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) $f : \mathcal{S} \times \mathcal{U} \to \mathbb{R}$ such that

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

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 $g_j : \mathcal{Z}_j \to P(\mathcal{U}_j)$ such that

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

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

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

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

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

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

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

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, $i \in \mathcal{I}_{\text{GD}}^s \; \forall \; s \in \mathcal{R}_{SO}^j$ 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

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

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

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

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

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

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

$$\begin{aligned}\nabla_{K_i}\bar{Q}_i &= \nabla_{K_i}\mathbb{E}_\pi\Bigg[\sum_{j\setminus\mathcal{I}^i_{\mathrm{GD}}}\sum_{t=0}^{\infty} c_j(x_{\mathcal{I}^j_C}(t), u_{\mathcal{I}^j_C}(t))|x(0)=x, u(0)=u\Bigg]\\ &\overset{(a)}{=} \mathbb{E}_\pi\Bigg[\nabla_{K_i}\sum_{j\setminus\mathcal{I}^i_{\mathrm{GD}}}\sum_{t=0}^{\infty} c_j(x_{\mathcal{I}^j_C}(t), u_{\mathcal{I}^j_C}(t))|x(0)=x, u(0)=u\Bigg] = 0, \end{aligned} \tag{23}$$

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 $\nabla_{K_i}Q(s,a) = \nabla_{K_i}[\widehat{Q}_i + \bar{Q}_i] = \nabla_{K_i}\widehat{Q}_i,$ as required. ■

Appendix D. Proof of Proposition 4.1

Proof From (9), we have

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

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

$$\begin{aligned}&\mathbb{E}\left[\widehat{Q}_i(x_{\mathcal{I}^i_{\widehat{Q}}}(t{+}1), u_{\mathcal{I}^i_{\widehat{Q}}}(t{+}1))\right]\\
&= \mathbb{E}\left[\begin{bmatrix} x_{\mathcal{I}^i_{\widehat{Q}}}(t{+}1) \\ u_{\mathcal{I}^i_{\widehat{Q}}}(t{+}1)\end{bmatrix}^\intercal \begin{bmatrix} S_{\mathcal{I}^i_{\widehat{Q}}} & 0 \\ 0 & R_{\mathcal{I}^i_{\widehat{Q}}}\end{bmatrix} \begin{bmatrix} x_{\mathcal{I}^i_{\widehat{Q}}}(t{+}1) \\ u_{\mathcal{I}^i_{\widehat{Q}}}(t{+}1)\end{bmatrix}\right] + \mathbb{E}\left[\mathbb{E}\left[\widehat{Q}_i(x_{\mathcal{I}^i_{\widehat{Q}}}(t{+}2), u_{\mathcal{I}^i_{\widehat{Q}}}(t{+}2))\right]\right]\\
&= \mathbb{E}\left[(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t) + w_{\mathcal{I}^i_{\widehat{Q}}}(t))^\intercal S_{\mathcal{I}^i_{\widehat{Q}}}(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t) + w_{\mathcal{I}^i_{\widehat{Q}}}(t))\right] +\\
&\mathbb{E}\Big[(K_{\mathcal{I}^i_{\widehat{Q}}}(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t) + w_{\mathcal{I}^i_{\widehat{Q}}}(t)))^\intercal R_{\mathcal{I}^i_{\widehat{Q}}}(K_{\mathcal{I}^i_{\widehat{Q}}}(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t) + w_{\mathcal{I}^i_{\widehat{Q}}}(t)))\Big]\\
&+ \mathbb{E}\left[\mathbb{E}\left[\widehat{Q}_i(x_{\mathcal{I}^i_{\widehat{Q}}}(t{+}2), u_{\mathcal{I}^i_{\widehat{Q}}}(t{+}2))\right]\right]\\
&= (A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t))^\intercal S_{\mathcal{I}^i_{\widehat{Q}}}(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t)) + \sigma_w^2\mathrm{Tr}\left(S_{\mathcal{I}^i_{\widehat{Q}}} + K^\intercal_{\mathcal{I}^i_{\widehat{Q}}}R_{\mathcal{I}^i_{\widehat{Q}}}K_{\mathcal{I}^i_{\widehat{Q}}}\right)\\
&+ (K_{\mathcal{I}^i_{\widehat{Q}}}(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t)))^\intercal R_{\mathcal{I}^i_{\widehat{Q}}}(K_{\mathcal{I}^i_{\widehat{Q}}}(A_{\mathcal{I}^i_{\widehat{Q}}}x_{\mathcal{I}^i_{\widehat{Q}}}(t) + B_{\mathcal{I}^i_{\widehat{Q}}}u_{\mathcal{I}^i_{\widehat{Q}}}(t)))\\
&+ \mathbb{E}\left[\mathbb{E}\left[\widehat{Q}_i(x_{\mathcal{I}^i_{\widehat{Q}}}(t{+}2), u_{\mathcal{I}^i_{\widehat{Q}}}(t{+}2))\right]\right]\end{aligned} \tag{25}$$

$$
\begin{aligned}
&= \begin{bmatrix} x_{\mathcal{I}^i_{\widehat{Q}}}(t) \\ u_{\mathcal{I}^i_{\widehat{Q}}}(t) \end{bmatrix}^\intercal \begin{bmatrix} A^\intercal_{\mathcal{I}^i_{\widehat{Q}}} \\ B^\intercal_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix} (S_{\mathcal{I}^i_{\widehat{Q}}} + K^\intercal_{\mathcal{I}^i_{\widehat{Q}}} R_{\mathcal{I}^i_{\widehat{Q}}} K_{\mathcal{I}^i_{\widehat{Q}}}) \begin{bmatrix} A_{\mathcal{I}^i_{\widehat{Q}}} & B_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix} \begin{bmatrix} x_{\mathcal{I}^i_{\widehat{Q}}}(t) \\ u_{\mathcal{I}^i_{\widehat{Q}}}(t) \end{bmatrix} + \sigma_w^2 \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix}^\intercal \begin{bmatrix} S_{\mathcal{I}^i_{\widehat{Q}}} & 0 \\ 0 & R_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix} \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix} \\
&+ \mathbb{E}\left[\mathbb{E}\left[\widehat{Q}_i(x_{\mathcal{I}^i_{\widehat{Q}}}(t+2), u_{\mathcal{I}^i_{\widehat{Q}}}(t+2))\right]\right]. \qquad (26)
\end{aligned}
$$

Recursive expansion of (26) yields

$$
\widehat{Q}_i(x_{\mathcal{I}^i_{\widehat{Q}}}, u_{\mathcal{I}^i_{\widehat{Q}}}) = \begin{bmatrix} x_{\mathcal{I}^i_{\widehat{Q}}}(t) \\ u_{\mathcal{I}^i_{\widehat{Q}}}(t) \end{bmatrix}^\intercal \widehat{Q}_i \begin{bmatrix} x_{\mathcal{I}^i_{\widehat{Q}}}(t) \\ u_{\mathcal{I}^i_{\widehat{Q}}}(t) \end{bmatrix} + \sigma_w^2 \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix}^\intercal \widehat{Q}_i \begin{bmatrix} \mathbb{I} \\ K_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix}, \qquad (27)
$$

where with a slight abuse of notation

$$
\widehat{Q}_i = \begin{bmatrix} S_{\mathcal{I}^i_{\widehat{Q}}} & 0 \\ 0 & R_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix} + \begin{bmatrix} A^\intercal_{\mathcal{I}^i_{\widehat{Q}}} \\ B^\intercal_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix} \mathcal{L}\left(A_{\mathcal{I}^i_{\widehat{Q}}} + B_{\mathcal{I}^i_{\widehat{Q}}} K_{\mathcal{I}^i_{\widehat{Q}}}, S_{\mathcal{I}^i_{\widehat{Q}}} + K^\intercal_{\mathcal{I}^i_{\widehat{Q}}} R_{\mathcal{I}^i_{\widehat{Q}}} K_{\mathcal{I}^i_{\widehat{Q}}}\right) \begin{bmatrix} A_{\mathcal{I}^i_{\widehat{Q}}} & B_{\mathcal{I}^i_{\widehat{Q}}} \end{bmatrix},
$$

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

Appendix E. Proof of Lemma 5.1

Proof Define $\mathcal{R}^i_{(SO)^\intercal} = \{j \in \mathcal{V} | j \xrightarrow{\mathcal{E}^\intercal_{SO}} i\}$, and $\mathcal{I}^i_{C^\intercal} = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}^\intercal_O\}$.

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

Consider the case where $i \in \{\mathcal{R}^q_{SO}\}_{q \in \mathcal{I}^j_C}$. Suppose that there exists an $r \in \mathcal{I}^j_C$ for which $i \in \mathcal{R}^r_{SO}$. Then as $k \notin \mathcal{I}^j_C$ and $k \notin \{\mathcal{R}^q_{SO}\}_{q \in \mathcal{I}^j_C}$, we have $\forall\, m \in \mathcal{R}^i_{(SO)^\intercal}$ and $\forall\, p \in \mathcal{R}^k_{(SO)^\intercal}$, $\mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \emptyset$. This is because otherwise for every $l \in \mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal}$, we obtain $l \in \mathcal{I}^i_{\text{GD}}$ and $k \in \mathcal{I}^l_Q$, implying $k \in \mathcal{I}^i_{\widehat{Q}}$, which contradicts our assumption.

Alternatively, if $i \in \mathcal{I}^j_C$, $k \notin \mathcal{I}^j_C$ and $k \notin \{\mathcal{R}^q_{SO}\}_{q \in \mathcal{I}^j_C}$ imply that $\forall\, p \in \mathcal{R}^k_{(SO)^\intercal}$, $\mathcal{I}^i_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \emptyset$. Otherwise for every $l \in \mathcal{I}^i_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal}$, we obtain $l \in \mathcal{I}^i_{\text{GD}}$, and $k \in \mathcal{I}^l_Q$ implying $k \in \mathcal{I}^i_{\widehat{Q}}$ which contradicts our assumption.

As $i \in \mathcal{R}^i_{(SO)^\intercal}$, we have that $\mathcal{I}^i_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \emptyset$ whenever $\mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \emptyset$. Therefore, we conclude that if $\mathcal{I}^i_{\widehat{Q}} \subset \mathcal{V}$, then $\forall\, m \in \mathcal{R}^i_{(SO)^\intercal}$, $\forall\, p \in \mathcal{R}^k_{(SO)^\intercal}$, $\mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \emptyset$.

Sufficient condition.

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

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

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

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

(b) **Necessary condition.** Consider an $i \in \mathcal{V}$ and assume that there exists a $j \in \mathcal{I}^i_{\text{GD}}$, such that $\mathcal{I}^j_Q \subset \mathcal{I}^i_{\widehat{Q}}$. This implies that $\exists\, k \in \mathcal{I}^i_{\widehat{Q}}$ such that $k \notin \mathcal{I}^j_Q$, and $k \in \bigcup_{h \in \mathcal{I}^i_{\text{GD}} \setminus \{j\}} \mathcal{I}^h_Q$. If $k \notin \mathcal{I}^j_Q$, then by definition, $k \notin \mathcal{I}^j_C$, and $k \notin \{\mathcal{R}^l_{SO}\}_{l \in \mathcal{I}^j_C}$. But, $k \in \bigcup_{h \in \mathcal{I}^i_{\text{GD}} \setminus \{j\}} \mathcal{I}^h_Q$ implies that $\exists\, h \in \mathcal{I}^i_{\text{GD}} \setminus \{j\}$ such that either $k \in \mathcal{I}^h_C$ or $k \in \{\mathcal{R}^m_{SO}\}_{m \in \mathcal{I}^h_C}$.

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

- If $i \in \mathcal{I}^h_C$, then $\mathcal{I}^i_{C^\intercal} \cap \mathcal{I}^k_{C^\intercal} = \{h\} \neq \emptyset$. or,
- If $i \in \{\mathcal{R}^l_{SO}\}_{l \in \mathcal{I}^j_C}$, then $\exists$ an $m \in \mathcal{I}^h_C \cap \mathcal{R}^i_{(SO)^\intercal}$. Hence, $\mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^k_{C^\intercal} = \{h\} \neq \emptyset$.

Case 2 Let $k \in \{\mathcal{R}^m_{SO}\}_{m \in \mathcal{I}^h_C}$. Then, $\exists$ an $p \in \mathcal{I}^h_C \cap \mathcal{R}^k_{(SO)^\intercal}$, and as $h \in \mathcal{I}^i_{\text{GD}}$, either $i \in \mathcal{I}^h_C$, or $i \in \{\mathcal{R}^l_{SO}\}_{l \in \mathcal{I}^h_C}$.

- If $i \in \mathcal{I}^j_C$, then $\mathcal{I}^i_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \{h\} \neq \emptyset$. or,
- If $i \in \{\mathcal{R}^l_{SO}\}_{l \in \mathcal{I}^h_C}$, then $\exists$ an $m \in \mathcal{I}^h_C \cap \mathcal{R}^i_{(SO)^\intercal}$. Hence, $\mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} = \{h\} \neq \emptyset$.

Therefore, in either case we conclude that if $\mathcal{I}^i_Q \subset \mathcal{I}^i_{\widehat{Q}}$, then $p \in \mathcal{R}^k_{(SO)^\intercal}$, $m \in \mathcal{R}^i_{(SO)^\intercal}$, such that $\mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal} \subset \mathcal{I}^i_{\text{GD}}$.

Sufficient condition. Consider an $i \in \mathcal{V}$ and assume that $\exists\, j \in \mathcal{I}^i_{\text{GD}}$ for which $\exists\, k \in \mathcal{V} \setminus \mathcal{I}^j_Q$. Let $h \in \mathcal{I}^i_{\text{GD}}$, $m \in \mathcal{R}^i_{(SO)^\intercal}$, and $p \in \mathcal{R}^k_{(SO)^\intercal}$, such that $h \in \mathcal{I}^m_{C^\intercal} \cap \mathcal{I}^p_{C^\intercal}$. Hence, as $p \in \mathcal{I}^h_C$, by definition $k \in \mathcal{I}^h_Q$. As $h \in \mathcal{I}^i_{\text{GD}}$, we have that $k \in \mathcal{I}^i_{\widehat{Q}}$. However, $k \notin \mathcal{I}^j_Q$ implies that $k \in \mathcal{I}^i_{\widehat{Q}} \setminus \mathcal{I}^i_Q$ or $\mathcal{I}^i_Q \subset \mathcal{I}^i_{\widehat{Q}}$ 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

$$\hat{q}_i = (\mathbf{\Phi}^\intercal(\mathbf{\Phi} - \mathbf{\Psi}_+ + \mathbf{F}))^{-1}\mathbf{\Phi}^\intercal \hat{\mathbf{c}}_i. \tag{28}$$

Rearranging the terms in (28) yields

$$\mathbf{\Phi}^\intercal(\mathbf{\Phi} - \mathbf{\Psi}_+ + \mathbf{F})\hat{q}_i = \mathbf{\Phi}^\intercal\hat{\mathbf{c}}_i \Rightarrow \mathbf{\Phi}\hat{q}_i = \mathbf{\Phi}(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-1}\mathbf{\Phi}^\intercal(\hat{\mathbf{c}}_i + (\mathbf{\Psi}_+ - \mathbf{F})\hat{q}_i). \tag{29}$$

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

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

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

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

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

$$\begin{aligned}
&\mathbf{\Phi}(\mathbb{I} - L\otimes_s L)^\intercal(\hat{q}_i^{\text{true}} - \hat{q}_i) = P_{\mathbf{\Phi}}(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i\\
&\Rightarrow (\mathbb{I} - L\otimes_s L)^\intercal(\hat{q}_i^{\text{true}} - \hat{q}_i) = (\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-1}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i.
\end{aligned} \tag{32}$$

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

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

$$\begin{aligned}
||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-1}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i|| &\leq \sigma_{\text{max}}((\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}})||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i||\\
&= \lambda_{\text{max}}((\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}})||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i||\\
&\quad (\because \mathbf{\Phi}^\intercal\mathbf{\Phi} \text{ is symmetric and P.S.D,} (\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}} \text{ is symmetric and P.S.D.})\\
&= \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i||}{\lambda_{\text{min}}((\mathbf{\Phi}^\intercal\mathbf{\Phi})^{\frac{1}{2}})}\\
&= \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i||}{\sqrt{\lambda_{\text{min}}(\mathbf{\Phi}^\intercal\mathbf{\Phi})}}\\
&= \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi} - \mathbf{\Psi}_+)\hat{q}_i||}{\sigma_{\text{min}}(\mathbf{\Phi})}
\end{aligned} \tag{34}$$

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

$$
\begin{aligned}
&\sigma_{\min}(\mathbb{I}-L\otimes_s L)||\hat{q}_i^{\text{true}}-\hat{q}_i|| \leq \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi}-\mathbf{\Psi}_+)\hat{q}_i||}{\sigma_{\min}(\mathbf{\Phi})} \\
&\Rightarrow ||\hat{q}_i^{\text{true}}-\hat{q}_i|| \leq \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi}-\mathbf{\Psi}_+)\hat{q}_i||}{\sigma_{\min}(\mathbf{\Phi})\sigma_{\min}(\mathbb{I}-L\otimes_s L)} \\
&\leq \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi}-\mathbf{\Psi}_+)||||\hat{q}_i^{\text{true}}-\hat{q}_i||}{\sigma_{\min}(\mathbf{\Phi})\sigma_{\min}(\mathbb{I}-L\otimes_s L)} + \frac{||(\mathbf{\Phi}^\intercal\mathbf{\Phi})^{-\frac{1}{2}}\mathbf{\Phi}^\intercal(\mathbf{\Xi}-\mathbf{\Psi}_+)\hat{q}_i^{\text{true}}||}{\sigma_{\min}(\mathbf{\Phi})\sigma_{\min}(\mathbb{I}-L\otimes_s L)} \\
&\qquad\qquad \text{(By triangle inequality and Cauchy-Scwartz inequality)}
\end{aligned} \tag{35}
$$

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

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

Observe that for each $i \in \mathcal{V}$, due to Lemma 3.1, (36) is analogous to a single agent setting with state dimension $n_{\hat{x}}^i = n_x|\mathcal{I}_{\widehat{Q}}^i|$, and control dimension $n_{\hat{u}}^i = n_u|\mathcal{I}_{\widehat{Q}}^i|$. 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

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

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

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

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

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 $x(0) \sim \mathcal{N}(x_0, \Sigma_0)$, $u(t) = K^{play}x(t)+\eta_t$, $\eta(t) \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2\mathbb{I}_{Nn_u})$, and $\sigma_\eta \leq \sigma_w$. For each $i \in \mathcal{V}$, let $K_{\mathcal{I}_Q^i}^{play}$, $K_{\mathcal{I}_Q^i}$ 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 $A_{\mathcal{I}_Q^i}+B_{\mathcal{I}_Q^i}K_{\mathcal{I}_Q^i}^{play}$ are (τ,ρ)-stable. Let $\mathfrak{P}_\infty = \mathcal{L}\left(A_{\mathcal{I}_Q^i}+B_{\mathcal{I}_Q^i}K_{\mathcal{I}_Q^i}, \sigma_w^2\mathbb{I}_{n_x^i}+\sigma_\eta^2 B_{\mathcal{I}_Q^i}B_{\mathcal{I}_Q^i}^\intercal\right)$*

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*

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

Then, with probability $1-\delta$*,*

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

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

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

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

where $\lambda \in \mathbb{R}$ is a free parameter to satisfy the fixed point equation. Let $\lambda = \left\langle Q_i, \sigma_w^2 \begin{bmatrix} \mathbb{I}_{n_x|I_Q^i|} \\ K^\intercal_{\mathcal{I}_Q^i}\end{bmatrix}\begin{bmatrix} \mathbb{I}_{n_x|I_Q^i|} \\ K^\intercal_{\mathcal{I}_Q^i}\end{bmatrix}^\intercal\right\rangle$. For a single trajectory $\left\{x_{\mathcal{I}_Q^i}(t), u_{\mathcal{I}_Q^i}(t), x_{\mathcal{I}_Q^i}(t{+}1)\right\}_{t=1}^{T_i}$, the bellman equation for agent i can be expressed in matrix form as

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

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

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

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,

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

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 $\mathcal{P}_Q^j = \text{blk\_diag}(P^{n_x}_{\mathcal{I}_Q^j,\mathcal{I}_{\hat{Q}}^i}, P^{n_u}_{\mathcal{I}_Q^j,\mathcal{I}_{\hat{Q}}^i})$, where $P^n_{\mathcal{S}_1,\mathcal{S}_2}$ is the projection defined in Section 4. Then, we have that
$\hat{q}_i^{\text{indirect}} = \sum_{j\in\mathcal{I}_{\text{GD}}^i} \text{svec}\left((\mathcal{P}_Q^j)^\intercal \text{smat}(q_j)\mathcal{P}_Q^j\right)$, and $\hat{q}_i^{\text{true}} = \sum_{j\in\mathcal{I}_{\text{GD}}^i} \text{svec}\left((\mathcal{P}_Q^j)^\intercal \text{smat}(q_j^{\text{true}})\mathcal{P}_Q^j\right)$.
Therefore, the error in estimation of $\hat{q}_i^{\text{true}}$ in the indirect case can be expressed as

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

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

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

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

Appendix H. Remark on the sample complexity of the indirect case

Define $w_1,\ w_2,\ \cdots,\ w_{|\mathcal{I}_{\text{GD}}^i|} \in [0,1]$ such that $\sum_{k=1}^{|\mathcal{I}_{\text{GD}}^i|} w_k = 1$. Then, from (43), observe that to achieve $\left\|\hat{q}_i^{\text{true}} - \hat{q}_i^{\text{indirect}}\right\| \leq \epsilon$, it is sufficient that every $j \in \mathcal{I}_{\text{GD}}^i$ satisfies $\left\|q_j^{\text{true}} - q_j\right\| \leq w_j\epsilon$. From (40), we have that for any $j \in \mathcal{I}_{\text{GD}}^i$, q_j to be $(w_j\epsilon)-$optimal requires

$$T_j \leq \max\left(\frac{(\tilde{O}(1))^2 W_j^2 (n_x^j + n_u^j)^3}{\sigma_\eta^4 w_j^2 \epsilon^2}||Q_j^{\text{true}}||^2, \frac{\tilde{O}(1) W_j^2 (n_x^j)^2 (n_x^j + n_u^j)^2}{\sigma_\eta^4}\right) \text{ samples.}$$

Therefore, we conclude that to ensure $\left\|\hat{q}_i^{\text{true}} - \hat{q}_i^{\text{indirect}}\right\| \leq \epsilon$, it is sufficient to have

$$T_{\text{indirect}} \leq \max_{j \in \mathcal{I}_{\text{GD}}^i} \left(\max \left(\frac{(\tilde{O}(1))^2 W_j^2 (n_x^j + n_u^j)^3}{\sigma_\eta^4 w_j^2 \epsilon^2} ||Q_j^{\text{true}}||^2, \frac{\tilde{O}(1) W_j^2 (n_x^j)^2 (n_x^j + n_u^j)^2}{\sigma_\eta^4} \right) \right) \text{ samples,}$$

where $W_j = ||K^{\text{play}}_{\mathcal{I}^j_{\widehat{Q}}}||_+^2 \sigma_w \bar{\sigma}_j \dfrac{\tau^2 ||K_{\mathcal{I}^j_{\widehat{Q}}}||_+^4 (||A_{\mathcal{I}^j_{\widehat{Q}}}||^2 + ||B_{\mathcal{I}^j_{\widehat{Q}}}||^2)}{\rho^2(1-\rho^2)}$.

For the same $\hat{q}_i^{\text{true}}$ in the direct case, we know from Theorem 5.1 that achieving $\epsilon-$optimal estimate requires at most

$$\begin{aligned} T_{\text{direct}} &\leq \max \left(\frac{(\tilde{O}(1))^2 W_i^2 (n_{\widehat{x}}^i + n_{\widehat{u}}^i)^3}{\sigma_\eta^4 \epsilon^2} ||\widehat{Q}_i^{\text{true}}||^2, \frac{\tilde{O}(1) W_i^2 (n_{\widehat{x}}^i)^2 (n_{\widehat{x}}^i + n_{\widehat{u}}^i)^2}{\sigma_\eta^4} \right) \\ &\leq \max \left(\frac{(\tilde{O}(1))^2 W_i^2 (n_{\widehat{x}}^i + n_{\widehat{u}}^i)^3}{\sigma_\eta^4 \epsilon^2} \left(\sum_{j \in \mathcal{I}_{\text{GD}}^i} \left\|Q_j^{\text{true}}\right\| \right)^2, \frac{\tilde{O}(1) W_i^2 (n_{\widehat{x}}^i)^2 (n_{\widehat{x}}^i + n_{\widehat{u}}^i)^2}{\sigma_\eta^4} \right) \end{aligned} \tag{44}$$

where $W_i = ||K^{\text{play}}_{\mathcal{I}^i_{\widehat{Q}}}||_+^2 \sigma_w \bar{\sigma}_i \dfrac{\tau^2 ||K_{\mathcal{I}^i_{\widehat{Q}}}||_+^4 (||A_{\mathcal{I}^i_{\widehat{Q}}}||^2 + ||B_{\mathcal{I}^i_{\widehat{Q}}}||^2)}{\rho^2(1-\rho^2)}$. We now provide an example on choosing the relative estimation weights w_j to achieve better sample efficiency in the indirect case compared to the direct case. Letting $w_j = ||Q_j^{\text{true}}|| \left(\sum_{j \in \mathcal{I}_{\text{GD}}^i} \left\|Q_j^{\text{true}}\right\|\right)^{-1}$ yields

$$T_{\text{indirect}} \leq \max_{j \in \mathcal{I}_{\text{GD}}^i} \left(\max \left(\frac{(\tilde{O}(1))^2 W_j^2 (n_x^j + n_u^j)^3}{\sigma_\eta^4 \epsilon^2} \left(\sum_{j \in \mathcal{I}_{\text{GD}}^i} \left\|Q_j^{\text{true}}\right\| \right)^2, \frac{\tilde{O}(1) W_j^2 (n_x^j)^2 (n_x^j + n_u^j)^2}{\sigma_\eta^4} \right) \right). \tag{45}$$

Note that the RHS in (44) is equal to (45) only if there exists $j \in \mathcal{I}_{\text{GD}}^i$ such that $\mathcal{I}_Q^j = \mathcal{I}_{\widehat{Q}}^i$, otherwise (44) is strictly greater. Thus, for any $i \in \mathcal{V}$, $\forall\ j \in \mathcal{I}_{GD}^i$, $w_j = ||Q_j^{\text{true}}|| \left(\sum_{j \in \mathcal{I}_{\text{GD}}^i} \left\|Q_j^{\text{true}}\right\|\right)^{-1}$ ensures that the worst case sample complexity of the indirect decomposition based Algorithm 1 is equal to the direct case and strictly better if $\mathcal{I}_Q^j \subset \mathcal{I}_{\widehat{Q}}^i$, $\forall\ j \in \mathcal{I}_{\text{GD}}^i$. We also note that the choice of w_j is not unique. Finding optimal weights w.r.t. the overall sample efficiency is a problem in its own interest and deferred to possible future work.

Appendix I. Illustration of the 10-agent leader follower network in the simulation example