal{I}_C^j}$ , we have  $\forall m \in \mathcal{R}_{(SO)\tau}^i$  and  $\forall p \in \mathcal{R}_{(SO)\tau}^k$ ,  $\mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p = \emptyset$ . This is because otherwise for every  $l \in \mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p$ , we obtain  $l \in \mathcal{I}_{\text{GD}}^i$  and  $k \in \mathcal{I}_Q^l$ , implying  $k \in \mathcal{I}_Q^i$ , which contradicts our assumption.

Alternatively, if  $i \in \mathcal{I}_C^j, k \notin \mathcal{I}_C^j$  and  $k \notin \{\mathcal{R}_{SO}^q\}_{q \in \mathcal{I}_C^j}$  imply that  $\forall p \in \mathcal{R}_{(SO)\tau}^k, \mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^p = \emptyset$ . Otherwise for every  $l \in \mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^p$ , we obtain  $l \in \mathcal{I}_{\text{GD}}^i$ , and  $k \in \mathcal{I}_Q^l$  implying  $k \in \mathcal{I}_Q^i$  which contradicts our assumption.

As  $i \in \mathcal{R}_{(SO)\tau}^i$ , we have that  $\mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^p = \emptyset$  whenever  $\mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p = \emptyset$ . Therefore, we conclude that if  $\mathcal{I}_{\hat{Q}}^i \subset \mathcal{V}$ , then  $\forall m \in \mathcal{R}_{(SO)\tau}^i, \forall p \in \mathcal{R}_{(SO)\tau}^k, \mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p = \emptyset$ .

**Sufficient condition.**

Consider an  $i \in \mathcal{V}$  and assume that there exists a  $k \in \mathcal{V}$  such that  $\forall m \in \mathcal{R}_{(SO)\tau}^i, \forall p \in \mathcal{R}_{(SO)\tau}^k, \mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p = \emptyset$ . Consider  $s \in \mathcal{I}_{\text{GD}}^i$ , which means  $i \in \mathcal{I}_Q^s$ . It then follows that either  $i \in \mathcal{I}_C^s$  or  $i \in \{\mathcal{R}_{SO}^n\}_{n \in \mathcal{I}_C^s}$ .

If  $i \in \mathcal{I}_C^s$ , then as  $i \in \mathcal{R}_{(SO)\tau}^i$ , we have that  $\forall p \in \mathcal{R}_{(SO)\tau}^k$ ,  $\mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^p = \emptyset$ , which results in  $p \notin \mathcal{I}_C^s$ . This is because otherwise  $s \in \mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^p$ . Also, as  $k \in \mathcal{R}_{(SO)\tau}^k$ , we have  $k \notin \mathcal{I}_C^s$ .

For any  $n \in \mathcal{I}_C^s$  such that  $i \in \mathcal{R}_{SO}^n$ , it follows that  $n \in \mathcal{R}_{(SO)\tau}^i$ . Therefore,  $\forall p \in \mathcal{R}_{(SO)\tau}^k$ ,  $\mathcal{I}_{C\tau}^n \cap \mathcal{I}_{C\tau}^p = \emptyset$ , which means  $k \notin \mathcal{R}_{SO}^n$  for any  $n \in \mathcal{I}_C^s$  such that  $i \in \mathcal{R}_{SO}^n$ . Let  $n_1, n_2 \in \mathcal{I}_C^s$ , where  $n_1 \neq n_2$  such that  $i \in \mathcal{R}_{SO}^{n_1}$  but  $i \notin \mathcal{R}_{SO}^{n_2}$ . Then, as  $n_1 \in \mathcal{R}_{(SO)\tau}^i$ , and  $n_1, n_2 \in \mathcal{I}_C^s$ , we have  $k \notin \mathcal{R}_{SO}^{n_2}$ . This is because otherwise  $s \in \mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p$  for  $m = n_1$  and  $p = n_2$ , which contradicts our assumption. Therefore, we conclude that  $\forall n \in \mathcal{I}_C^s$ ,  $k \notin \mathcal{R}_{SO}^n$ .

It follows from  $k \notin \mathcal{I}_C^s$  and  $k \notin \{\mathcal{R}_{SO}^n\}_{n \in \mathcal{I}_C^s}$  that  $k \notin \mathcal{I}_Q^s \forall s \in \mathcal{I}_{GD}^i$ , i.e.,  $k \notin \mathcal{I}_Q^i$ . As  $k \in \mathcal{V} \setminus \mathcal{I}_Q^i$ ,  $\mathcal{V} \setminus \mathcal{I}_Q^i$  is non-empty, i.e.,  $\mathcal{I}_Q^i \subset \mathcal{V}$ .

- (b) **Necessary condition.** Consider an  $i \in \mathcal{V}$  and assume that there exists a  $j \in \mathcal{I}_{GD}^i$ , such that  $\mathcal{I}_Q^j \subset \mathcal{I}_Q^i$ . This implies that  $\exists k \in \mathcal{I}_Q^i$  such that  $k \notin \mathcal{I}_Q^j$ , and  $k \in \bigcup_{h \in \mathcal{I}_{GD}^i \setminus \{j\}} \mathcal{I}_Q^h$ . If  $k \notin \mathcal{I}_Q^j$ , then by definition,  $k \notin \mathcal{I}_C^j$ , and  $k \notin \{\mathcal{R}_{SO}^l\}_{l \in \mathcal{I}_C^j}$ . But,  $k \in \bigcup_{h \in \mathcal{I}_{GD}^i \setminus \{j\}} \mathcal{I}_Q^h$  implies that  $\exists h \in \mathcal{I}_{GD}^i \setminus \{j\}$  such that either  $k \in \mathcal{I}_C^h$  or  $k \in \{\mathcal{R}_{SO}^m\}_{m \in \mathcal{I}_C^h}$ .

**Case 1** Let  $k \in \mathcal{I}_C^h$ . Then, as  $h \in \mathcal{I}_{GD}^i$ , either  $i \in \mathcal{I}_C^h$ , or  $i \in \{\mathcal{R}_{SO}^l\}_{l \in \mathcal{I}_C^h}$ .

- If  $i \in \mathcal{I}_C^h$ , then  $\mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^k = \{h\} \neq \emptyset$ . or,
- If  $i \in \{\mathcal{R}_{SO}^l\}_{l \in \mathcal{I}_C^h}$ , then  $\exists$  an  $m \in \mathcal{I}_C^h \cap \mathcal{R}_{(SO)\tau}^i$ . Hence,  $\mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^k = \{h\} \neq \emptyset$ .

**Case 2** Let  $k \in \{\mathcal{R}_{SO}^m\}_{m \in \mathcal{I}_C^h}$ . Then,  $\exists$  an  $p \in \mathcal{I}_C^h \cap \mathcal{R}_{(SO)\tau}^k$ , and as  $h \in \mathcal{I}_{GD}^i$ , either  $i \in \mathcal{I}_C^h$ , or  $i \in \{\mathcal{R}_{SO}^l\}_{l \in \mathcal{I}_C^h}$ .

- If  $i \in \mathcal{I}_C^h$ , then  $\mathcal{I}_{C\tau}^i \cap \mathcal{I}_{C\tau}^p = \{h\} \neq \emptyset$ . or,
- If  $i \in \{\mathcal{R}_{SO}^l\}_{l \in \mathcal{I}_C^h}$ , then  $\exists$  an  $m \in \mathcal{I}_C^h \cap \mathcal{R}_{(SO)\tau}^i$ . Hence,  $\mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p = \{h\} \neq \emptyset$ .

Therefore, in either case we conclude that if  $\mathcal{I}_Q^i \subset \mathcal{I}_Q^j$ , then  $p \in \mathcal{R}_{(SO)\tau}^k$ ,  $m \in \mathcal{R}_{(SO)\tau}^i$ , such that  $\mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p \subset \mathcal{I}_{GD}^i$ .

**Sufficient condition.** Consider an  $i \in \mathcal{V}$  and assume that  $\exists j \in \mathcal{I}_{GD}^i$  for which  $\exists k \in \mathcal{V} \setminus \mathcal{I}_Q^j$ . Let  $h \in \mathcal{I}_{GD}^i$ ,  $m \in \mathcal{R}_{(SO)\tau}^i$ , and  $p \in \mathcal{R}_{(SO)\tau}^k$ , such that  $h \in \mathcal{I}_{C\tau}^m \cap \mathcal{I}_{C\tau}^p$ . Hence, as  $p \in \mathcal{I}_C^h$ , by definition  $k \in \mathcal{I}_Q^h$ . As  $h \in \mathcal{I}_{GD}^i$ , we have that  $k \in \mathcal{I}_Q^i$ . However,  $k \notin \mathcal{I}_Q^j$  implies that  $k \in \mathcal{I}_Q^i \setminus \mathcal{I}_Q^j$  or  $\mathcal{I}_Q^i \subset \mathcal{I}_Q^j$  as required. ■

## Appendix F. Proof of Theorem 5.1

**Proof** For the analysis of the direct case, we first show that for each  $i \in \mathcal{V}$ ,  $\|\hat{q}_i^{\text{true}} - \hat{q}_i^{\text{direct}}\|$  is analogous to Lemma A.1 Krauth et al. (2019) in the single-agent case. For brevity, in the remainder

of the proof we denote  $\hat{q}_i^{\text{direct}}$  by  $\hat{q}_i$ . From (12), the solution error-in-variables least squares is given by

$$\hat{q}_i = (\Phi^\top(\Phi - \Psi_+ + \mathbf{F}))^{-1} \Phi^\top \hat{\mathbf{c}}_i. \quad (28)$$

Rearranging the terms in (28) yields

$$\Phi^\top(\Phi - \Psi_+ + \mathbf{F})\hat{p} = \Phi^\top \hat{\mathbf{c}}_i \Rightarrow \Phi \hat{q}_i = \Phi(\Phi^\top \Phi)^{-1} \Phi^\top(\hat{\mathbf{c}}_i + (\Psi_+ - \mathbf{F})\hat{q}_i). \quad (29)$$

Define  $P_\Phi = \Phi(\Phi^\top \Phi)^{-1} \Phi^\top$  as the orthogonal projection onto the columns of  $\Phi$ . Combining (11), (29), and using the fact that  $P_\Phi \Phi = \Phi$  yields

$$P_\Phi(\Phi - \Xi + \mathbf{F})(\hat{q}_i^{\text{true}} - \hat{q}_i) = P_\Phi(\Xi - \Psi_+)\hat{q}_i. \quad (30)$$

The  $i^{\text{th}}$  row of  $\Phi - \Xi + \mathbf{F}$  can be expressed as,

$$\begin{aligned} & \text{svec} \left( \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix}^\top - \mathbb{E} \left[ \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t+1) \\ K_{\mathcal{I}_{\hat{Q}}}^i(x_{\mathcal{I}_{\hat{Q}}}^i(t+1)) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t+1) \\ K_{\mathcal{I}_{\hat{Q}}}^i(x_{\mathcal{I}_{\hat{Q}}}^i(t+1)) \end{bmatrix}^\top \right] + \sigma_w^2 \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix} \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix}^\top \right), \\ & = \text{svec} \left( \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix}^\top - L \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_{\hat{Q}}}^i(t) \\ u_{\mathcal{I}_{\hat{Q}}}^i(t) \end{bmatrix}^\top L^\top \right) = (\mathbb{I} - L \otimes_s L) \phi_t, \\ & \text{where } L = \begin{bmatrix} A_{\mathcal{I}_{\hat{Q}}}^i & B_{\mathcal{I}_{\hat{Q}}}^i \\ K_{\mathcal{I}_{\hat{Q}}}^i A_{\mathcal{I}_{\hat{Q}}}^i & K_{\mathcal{I}_{\hat{Q}}}^i B_{\mathcal{I}_{\hat{Q}}}^i \end{bmatrix}. \end{aligned} \quad (31)$$

Combining (31) and (30) and assuming that  $\Phi$  is full column rank, we obtain

$$\begin{aligned} & \Phi(\mathbb{I} - L \otimes_s L)^\top(\hat{q}_i^{\text{true}} - \hat{q}_i) = P_\Phi(\Xi - \Psi_+)\hat{q}_i \\ & \Rightarrow (\mathbb{I} - L \otimes_s L)^\top(\hat{q}_i^{\text{true}} - \hat{q}_i) = (\Phi^\top \Phi)^{-1} \Phi^\top(\Xi - \Psi_+)\hat{q}_i. \end{aligned} \quad (32)$$

Let  $\sigma_{\min}(\cdot)$  denote the minimum singular value of a matrix. Then, we have that

$$\|(\mathbb{I} - L \otimes_s L)^\top(\hat{q}_i^{\text{true}} - \hat{q}_i)\| \geq \sigma_{\min}(\mathbb{I} - L \otimes_s L) \|\hat{q}_i^{\text{true}} - \hat{q}_i\|, \quad (33)$$

$$\begin{aligned} \|(\Phi^\top \Phi)^{-1} \Phi^\top(\Xi - \Psi_+)\hat{q}_i\| & \leq \sigma_{\max}((\Phi^\top \Phi)^{-\frac{1}{2}}) \|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top(\Xi - \Psi_+)\hat{q}_i\| \\ & = \lambda_{\max}((\Phi^\top \Phi)^{-\frac{1}{2}}) \|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top(\Xi - \Psi_+)\hat{q}_i\| \end{aligned}$$

( $\because \Phi^\top \Phi$  is symmetric and P.S.D.,  $(\Phi^\top \Phi)^{-\frac{1}{2}}$  is symmetric and P.S.D.)

$$\begin{aligned} & = \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top(\Xi - \Psi_+)\hat{q}_i\|}{\lambda_{\min}((\Phi^\top \Phi)^{\frac{1}{2}})} \\ & = \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top(\Xi - \Psi_+)\hat{q}_i\|}{\sqrt{\lambda_{\min}(\Phi^\top \Phi)}} \\ & = \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top(\Xi - \Psi_+)\hat{q}_i\|}{\sigma_{\min}(\Phi)} \end{aligned} \quad (34)$$

Combining (36), (33), (34) yields

$$\begin{aligned}
 \sigma_{\min}(\mathbb{I} - L \otimes_s L) \|\hat{q}_i^{\text{true}} - \hat{q}_i\| &\leq \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top (\Xi - \Psi_+) \hat{q}_i\|}{\sigma_{\min}(\Phi)} \\
 \Rightarrow \|\hat{q}_i^{\text{true}} - \hat{q}_i\| &\leq \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top (\Xi - \Psi_+) \hat{q}_i\|}{\sigma_{\min}(\Phi) \sigma_{\min}(\mathbb{I} - L \otimes_s L)} \\
 &\leq \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top (\Xi - \Psi_+)\| \|\hat{q}_i^{\text{true}} - \hat{q}_i\|}{\sigma_{\min}(\Phi) \sigma_{\min}(\mathbb{I} - L \otimes_s L)} + \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top (\Xi - \Psi_+) \hat{q}_i^{\text{true}}\|}{\sigma_{\min}(\Phi) \sigma_{\min}(\mathbb{I} - L \otimes_s L)} \\
 &\quad \text{(By triangle inequality and Cauchy-Schwartz inequality)}
 \end{aligned} \tag{35}$$

If  $\frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top (\Xi - \Psi_+)\|}{\sigma_{\min}(\Phi) \sigma_{\min}(\mathbb{I} - L \otimes_s L)} < \frac{1}{2}$ , then

$$\|\hat{q}_i^{\text{true}} - \hat{q}_i\| \leq 2 \frac{\|(\Phi^\top \Phi)^{-\frac{1}{2}} \Phi^\top (\Xi - \Psi_+) \hat{q}_i^{\text{true}}\|}{\sigma_{\min}(\Phi) \sigma_{\min}(\mathbb{I} - L \otimes_s L)}. \tag{36}$$

Observe that for each  $i \in \mathcal{V}$ , due to Lemma 3.1, (36) is analogous to a single agent setting with state dimension  $n_{\hat{x}}^i = n_x |\mathcal{I}_{\hat{Q}}^i|$ , and control dimension  $n_{\hat{u}}^i = n_u |\mathcal{I}_{\hat{Q}}^i|$ . In the interest of space, we omit the details of the proof and provide the bound analogous to Krauth et al. (2019) for the direct case. Under the pre-conditions stated in Theorem 5.1, if the trajectory length  $T$  satisfies

$$T \geq \tilde{O}(1) \max \left\{ (n_{\hat{x}}^i + n_{\hat{u}}^i)^2, \frac{(n_{\hat{x}}^i)^2 (n_{\hat{x}}^i + n_{\hat{u}}^i)^2 \|\hat{K}_{\mathcal{I}_{\hat{Q}}^i}^{\text{play}}\|_+^4}{\sigma_\eta^4} - \sigma_w^2 \bar{\sigma}_i^2 \frac{\tau^4 \|K_{\mathcal{I}_{\hat{Q}}^i}\|_+^8 (\|A_{\mathcal{I}_{\hat{Q}}^i}\|^2 + \|B_{\mathcal{I}_{\hat{Q}}^i}\|^2)^2}{\rho^4 (1 - \rho^2)^2} \right\},$$

then with probability at least  $1 - \delta$ , we have taht

$$\|\hat{q}_i^{\text{true}} - \hat{q}_i\| \leq \frac{\tilde{O}(1) (n_{\hat{x}}^i + n_{\hat{u}}^i) \|K_{\mathcal{I}_{\hat{Q}}^i}^{\text{play}}\|_+^2}{\sigma_\eta^2 \sqrt{T}} - \sigma_w \bar{\sigma}_i \|\hat{Q}_i^{\text{true}}\|_F \frac{\tau^2 \|K_{\mathcal{I}_{\hat{Q}}^i}\|_+^4 (\|A_{\mathcal{I}_{\hat{Q}}^i}\|^2 + \|B_{\mathcal{I}_{\hat{Q}}^i}\|^2)}{\rho^2 (1 - \rho^2)},$$

where  $\tilde{O}(1)$  hides  $\text{polylog} \left( \frac{T}{\delta}, \frac{1}{\sigma_\eta^4}, \tau, n_{\hat{x}}^i, \|\Sigma_0\|, \|K_{\mathcal{I}_{\hat{Q}}^i}^{\text{play}}\|, \|\mathfrak{P}_\infty\| \right)$ . ■

## Appendix G. Analysis of the indirect case

Define  $n_x^i = n_x |\mathcal{I}_Q^i|$ , and  $n_u^i = n_u |\mathcal{I}_Q^i|$ .

**Corollary 1** Consider  $\delta \in (0, 1)$ . Let the initial global state and the global control (during sample generation)  $\forall t$  satisfy  $x(0) \sim \mathcal{N}(x_0, \Sigma_0)$ ,  $u(t) = K^{\text{play}} x(t) + \eta_t$ ,  $\eta(t) \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 \mathbb{I}_{N_{n_u}})$ , and  $\sigma_\eta \leq \sigma_w$ . For each  $i \in \mathcal{V}$ , let  $K_{\mathcal{I}_Q^i}^{\text{play}}$ ,  $K_{\mathcal{I}_Q^i}$  stabilize  $(A_{\mathcal{I}_Q^i}, B_{\mathcal{I}_Q^i})$ . Assume that  $A_{\mathcal{I}_Q^i} + B_{\mathcal{I}_Q^i} K_{\mathcal{I}_Q^i}$  and  $A_{\mathcal{I}_Q^i} + B_{\mathcal{I}_Q^i} K_{\mathcal{I}_Q^i}^{\text{play}}$  are  $(\tau, \rho)$ -stable. Let  $\mathfrak{P}_\infty = \mathcal{L} \left( A_{\mathcal{I}_Q^i} + B_{\mathcal{I}_Q^i} K_{\mathcal{I}_Q^i}, \sigma_w^2 \mathbb{I}_{n_x^i} + \sigma_\eta^2 B_{\mathcal{I}_Q^i} B_{\mathcal{I}_Q^i}^\top \right)$

and  $\bar{\sigma}_i = \sqrt{\tau^2 \rho^4 \|\Sigma_0^x\| + \|\mathfrak{P}_\infty\| + \sigma_w^2 + \sigma_\eta^2 \|B_{\mathcal{I}_Q^i}\|^2}$ . Further,  $\forall i \in \mathcal{V}$ , let  $T_i$  denote the minimum number of samples required during learning. Suppose that

$$T_i \geq \tilde{O}(1) \max \left\{ (n_x^i + n_u^i)^2, \frac{(n_x^i)^2 (n_x^i + n_u^i)^2 \|K_{\mathcal{I}_Q^i}^{\text{play}}\|_+^4}{\sigma_\eta^4} \sigma_w^2 \bar{\sigma}_i^2 \frac{\tau_i^4 \|K_{\mathcal{I}_Q^i}\|_+^8 (\|A_{\mathcal{I}_Q^i}\|^2 + \|B_{\mathcal{I}_Q^i}\|^2)^2}{\rho_i^4 (1 - \rho_i^2)^2} \right\}.$$

Then, with probability  $1 - \delta$ ,

$$\|\hat{q}_i^{\text{true}} - \hat{q}_i^{\text{indirect}}\| \leq \sum_{j \in \mathcal{I}_{GD}^i} \frac{\tilde{O}(1) (n_x^j + n_u^j) \|K_{\mathcal{I}_Q^j}^{\text{play}}\|_+^2}{\sigma_\eta^2 \sqrt{T}} \sigma_w \bar{\sigma}_j \|Q_j^{\text{true}}\|_F \frac{\tau_j^2 \|K_{\mathcal{I}_Q^j}\|_+^4 (\|A_{\mathcal{I}_Q^j}\|^2 + \|B_{\mathcal{I}_Q^j}\|^2)}{\rho_j^2 (1 - \rho_j^2)}$$

whenever  $T \geq \max \{T_j\}_{j \in \mathcal{I}_{GD}^i}$ , where  $\tilde{O}(1)$  hides  $\text{polylog} \left( \frac{T}{\delta}, \frac{1}{\sigma_\eta^4}, \tau, n_x, \|\Sigma_0\|, \|K_{\mathcal{I}_Q^i}^{\text{play}}\|, \|\mathfrak{P}_\infty\| \right)$ .

**Proof** For brevity, in the remainder of the proof denote  $\phi_t = \text{svec} \left( \begin{bmatrix} x_{\mathcal{I}_Q^i}(t) \\ u_{\mathcal{I}_Q^i}(t) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_Q^i}(t) \\ u_{\mathcal{I}_Q^i}(t) \end{bmatrix}^\top \right)$ ,  
 $\psi_t = \text{svec} \left( \begin{bmatrix} x_{\mathcal{I}_Q^i}(t) \\ K_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_Q^i}(t) \\ K_{\mathcal{I}_Q^i} x_{\mathcal{I}_Q^i}(t) \end{bmatrix}^\top \right)$ ,  $f = \text{svec} \left( \begin{bmatrix} \Sigma_{\mathcal{I}_Q^i}^x & \Sigma_{\mathcal{I}_Q^i}^x K_{\mathcal{I}_Q^i}^\top \\ K_{\mathcal{I}_Q^i} \Sigma_{\mathcal{I}_Q^i}^x & K_{\mathcal{I}_Q^i} \Sigma_{\mathcal{I}_Q^i}^x K_{\mathcal{I}_Q^i}^\top \end{bmatrix} \right)$ , and  $\xi_t =$   
 $\mathbb{E} \left[ \text{svec} \left( \begin{bmatrix} x_{\mathcal{I}_Q^i}(t+1) \\ u_{\mathcal{I}_Q^i}(t+1) \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}_Q^i}(t+1) \\ u_{\mathcal{I}_Q^i}(t+1) \end{bmatrix}^\top \right) \right]$ . Employing a *linear architecture*, the Q-function for each agent  $i$  can be expressed as

$$c_i(x_{\mathcal{I}_Q^i}(t), u_{\mathcal{I}_Q^i}(t)) = \lambda + [\phi_t - \xi_t] \text{svec}(Q_i), \quad (37)$$

where  $\lambda \in \mathbb{R}$  is a free parameter to satisfy the fixed point equation. Let  $\lambda = \left\langle Q_i, \sigma_w^2 \begin{bmatrix} \mathbb{I}_{n_x | \mathcal{I}_Q^i} \\ K_{\mathcal{I}_Q^i}^\top \end{bmatrix} \begin{bmatrix} \mathbb{I}_{n_x | \mathcal{I}_Q^i} \\ K_{\mathcal{I}_Q^i}^\top \end{bmatrix}^\top \right\rangle$ .

For a single trajectory  $\{x_{\mathcal{I}_Q^i}(t), u_{\mathcal{I}_Q^i}(t), x_{\mathcal{I}_Q^i}(t+1)\}_{t=1}^{T_i}$ , the bellman equation for agent  $i$  can be expressed in matrix form as

$$\mathbf{c}_i = (\Phi - \Xi + \mathbf{F}) \mathbf{q}_i, \quad (38)$$

where  $\Phi^\top = [\phi_1, \phi_2, \dots, \phi_{T_i}]$ ,  $\Xi^\top = [\xi_1, \xi_2, \dots, \xi_{T_i}]$ ,  $\mathbf{c}_i^\top = [c_i(1), c_i(2), \dots, c_i(T_i)]$ ,  $\mathbf{F}^\top = [f_1, f_2, \dots, f_{T_i}]$ . Observe that (38) is analogous to (11). Thus, using Theorem 5.1, we can conclude that if  $T_i$  satisfies

$$T_i \geq \tilde{O}(1) \max \left\{ (n_x^i + n_u^i)^2, \frac{(n_x^i)^2 (n_x^i + n_u^i)^2 \|K_{\mathcal{I}_Q^i}^{\text{play}}\|_+^4}{\sigma_\eta^4} \sigma_w^2 \bar{\sigma}_i^2 \frac{\tau_i^4 \|K_{\mathcal{I}_Q^i}\|_+^8 (\|A_{\mathcal{I}_Q^i}\|^2 + \|B_{\mathcal{I}_Q^i}\|^2)^2}{\rho_i^4 (1 - \rho_i^2)^2} \right\}, \quad (39)$$

where  $\bar{\sigma}_i = \sqrt{\tau^2 \rho^4 \|\Sigma^x(0)\| + \|\mathfrak{P}_\infty\| + \sigma_w^2 + \sigma_\eta^2 \|B_{\mathcal{I}_Q^i}\|^2}$ . Then,

$$\|\hat{q}_i^{\text{true}} - q_i\| \leq \frac{\tilde{O}(1) (n_x^i + n_u^i) \|K_{\mathcal{I}_Q^i}^{\text{play}}\|_+^2}{\sigma_\eta^2 \sqrt{T_i}} \sigma_w \bar{\sigma}_i \|Q_i^{\text{true}}\|_F \frac{\tau_i^2 \|K_{\mathcal{I}_Q^i}\|_+^4 (\|A_{\mathcal{I}_Q^i}\|^2 + \|B_{\mathcal{I}_Q^i}\|^2)}{\rho_i^2 (1 - \rho_i^2)} \quad (40)$$



However, note that in general  $\hat{q}_i^{\text{indirect}} \neq \sum_{j \in \mathcal{I}_{\text{GD}}^i} q_j$  as the agents might not correspond to each other if  $\mathcal{I}_Q^j \neq \mathcal{I}_Q^k, \forall j, k \in \mathcal{I}_{\text{GD}}^i$ . Hence, to make the dimensions consistent, and compute the estimated local Q-function for each  $j \in \mathcal{I}_{\text{GD}}^i$ , we define a projection operator  $\mathcal{P}_Q^j = \text{blk\_diag}(P_{\mathcal{I}_Q^j, \mathcal{I}_Q^j}^{n_x}, P_{\mathcal{I}_Q^j, \mathcal{I}_Q^j}^{n_u})$ , where  $P_{S_1, S_2}^n$  is the projection defined in Section 4. Then, we have that  $\hat{q}_i^{\text{indirect}} = \sum_{j \in \mathcal{I}_{\text{GD}}^i} \text{svec} \left( (\mathcal{P}_Q^j)^\top \text{smat}(q_j) \mathcal{P}_Q^j \right)$ , and  $\hat{q}_i^{\text{true}} = \sum_{j \in \mathcal{I}_{\text{GD}}^i} \text{svec} \left( (\mathcal{P}_Q^j)^\top \text{smat}(q_j^{\text{true}}) \mathcal{P}_Q^j \right)$ . Therefore, the error in estimation of  $\hat{q}_i^{\text{true}}$  in the indirect case can be expressed as

$$\begin{aligned}
 \|\hat{q}_i^{\text{true}} - \hat{q}_i^{\text{indirect}}\| &= \left\| \sum_{j \in \mathcal{I}_{\text{GD}}^i} \text{svec} \left( (\mathcal{P}_Q^j)^\top \text{smat}(q_j^{\text{true}}) \mathcal{P}_Q^j \right) - \sum_{j \in \mathcal{I}_{\text{GD}}^i} \text{svec} \left( (\mathcal{P}_Q^j)^\top \text{smat}(q_j) \mathcal{P}_Q^j \right) \right\| \\
 &= \left\| \sum_{j \in \mathcal{I}_{\text{GD}}^i} \text{svec} \left( (\mathcal{P}_Q^j)^\top (\text{smat}(q_j^{\text{true}} - q_j)) \mathcal{P}_Q^j \right) \right\| \quad (\text{Due to the linearity of } \text{svec}(\cdot), \text{smat}(\cdot)) \\
 &= \left\| \sum_{j \in \mathcal{I}_{\text{GD}}^i} (q_j^{\text{true}} - q_j) \right\| \quad (\because \|q_j\| = \|(\mathcal{P}_Q^j)^\top \text{smat}(q_j) \mathcal{P}_Q^j\| \forall j) \\
 &\leq \sum_{j \in \mathcal{I}_{\text{GD}}^i} \|q_j^{\text{true}} - q_j\| \quad (\text{Using triangle inequality}). \tag{41}
 \end{aligned}$$

Combining (39), (40), and (41), we obtain that whenever the length of trajectory (number of samples) satisfies

$$T \geq \max \{T_j\}_{j \in \mathcal{I}_{\text{GD}}^i} \tag{42}$$

then with probability  $1 - \delta$ , we have

$$\|\hat{q}_i^{\text{true}} - \hat{q}_i^{\text{indirect}}\| \leq \sum_{j \in \mathcal{I}_{\text{GD}}^i} \frac{\tilde{O}(1)(n_x^j + n_u^j) \|K_{\mathcal{I}_Q^j}^{\text{play}}\|_+^2}{\sigma_\eta^2 \sqrt{T_j}} \sigma_w \bar{\sigma}_j \|Q_j^{\text{true}}\|_F \frac{\tau_j^2 \|K_{\mathcal{I}_Q^j}\|_+^4 (\|A_{\mathcal{I}_Q^j}\|^2 + \|B_{\mathcal{I}_Q^j}\|^2)}{\rho_j^2 (1 - \rho_j^2)}, \tag{43}$$

where  $\tilde{O}(1)$  hides  $\text{polylog} \left( \frac{T}{\delta}, \frac{1}{\sigma_\eta^4}, \tau, n_x, \|\Sigma_0\|, \|K_{\mathcal{I}_Q}^{\text{play}}\|, \|\mathfrak{P}_\infty\| \right)$ . ■