Supplementary Materials

A Convergence Analysis of Two Time-scale Linear TDC

We first provide the following lemma that is useful for the proof of Theorem 1, which is proved in [Xu et al., 2020a]. Throughout the paper, for two matrices $M, N \in \mathbb{R}^{d\times d}$, we define $\langle M, N\rangle = \sum_{i=1}^{d}\sum_{j=1}^{d} M_{i,j}N_{i,j}$.

Lemma 2. *Consider a sequence $\{s_t\}_{t\geq 0}$ generated by the MDP defined in Section 2. Suppose Assumption 2 holds. Let $X(s)$ be either a matrix or a vector that satisfies the following conditions:*

$$\|X(s)\|_2 \text{ (vector) or } \|X(S)\|_F \text{ (matrix)} \leq C_x \quad \text{for all} \quad s\in\mathcal{S},$$

and

$$\mathbb{E}_\nu[X(s)] = \widetilde{X}.$$

For any $t_0 \geq 0$ and $M > 0$, define $X(\mathcal{M}) = \frac{1}{M}\sum_{i=t_0}^{t_0+M-1} X(s_i)$. We have

$$\mathbb{E}\left[\left\|X(\mathcal{M}) - \widetilde{X}\right\|_2^2\right] \leq \frac{8C_x^2[1+(\kappa-1)\rho]}{(1-\rho)M}.$$

We next proceed to prove Theorem 1.

Proof of Theorem 1. We define $w^*(\theta) = -C^{-1}(A\theta + b)$, $\theta^* = -A^{-1}b$, and

$$g(\theta_t) = (A - BC^{-1}A)\theta_t + (b - BC^{-1}b), \tag{15}$$
$$f(w_t) = C(w_t - w^*(\theta_t)), \tag{16}$$

where $B = -\gamma\mathbb{E}[\mathbb{E}_\pi[\phi(s')|s]\phi(s)^\top]$. We further define

$$g_t(\theta_t) = (A_t - B_tC^{-1}A)\theta_t + (b_t - B_tC^{-1}b), \tag{17}$$
$$f_t(w_t) = C_t(w_t - w^*(\theta_t)), \tag{18}$$
$$h_t(\theta_t) = (A_t - C_tC^{-1}A)\theta_t + (b_t - C_tC^{-1}b), \tag{19}$$

where $A_t = (\gamma\rho(s_t,a_t)\phi(s_{t+1}) - \phi(s_t))\phi(s_t)$, $B_t = -\gamma\rho(s_t,a_t)\phi(s_{t+1})\phi(s_t)^\top$, $C_t = -\phi(s_t)\phi(s_t)^\top$ and $b_t = \rho(s_t,a_t)r(s_t,a_t,s_{t+1})\phi(s_t)$. The update of two time-scale linear TDC (line 5-6 of Algorithm 1) can be rewritten as

$$\theta_{t+1} = \theta_t + \alpha[g_t(\theta_t) + B_t(w_t - w^*(\theta_t))], \tag{20}$$
$$w_{t+1} = w_t + \beta[f_t(w_t) + h_t(\theta_t)]. \tag{21}$$

Considering the iteration of w_t, we proceed as follows:

$$\begin{aligned}
&\|w_{t+1} - w^*(\theta_t)\|_2^2 \\
&= \|w_t + \beta[f_t(w_t) + h_t(\theta_t)] - w^*(\theta_t)\|_2^2 \\
&= \|w_t - w^*(\theta_t)\|_2^2 + 2\beta\langle w_t - w^*(\theta_t), f_t(w_t)\rangle + 2\beta\langle w_t - w^*(\theta_t), h_t(w_t)\rangle \\
&\quad + \beta^2\|f_t(w_t) + h_t(\theta_t)\|_2^2 \\
&= \|w_t - w^*(\theta_t)\|_2^2 + 2\beta\langle w_t - w^*(\theta_t), f(w_t)\rangle + 2\beta\langle w_t - w^*(\theta_t), f_t(w_t) - f(w_t)\rangle \\
&\quad + 2\beta\langle w_t - w^*(\theta_t), h_t(w_t)\rangle + \beta^2\|f_t(w_t) + h_t(\theta_t)\|_2^2 \\
&\overset{(i)}{\leq} (1 - 2\lambda_2\beta)\|w_t - w^*(\theta_t)\|_2^2 + 2\beta\left[\frac{\lambda_2}{4}\|w_t - w^*(\theta_t)\|_2^2 + \frac{1}{\lambda_2}\|f_t(w_t) - f(w_t)\|_2^2\right] \\
&\quad + 2\beta\left[\frac{\lambda_2}{4}\|w_t - w^*(\theta_t)\|_2^2 + \frac{1}{\lambda_2}\|h_t(\theta_t)\|_2^2\right] + 2\beta^2\|f_t(w_t)\|_2^2 + 2\beta^2\|h_t(\theta_t)\|_2^2 \\
&\overset{(ii)}{\leq} (1 - \lambda_2\beta + 2\beta^2)\|w_t - w^*(\theta_t)\|_2^2 + \frac{2\beta}{\lambda_2}\|f_t(w_t) - f(w_t)\|_2^2 + \left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\|h_t(\theta_t)\|_2^2,
\end{aligned} \tag{22}$$

where (i) follows from the fact that $\langle w_t - w^*(\theta_t), f(w_t)\rangle = \langle w_t - w^*(\theta_t), C(w_t - w^*(\theta_t))\rangle \le -\lambda_2 \|w_t - w^*(\theta_t)\|_2^2$ and Young's inequality, (ii) follows from the fact that $\|f_t(w_t)\|_2 = \|C_t(w_t - w^*(\theta_t))\|_2 \le \|C_t\|_2 \|w_t - w^*(\theta_t)\|_2 \le \|w_t - w^*(\theta_t)\|_2$. Taking expectation conditioned on $\mathcal{F}_t$ on both sides of eq. (22) yields

$$
\begin{aligned}
&\mathbb{E}[\|w_{t+1} - w^*(\theta_t)\|_2^2 | \mathcal{F}_t] \\
&\le (1 - \lambda_2\beta + 2\beta^2) \|w_t - w^*(\theta_t)\|_2^2 + \frac{2\beta}{\lambda_2}\mathbb{E}[\|f_t(w_t) - f(w_t)\|_2^2 | \mathcal{F}_t] \\
&\quad + \left(\frac{2\beta}{\lambda_2} + 2\beta^2\right) \mathbb{E}[\|h_t(\theta_t)\|_2^2 | \mathcal{F}_t] \\
&\overset{(i)}{\le} \left[1 - \lambda_2\beta + 2\beta^2 + \frac{16\beta}{\lambda_2}\frac{1 + (\kappa - 1)\rho}{(1-\rho)M}\right] \|w_t - w^*(\theta_t)\|_2^2 \\
&\quad + 128\left(1 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M} \|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M} \\
&\overset{(ii)}{\le} \left(1 - \frac{\lambda_2\beta}{2}\right)\|w_t - w^*(\theta_t)\|_2^2 + 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M},
\end{aligned} \tag{23}
$$

where (i) follows from the facts that

$$
\begin{aligned}
\mathbb{E}[\|f_t(w_t) - f(w_t)\|_2^2 | \mathcal{F}_t] &= \mathbb{E}[\|(C_t - C)(w_t - w^*(\theta_t))\|_2^2 | \mathcal{F}_t] \\
&\le \mathbb{E}[\|(C_t - C)\|_2^2 | \mathcal{F}_t] \|w_t - w^*(\theta_t)\|_2^2 \\
&\overset{(a)}{\le} \frac{8[1 + (\kappa - 1)\rho]}{(1-\rho)M} \|w_t - w^*(\theta_t)\|_2^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}[\|h_t(\theta_t)\|_2^2 | \mathcal{F}_t] \\
&= \mathbb{E}[\left\|(A_t - C_tC^{-1}A)\theta_t + (b_t - C_tC^{-1}b)\right\|_2^2 | \mathcal{F}_t] \\
&= \mathbb{E}[\left\|(A_t - A)(\theta_t - \theta^*) + (A_t - A)\theta^* + b_t - b + (C - C_t)C^{-1}A(\theta_t - \theta^*)\right\|_2^2 | \mathcal{F}_t] \\
&\le 4\mathbb{E}[\|(A_t - A)(\theta_t - \theta^*)\|_2^2 | \mathcal{F}_t] + 4\mathbb{E}[\|(A_t - A)\theta^*\|_2^2 | \mathcal{F}_t] + 4\mathbb{E}[\|b_t - b\|_2^2 | \mathcal{F}_t] \\
&\quad + 4\mathbb{E}[\left\|(C - C_t)C^{-1}A(\theta_t - \theta^*)\right\|_2^2 | \mathcal{F}_t] \\
&\le 4\mathbb{E}[\|A_t - A\|_2^2 | \mathcal{F}_t] \|\theta_t - \theta^*\|_2^2 + 4\mathbb{E}[\|A_t - A\|_2^2 | \mathcal{F}_t] \|\theta^*\|_2^2 + 4\mathbb{E}[\|b_t - b\|_2^2 | \mathcal{F}_t] \\
&\quad + 4\mathbb{E}[\|(C - C_t)\|_2^2 | \mathcal{F}_t] \left\|C^{-1}\right\|_2^2 \|A\|_2^2 \|\theta_t - \theta^*\|_2^2 \\
&\overset{(b)}{\le} 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M},
\end{aligned}
$$

where (a) and (b) follow from Lemma 2 and the fact that $\|\theta^*\|_2 \le R_\theta$, where $R_\theta = \frac{r_{\max}}{\lambda_1}$, and (ii) follows from the fact that $\beta \le \frac{\lambda_2}{4}$ and $M \ge \frac{64[1+(\kappa-1)\rho]}{\lambda_2^2(1-\rho)}$. Then, we upper bound the term $\mathbb{E}[\|w_{t+1} - w^*(\theta_{t+1})\|_2^2 | \mathcal{F}_t]$ as follows:

$$
\begin{aligned}
&\mathbb{E}[\|w_{t+1} - w^*(\theta_{t+1})\|_2^2 | \mathcal{F}_t] \\
&\le \left(1 + \frac{1}{2(2/(\lambda_2\beta) - 1)}\right)\mathbb{E}[\|w_{t+1} - w^*(\theta_t)\|_2^2] + (1 + 2(2/(\lambda_2\beta) - 1))\mathbb{E}[\|w^*(\theta_{t+1}) - w^*(\theta_t)\|_2^2] \\
&\overset{(i)}{\le} \left(\frac{4/(\lambda_2\beta) - 1}{4/(\lambda_2\beta) - 2}\right)\left(1 - \frac{\lambda_2\beta}{2}\right)\|w_t - w^*(\theta_t)\|_2^2 + \frac{8}{\lambda_2^2\beta}\mathbb{E}[\|\theta_{t+1} - \theta_t\|_2^2] \\
&\quad + 128\left(\frac{4/(\lambda_2\beta) - 1}{4/(\lambda_2\beta) - 2}\right)\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1 + (\kappa - 1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2
\end{aligned}
$$

$$
\begin{aligned}
&\quad + 32\left(\frac{4/(\lambda_2\beta)-1}{4/(\lambda_2\beta)-2}\right)(4R_\theta^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M} \\
&\overset{(ii)}{\leq} \left(1 - \frac{\lambda_2\beta}{4}\right)\|w_t - w^*(\theta_t)\|_2^2 + \frac{8}{\lambda_2^2\beta}\mathbb{E}[\|\theta_{t+1} - \theta_t\|_2^2] \\
&\quad + 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M} \\
&\leq \left(1 - \frac{\lambda_2\beta}{4}\right)\|w_t - w^*(\theta_t)\|_2^2 + \frac{16\alpha^2}{\lambda_2^2\beta}\mathbb{E}[\|B_t(w_t - w^*(\theta_t))\|_2^2] + \frac{16\alpha^2}{\lambda_2^2\beta}\mathbb{E}[\|g_t(\theta_t)\|_2^2] \\
&\quad + 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M} \\
&\leq \left(1 - \frac{\lambda_2\beta}{4} + \frac{16\rho_{\max}^2\alpha^2}{\lambda_2^2\beta}\right)\|w_t - w^*(\theta_t)\|_2^2 + \frac{32\alpha^2}{\lambda_2^2\beta}\mathbb{E}[\|g_t(\theta_t) - g(\theta_t)\|_2^2] \\
&\quad + \frac{32\alpha^2}{\lambda_2^2\beta}\mathbb{E}[\|g(\theta_t)\|_2^2] + 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M} \\
&\overset{(iii)}{\leq} \left(1 - \frac{\lambda_2\beta}{4} + \frac{16\rho_{\max}^2\alpha^2}{\lambda_2^2\beta}\right)\|w_t - w^*(\theta_t)\|_2^2 \\
&\quad + \frac{32\alpha^2}{\lambda_2^2\beta}\left[128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\right] \\
&\quad + \frac{64\alpha^2}{\lambda_2^2\beta}\|\theta_t - \theta^*\|_2^2 + 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M} \\
&\leq \left(1 - \frac{\lambda_2\beta}{4} + \frac{16\rho_{\max}^2\alpha^2}{\lambda_2^2\beta}\right)\|w_t - w^*(\theta_t)\|_2^2 + \left(\frac{96\alpha^2}{\lambda_2^2\beta} + \frac{\lambda_1\alpha}{4}\right)\|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M},
\end{aligned} \tag{24}
$$

where (i) follows Yong's inequality, (ii) follows from the fact that $\beta \leq \min\left\{\frac{1}{8\lambda_2}, \frac{\lambda_2}{4}\right\}$, and (iii) follows from the fact that

$$
\begin{aligned}
&\mathbb{E}[\|g_t(\theta_t) - g(\theta_t)\|_2^2\,|\mathcal{F}_t] \\
&= \mathbb{E}[\left\|(A_t - A)(\theta_t - \theta^*) + (A_t - A)\theta^* + (b_t - b) + (B - B_t)C^{-1}A(\theta_t - \theta^*)\right\|_2^2\,|\mathcal{F}_t] \\
&\leq 4\mathbb{E}[\|A_t - A\|_2^2\,|\mathcal{F}_t]\,\|\theta_t - \theta^*\|_2^2 + 4\mathbb{E}[\|A_t - A\|_2^2\,|\mathcal{F}_t]\,\|\theta^*\|_2^2 + 4\mathbb{E}[\|b_t - b\|_2^2\,|\mathcal{F}_t] \\
&\quad + 4\mathbb{E}[\|B - B_t\|_2^2\,|\mathcal{F}_t]\,\left\|C^{-1}\right\|_2^2\|A\|_2^2\|\theta_t - \theta^*\|_2^2 \\
&\overset{(a)}{\leq} 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\|\theta_t - \theta^*\|_2^2 + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\frac{1+(\kappa-1)\rho}{(1-\rho)M},
\end{aligned} \tag{25}
$$

where (a) follows from Lemma 2, and (ii) follows from the fact that $M \geq 128\left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right)\frac{1+(\kappa-1)\rho}{1-\rho}\max\{1, \frac{8\beta+8\lambda_2\beta^2}{\lambda_1\lambda_2\alpha}\}$. Considering the iterate of θ_t, we proceed as follows:

$$
\begin{aligned}
&\|\theta_{t+1} - \theta^*\|_2^2 \\
&= \|\theta_t + \alpha[g_t(\theta_t) + B_t(w_t - w^*(\theta_t))] - \theta^*\|_2^2
\end{aligned}
$$

$$
\begin{aligned}
&= \|\theta_t - \theta^*\|_2^2 + 2\alpha\langle\theta_t - \theta^*, g_t(\theta_t)\rangle + 2\alpha\langle\theta_t - \theta^*, B_t(w_t - w^*(\theta_t))\rangle \\
&\quad + \alpha^2 \|g_t(\theta_t) + B_t(w_t - w^*(\theta_t))\|_2^2 \\
&= \|\theta_t - \theta^*\|_2^2 + 2\alpha\langle\theta_t - \theta^*, g(\theta_t)\rangle + 2\alpha\langle\theta_t - \theta^*, g_t(\theta_t) - g(\theta_t)\rangle \\
&\quad + 2\alpha\langle\theta_t - \theta^*, B_t(w_t - w^*(\theta_t))\rangle + \alpha^2 \|g_t(\theta_t) + B_t(w_t - w^*(\theta_t))\|_2^2 \\
&\overset{(i)}{\leq} (1 - 2\lambda_1\alpha) \|\theta_t - \theta^*\|_2^2 + 2\alpha \left[\frac{\lambda_1}{4} \|\theta_t - \theta^*\|_2^2 + \frac{1}{\lambda_1} \|g_t(\theta_t) - g(\theta_t)\|_2^2\right] \\
&\quad + 2\alpha \left[\frac{\lambda_1}{4} \|\theta_t - \theta^*\|_2^2 + \frac{1}{\lambda_1} \|B_t(w_t - w^*(\theta_t))\|_2^2\right] + 3\alpha^2 \|g(\theta_t)\|_2^2 \\
&\quad + 3\alpha^2 \|g_t(\theta_t) - g(\theta_t)\|_2^2 + 3\alpha^2 \|B_t(w_t - w^*(\theta_t))\|_2^2 \\
&\overset{(ii)}{\leq} (1 - \lambda_1\alpha) \|\theta_t - \theta^*\|_2^2 + \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|g_t(\theta_t) - g(\theta_t)\|_2^2 \\
&\quad + \rho_{\max}^2 \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|w_t - w^*(\theta_t)\|_2^2 + 3\alpha^2 \|g(\theta_t)\|_2^2 \\
&\overset{(iii)}{\leq} \left(1 - \lambda_1\alpha + \frac{3\alpha^2}{\lambda_2}\right) \|\theta_t - \theta^*\|_2^2 + \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|g_t(\theta_t) - g(\theta_t)\|_2^2 \\
&\quad + \rho_{\max}^2 \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|w_t - w^*(\theta_t)\|_2^2, \qquad (26)
\end{aligned}
$$

where (i) follows from the fact that $\langle\theta_t - \theta^*, g(\theta_t)\rangle = \langle\theta_t - \theta^*, A^\top C^{-1} A(\theta_t - \theta^*)\rangle \leq -\lambda_1 \|\theta_t - \theta^*\|_2$ and Young's inequality, (ii) follows from the fact that $\|B_t(w_t - w^*(\theta_t))\|_2 \leq \|B_t\|_2 \|w_t - w^*(\theta_t)\|_2 \leq \rho_{\max} \|w_t - w^*(\theta_t)\|_2$, and (iii) follows from the fact that

$$
\|g(\theta_t)\|_2 = \left\|A^\top C^{-1} A(\theta_t - \theta^*)\right\|_2 \leq \left\|A^\top\right\|_2 \left\|C^{-1}\right\|_2 \|A\|_2 \|\theta_t - \theta^*\|_2 \leq \frac{1}{\lambda_2} \|\theta_t - \theta^*\|_2 .
$$

Taking expectation conditioned on $\mathcal{F}_t$ on both sides of eq. (26) yields

$$
\begin{aligned}
&\mathbb{E}[\|\theta_{t+1} - \theta^*\|_2^2 \,|\mathcal{F}_t] \\
&\leq \left(1 - \lambda_1\alpha + \frac{3\alpha^2}{\lambda_2}\right) \|\theta_t - \theta^*\|_2^2 + \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \mathbb{E}[\|g_t(\theta_t) - g(\theta_t)\|_2^2 \,|\mathcal{F}_t] \\
&\quad + \rho_{\max}^2 \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|w_t - w^*(\theta_t)\|_2^2 \\
&\overset{(i)}{\leq} \left[1 - \lambda_1\alpha + \frac{3\alpha^2}{\lambda_2} + 128 \left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right) \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \frac{1 + (\kappa - 1)\rho}{(1 - \rho)M}\right] \|\theta_t - \theta^*\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2) \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \frac{1 + (\kappa - 1)\rho}{(1 - \rho)M} + \rho_{\max}^2 \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|w_t - w^*(\theta_t)\|_2^2 \\
&\overset{(ii)}{\leq} \left(1 - \frac{3}{4}\lambda_1\alpha + \frac{3\alpha^2}{\lambda_2}\right) \|\theta_t - \theta^*\|_2^2 + \rho_{\max}^2 \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \|w_t - w^*(\theta_t)\|_2^2 \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2) \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right) \frac{1 + (\kappa - 1)\rho}{(1 - \rho)M}, \qquad (27)
\end{aligned}
$$

where (i) follows from eq. (25) and (ii) follows from the fact that $M \geq 128 \left(\rho_{\max}^2 + \frac{1}{\lambda_2^2}\right) \frac{1+(\kappa-1)\rho}{1-\rho} \max\{1, \frac{8+12\lambda_1\alpha}{\lambda_1}\}$. Combining eq. (23) and eq. (27) yields

$$
\begin{aligned}
&\mathbb{E}[\|w_{t+1} - w^*(\theta_{t+1})\|_2^2 \,|\mathcal{F}_t] + \mathbb{E}[\|\theta_{t+1} - \theta^*\|_2^2 \,|\mathcal{F}_t] \\
&\leq \left[1 - \frac{\lambda_2\beta}{4} + \frac{16\rho_{\max}^2\alpha^2}{\lambda_2^2\beta} + \rho_{\max}^2 \left(\frac{2\alpha}{\lambda_1} + 3\alpha^2\right)\right] \|w_t - w^*(\theta_t)\|_2^2 \\
&\quad + \left(1 - \frac{1}{2}\lambda_1\alpha + \frac{3\alpha^2}{\lambda_2} + \frac{96\alpha^2}{\lambda_2^2\beta}\right) \|\theta_t - \theta^*\|_2^2
\end{aligned}
$$

$$+ 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2 + \frac{2\alpha}{\lambda_1} + 3\alpha^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}.$$

If we further let

$$\alpha \le \min\left\{\frac{1}{8\lambda_1}, \frac{\lambda_1\lambda_2}{12}, \frac{\sqrt{\lambda_2\beta}}{4\sqrt{6}\rho_{\max}}\frac{\lambda_2\sqrt{\lambda_2}\beta}{16\rho_{\max}^2}, \frac{\lambda_1\lambda_2\beta}{64\rho_{\max}^2}, \frac{\lambda_1\lambda_2^2\beta}{768}\right\}, \quad \beta \le \frac{1}{8\lambda_2}.$$

We have

$$\begin{aligned}
&\mathbb{E}[\|w_{t+1} - w^*(\theta_{t+1})\|_2^2 \,|\mathcal{F}_t] + \mathbb{E}[\|\theta_{t+1} - \theta^*\|_2^2 \,|\mathcal{F}_t] \\
&\le \left(1 - \frac{1}{8}\min\{\lambda_2\beta, \lambda_1\alpha\}\right)\left(\|w_t - w^*(\theta_t)\|_2^2 + \|\theta_t - \theta^*\|_2^2\right) \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2 + \frac{2\alpha}{\lambda_1} + 3\alpha^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}. \qquad (28)
\end{aligned}$$

Taking expectation on both sides of eq. (28) and applying the relation recursively from $t = T-1$ to 0 yield

$$\begin{aligned}
&\mathbb{E}[\|w_T - w^*(\theta_T)\|_2^2] + \mathbb{E}[\|\theta_T - \theta^*\|_2^2] \\
&\le \left(1 - \frac{1}{8}\min\{\lambda_2\beta, \lambda_1\alpha\}\right)^T\left(\|w_0 - w^*(\theta_0)\|_2^2 + \|\theta_0 - \theta^*\|_2^2\right) \\
&\quad + 32(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2 + \frac{2\alpha}{\lambda_1} + 3\alpha^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\sum_{t=0}^{T-1}\left(1 - \frac{1}{8}\min\{\lambda_2\beta, \lambda_1\alpha\}\right)^t \\
&\le \left(1 - \frac{1}{8}\min\{\lambda_2\beta, \lambda_1\alpha\}\right)^T\left(\|w_0 - w^*(\theta_0)\|_2^2 + \|\theta_0 - \theta^*\|_2^2\right) \\
&\quad + \frac{256(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)}{\min\{\lambda_2\beta, \lambda_1\alpha\}}\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2 + \frac{2\alpha}{\lambda_1} + 3\alpha^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}, \qquad (29)
\end{aligned}$$

which implies

$$\begin{aligned}
\mathbb{E}[\|\theta_T - \theta^*\|_2^2] &\le \left(1 - \frac{1}{8}\min\{\lambda_2\beta, \lambda_1\alpha\}\right)^T\left(\|w_0 - w^*(\theta_0)\|_2^2 + \|\theta_0 - \theta^*\|_2^2\right) \\
&\quad + \frac{256(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)}{\min\{\lambda_2\beta, \lambda_1\alpha\}}\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2 + \frac{2\alpha}{\lambda_1} + 3\alpha^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}. \qquad (30)
\end{aligned}$$

□

B Convergence Analysis of Two Time-scale Nonlinear TDC

Before we present our technical proof of Theorem 2, we first introduce some notations and definitions. Recall that

$$\frac{1}{2}\nabla J(\theta_t) = -\mathbb{E}[\delta(\theta_t)\phi_{\theta_t}(s)] - \gamma\mathbb{E}[\phi_{\theta_t}(s')\phi_{\theta_t}(s)^\top]w(\theta_t) + h(\theta_t, w(\theta_t)).$$

For any $x_j = (s_j, a_j, s_{j+1})$, we define

$$g(\theta_t, w, x_j) = -\delta_j(\theta_t)\phi_{\theta_t}(s_j) - \gamma\phi_{\theta_t}(s_{j+1})\phi_{\theta_t}(s_j)^\top w + h(\theta_t, w, x_j),$$

where $h(\theta_t, w, x_j) = (\delta_j(\theta_t) - \phi_{\theta_t}(s_j)^\top w)\nabla_{\theta_t}^2 V_{\theta_t}(s_j)w$. We also define the mini-batch gradient estimator as $g(\theta_t, w, \mathcal{B}_t) = \frac{1}{M}\sum_{j\in\mathcal{B}_t} g(\theta_t, w, x_j)$, where $\mathcal{B}_t = \{i_t, i_t+1, \cdots, i_t+M-1\}$.

For critic's update, we define $A_{\theta_t,x_j} = -\phi_{\theta_t}(s_j)\phi_{\theta_t}(s_j)^\top$, $b_{\theta_t,x_j} = \delta_j(\theta_t)\phi_{\theta_t}(s_j)$, $A_{\theta_t,\mathcal{B}_k} = -\frac{1}{M}\sum_{j=i_k}^{i_k+M-1}\phi_{\theta_t}(s_j)\phi_{\theta_t}(s_j)^\top$, $b_{\theta_t,\mathcal{B}_k} = \frac{1}{M}\sum_{j=i_k}^{i_k+M-1}\delta_j(\theta_t)\phi_{\theta_t}(s_j)$, $A_{\theta_t} = -\mathbb{E}_{\mu_\pi}[\phi_{\theta_t}(s)\phi_{\theta_t}(s)^\top]$ and $b_{\theta_t} = \mathbb{E}_{\mu_\pi}[\delta(\theta_t)\phi_{\theta_t}(s)]$. We also define $f_{\theta_t}(w_k', x_j) = A_{\theta_t,x_j}w_k' + b_{\theta_t,x_j}$, $f_{\theta_t}(w_k', \mathcal{B}_k) = A_{\theta_t,\mathcal{B}_k}w_k' + b_{\theta_t,\mathcal{B}_k}$ and $f_{\theta_t}(w_k') = A_{\theta_t}w_k' + b_{\theta_t}$. It can be checked easily that for all $\theta \in \mathbb{R}^d$, we have $w(\theta) \le R_w$, where $R_w = \frac{C_\phi(r_{\max}+2C_v)}{\lambda_v}$.

B.1 Preliminaries

In this subsection, we provide some supporting lemmas, which are useful to the proof of Theorem 2.

Lemma 3. *Suppose Assumptions 1-5 hold. For any $t \geq 0$ and j, we have $\|g(\theta_t, w(\theta_t), x_j)\|_2 \leq C_g$, where*

$$C_g = [r_{\max} + (\gamma+1)C_v]C_\phi + \gamma C_\phi^2 R_w + [r_{\max} + (\gamma+1)C_v + C_\phi R_w]D_v R_w.$$

Proof. According to the definition of $g(\theta_t, w_t, x_j)$, we have

$$\begin{aligned}
\|g(\theta_t, w(\theta_t), x_j)\|_2 &\leq \|\delta_j(\theta_t)\phi_{\theta_t}(s_j)\|_2 + \gamma \left\|\phi_{\theta_t}(s_{j+1})\phi_{\theta_t}(s_j)^\top w(\theta_t)\right\|_2 \\
&\quad + \left\|(\delta_j(\theta_t) - \phi_{\theta_t}(s_j)^\top w(\theta_t)\nabla^2_{\theta_t} V_{\theta_t}(s_j) w(\theta_t)\right\|_2 \\
&\leq |\delta_j(\theta_t)| \, \|\phi_{\theta_t}(s_j)\|_2 + \gamma \|\phi_{\theta_t}(s_{j+1})\|_2 \|\phi_{\theta_t}(s_j)\|_2 \|w(\theta_t)\|_2 \\
&\quad + \left(|\delta_j(\theta_t)| + \left|\phi_{\theta_t}(s_j)^\top w(\theta_t)\right| \right) \left\|\nabla^2_{\theta_t} V_{\theta_t}(s_j)\right\|_2 \|w(\theta_t)\|_2 \\
&\overset{(i)}{\leq} [r_{\max} + (\gamma+1)C_v]C_\phi + \gamma C_\phi^2 R_w + [r_{\max} + (\gamma+1)C_v + C_\phi R_w]D_v R_w.
\end{aligned}$$

where (i) follows from the fact that $w(\theta_t) \leq R_w$. ☐

Lemma 4. *Suppose Assumptions 1-5 hold, for any $\theta, \theta' \in \mathbb{R}^d$, we have $\|w(\theta) - w(\theta')\|_2 \leq L_w \|\theta - \theta'\|_2$, where $L_w = \left\{ \frac{2C_\phi L_\phi}{\lambda_v^2}[r_{\max} + (1+\gamma)C_v] + \frac{1}{\lambda_v}[L_v C_\phi(1+\gamma) + L_\phi(r_{\max} + (1+\gamma)C_v)] \right\}$.*

Proof. According to the definition of $w(\theta)$, we have

$$\begin{aligned}
\|w(\theta) - w(\theta')\|_2 &= \left\|A_\theta^{-1} b_\theta - A_{\theta'}^{-1} b_{\theta'}\right\|_2 = \left\|A_\theta^{-1} b_\theta - A_{\theta'}^{-1} b_\theta + A_{\theta'}^{-1} b_\theta - A_{\theta'}^{-1} b_{\theta'}\right\|_2 \\
&\leq \left\|A_\theta^{-1} b_\theta - A_{\theta'}^{-1} b_\theta\right\|_2 + \left\|A_{\theta'}^{-1} b_\theta - A_{\theta'}^{-1} b_{\theta'}\right\|_2 \\
&= \left\|A_{\theta'}^{-1} A_{\theta'} A_\theta^{-1} b_\theta - A_{\theta'}^{-1} A_\theta A_\theta^{-1} b_\theta\right\|_2 + \left\|A_{\theta'}^{-1} b_\theta - A_{\theta'}^{-1} b_{\theta'}\right\|_2 \\
&= \left\|A_{\theta'}^{-1}(A_{\theta'} - A_\theta) A_\theta^{-1} b_\theta\right\|_2 + \left\|A_{\theta'}^{-1}(b_\theta - b_{\theta'})\right\|_2 \\
&\leq \left\|A_{\theta'}^{-1}\right\|_2 \|A_{\theta'} - A_\theta\|_2 \left\|A_\theta^{-1}\right\|_2 \|b_\theta\|_2 + \left\|A_{\theta'}^{-1}\right\|_2 \|b_\theta - b_{\theta'}\|_2 \\
&\leq \frac{r_{\max} + (1+\gamma)C_v}{\lambda_v^2} \|A_{\theta'} - A_\theta\|_2 + \frac{1}{\lambda_v} \|b_\theta - b_{\theta'}\|_2 .
\end{aligned} \tag{31}$$

Considering the term $\|A_{\theta'} - A_\theta\|_2$, by definition we can obtain

$$\begin{aligned}
\|A_{\theta'} - A_\theta\|_2 &= \left\|\mathbb{E}[\phi_\theta \phi_\theta^\top] - \mathbb{E}[\phi_{\theta'} \phi_{\theta'}^\top]\right\|_2 = \left\|\mathbb{E}[\phi_\theta \phi_\theta^\top] - \mathbb{E}[\phi_{\theta'} \phi_\theta^\top] + \mathbb{E}[\phi_{\theta'} \phi_\theta^\top] - \mathbb{E}[\phi_{\theta'} \phi_{\theta'}^\top]\right\|_2 \\
&\leq \left\|\mathbb{E}[\phi_\theta \phi_\theta^\top] - \mathbb{E}[\phi_{\theta'} \phi_\theta^\top]\right\|_F + \left\|\mathbb{E}[\phi_{\theta'} \phi_\theta^\top] - \mathbb{E}[\phi_{\theta'} \phi_{\theta'}^\top]\right\|_F \\
&\leq 2\mathbb{E}[\|\phi_\theta - \phi_{\theta'}\|_2 \|\phi_\theta\|_2] \leq 2C_\phi L_\phi \|\theta - \theta'\|_2 .
\end{aligned} \tag{32}$$

Considering the term $\|b_\theta - b_{\theta'}\|_2$, by definition we obtain

$$\begin{aligned}
\|b_\theta - b_{\theta'}\|_2 &= \|\mathbb{E}[\delta(\theta)\phi_\theta] - \mathbb{E}[\delta(\theta')\phi_{\theta'}]\|_2 = \|\mathbb{E}[\delta(\theta)\phi_\theta] - \mathbb{E}[\delta(\theta')\phi_\theta] + \mathbb{E}[\delta(\theta')\phi_\theta] - \mathbb{E}[\delta(\theta')\phi_{\theta'}]\|_2 \\
&\leq \|\mathbb{E}[\delta(\theta)\phi_\theta] - \mathbb{E}[\delta(\theta')\phi_\theta]\|_2 + \|\mathbb{E}[\delta(\theta')\phi_\theta] - \mathbb{E}[\delta(\theta')\phi_{\theta'}]\|_2 \\
&\leq \mathbb{E}[|\delta(\theta) - \delta(\theta')| \, \|\phi_\theta\|_2] + \mathbb{E}[|\delta(\theta')| \, \|\phi_{\theta'} - \phi_\theta\|_2] \\
&= \mathbb{E}[|(\gamma V(s', \theta) - V(s, \theta)) - (\gamma V(s', \theta') - V(s, \theta'))| \, \|\phi_\theta\|_2] + \mathbb{E}[|\delta(\theta')| \, \|\phi_{\theta'} - \phi_\theta\|_2] \\
&\leq [L_v C_\phi(1+\gamma) + L_\phi(r_{\max} + (1+\gamma)C_v)] \|\theta - \theta'\|_2 .
\end{aligned} \tag{33}$$

Substituting eq. (32) and eq. (33) into eq. (31) yields

$$\begin{aligned}
&\|w(\theta) - w(\theta')\|_2 \\
&\leq \left\{ \frac{2C_\phi L_\phi}{\lambda_v^2}[r_{\max} + (1+\gamma)C_v] + \frac{1}{\lambda_v}[L_v C_\phi(1+\gamma) + L_\phi(r_{\max} + (1+\gamma)C_v)] \right\} \|\theta - \theta'\|_2 .
\end{aligned}$$

☐

Lemma 5. *Suppose Assumptions 1-5 hold. Consider the iteration of w_t in Algorithm 2. Let the stepsize $\beta \leq \min\{\frac{\lambda_v}{8C_\phi^4}, \frac{8}{\lambda_v}\}$ and $\alpha \leq \frac{\lambda_v}{8\sqrt{2}L_w L_e}\beta$ and the batch size $M \geq (\frac{1}{\lambda_v} + 2\beta)\frac{96C_\phi^4[1-(\kappa-1)\rho]}{\lambda_v(1-\rho)}$. For any $t > 0$, we have*

$$
\begin{aligned}
&\mathbb{E}[\|w_t - w(\theta_t)\|_2^2] \\
&\leq \left(1 - \frac{\lambda_v}{8}\beta\right)\mathbb{E}[\|w_{t-1} - w(\theta_{t-1})\|_2^2] + \frac{2L_w^2\alpha^2}{\lambda_v\beta}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2] + \frac{D_1[1+(\kappa-1)\rho]}{M(1-\rho)},
\end{aligned}
$$

where $D_1 = \frac{128L_w^2C_g^2\alpha^2}{\lambda_v\beta} + 4C_f^2\left(\frac{\beta}{\lambda_v} + 2\beta^2\right)$.

Proof. We proceed as follows:

$$
\begin{aligned}
&\|w_t - w(\theta_{t-1})\|_2^2 \\
&= \left\|w_{t-1} + \beta f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - w(\theta_{t-1})\right\|_2^2 \\
&= \|w_{t-1} - w(\theta_{t-1})\|_2^2 + 2\beta\langle w_{t-1} - w(\theta_{t-1}), f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t)\rangle + \beta^2\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t)\right\|_2^2 \\
&= \|w_{t-1} - w(\theta_{t-1})\|_2^2 + 2\beta\langle w_{t-1} - w(\theta_{t-1}), f_{\theta_{t-1}}(w_{t-1})\rangle \\
&\quad + 2\beta\langle w_{t-1} - w(\theta_{t-1}), f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\rangle \\
&\quad + \beta^2\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1}) + f_{\theta_{t-1}}(w_{t-1})\right\|_2^2 \\
&\overset{(i)}{\leq} (1 - 2\lambda_v\beta)\left\|w_{t-1} - w(\theta_{t-1})\right\|_2^2 + 2\beta\langle w_{t-1} - w(\theta_{t-1}), f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\rangle \\
&\quad + \beta^2\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1}) + f_{\theta_{t-1}}(w_{t-1})\right\|_2^2 \\
&\overset{(ii)}{\leq} (1 - 2\lambda_v\beta)\left\|w_{t-1} - w(\theta_{t-1})\right\|_2^2 + \lambda_v\beta\left\|w_{t-1} - w(\theta_{t-1})\right\|_2^2 + \frac{\beta}{\lambda_v}\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2^2 \\
&\quad + 2\beta^2\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2 + 2\beta^2\left\|f_{\theta_{t-1}}(w_{t-1})\right\|_2^2 \\
&\overset{(iii)}{=} (1 - \lambda_v\beta + 2C_\phi^4\beta^2)\left\|w_{t-1} - w(\theta_{t-1})\right\|_2^2 + \left(\frac{\beta}{\lambda_v} + 2\beta^2\right)\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2^2,
\end{aligned} \tag{34}
$$

where (i) follows from the fact that

$$
\begin{aligned}
\langle w_{t-1} - w(\theta_{t-1}), f_{\theta_{t-1}}(w_{t-1})\rangle &= \langle w_{t-1} - w(\theta_{t-1}), A_{\theta_{t-1}}(w_{t-1} - w(\theta_{t-1}))\rangle \\
&\leq -\lambda_v\left\|w_{t-1} - w(\theta_{t-1})\right\|_2^2,
\end{aligned}
$$

(ii) follows from the fact that $\langle a, b\rangle \leq \frac{\lambda_v}{2}a^2 + \frac{1}{2\lambda_v}b^2$, and (iii) follows from the fact that $\left\|f_{\theta_{t-1}}(w_{t-1})\right\|_2^2 = \left\|A_{\theta_{t-1}}(w_{t-1} - w(\theta_{t-1}))\right\|_2^2 \leq C_\phi^4\left\|w_{t-1} - w(\theta_{t-1})\right\|_2^2$. Taking expectation on both side of eq. (34) yields

$$
\begin{aligned}
&\mathbb{E}[\|w_t - w(\theta_{t-1})\|_2^2] \\
&\leq (1 - \lambda_v\beta + 2C_\phi^4\beta^2)\mathbb{E}[\|w_{t-1} - w(\theta_{t-1})\|_2^2] + \left(\frac{\beta}{\lambda_v} + 2\beta^2\right)\mathbb{E}\left[\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2^2\right].
\end{aligned} \tag{35}
$$

Next we bound the term $\mathbb{E}\left[\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2^2\right]$ in eq. (35) as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\left\|f_{\theta_{t-1}}(w_{t-1}, \mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2^2\right] \\
&= \mathbb{E}\left[\left\|(A_{\theta_{t-1},\mathcal{B}_{t-1}} - A_{\theta_{t-1}})w_{t-1} + b_{\theta_{t-1},\mathcal{B}_{t-1}} - b_{\theta_{t-1}}\right\|_2^2\right] \\
&= \mathbb{E}\left[\left\|(A_{\theta_{t-1},\mathcal{B}_{t-1}} - A_{\theta_{t-1}})(w_{t-1} - w(\theta_{t-1})) + (A_{\theta_{t-1},\mathcal{B}_{t-1}} - A_{\theta_{t-1}})w(\theta_{t-1}) + b_{\theta_{t-1},\mathcal{B}_{t-1}} - b_{\theta_{t-1}}\right\|_2^2\right] \\
&\leq 3\mathbb{E}\left[\left\|(A_{\theta_{t-1},\mathcal{B}_{t-1}} - A_{\theta_{t-1}})(w_{t-1} - w(\theta_{t-1}))\right\|_2^2\right] + 3\mathbb{E}\left[\left\|(A_{\theta_{t-1},\mathcal{B}_{t-1}} - A_{\theta_{t-1}})w(\theta_{t-1})\right\|_2^2\right] \\
&\quad + 3\mathbb{E}\left[\left\|b_{\theta_{t-1},\mathcal{B}_{t-1}} - b_{\theta_{t-1}}\right\|_2^2\right].
\end{aligned} \tag{36}
$$

From Assumption 2, we have $\left\|A_{\theta_t,x_j}\right\|_F \leq C_\phi^2$ and $\left\|b_{\theta_t,x_j}\right\|_2 \leq C_\phi(r_{\max}+2C_v)$. Following from Lemma 2, we can obtain the following two upper bounds:

$$\mathbb{E}\left[\left\|(A_{\theta_{t-1},\mathcal{B}_{t-1}} - A_{\theta_{t-1}})\right\|_2^2\right] \leq \frac{8C_\phi^4[1-(\kappa-1)\rho]}{(1-\rho)M}, \tag{37}$$

and

$$\mathbb{E}\left[\left\|b_{\theta_{t-1},\mathcal{B}_{t-1}} - b_{\theta_{t-1}}\right\|_2^2\right] \leq \frac{8C_\phi^2(r_{\max}+2C_v)^2[1-(\kappa-1)\rho]}{(1-\rho)M}. \tag{38}$$

Substituting eq. (37) and eq. (38) into eq. (36) yields

$$\begin{aligned}
&\mathbb{E}\left[\left\|f_{\theta_{t-1}}(w_{t-1},\mathcal{B}_t) - f_{\theta_{t-1}}(w_{t-1})\right\|_2^2\right] \\
&= \frac{24C_\phi^4[1-(\kappa-1)\rho]}{(1-\rho)M}\mathbb{E}[\|w_{t-1}-w(\theta_{t-1})\|_2^2] + \frac{24[C_\phi^2(r_{\max}+2C_v)^2 + C_\phi^4 R_w][1-(\kappa-1)\rho]}{(1-\rho)M}.
\end{aligned} \tag{39}$$

Substituting eq. (39) into eq. (34) yields

$$\begin{aligned}
&\mathbb{E}[\|w_t - w(\theta_{t-1})\|_2^2] \\
&\leq \left(1-\lambda_v\beta + 2C_\phi^4\beta^2 + \left(\frac{\beta}{\lambda_v}+2\beta^2\right)\frac{24C_\phi^4[1-(\kappa-1)\rho]}{(1-\rho)M}\right)\mathbb{E}[\|w_{t-1}-w(\theta_{t-1})\|_2^2] \\
&\quad + \left(\frac{\beta}{\lambda_v}+2\beta^2\right)\frac{24[C_\phi^2(r_{\max}+2C_v)^2 + C_\phi^4 R_w][1-(\kappa-1)\rho]}{(1-\rho)M} \\
&\overset{(i)}{\leq} \left(1-\frac{\lambda_v}{2}\beta\right)\mathbb{E}[\|w_{t-1}-w(\theta_{t-1})\|_2^2] + \left(\frac{\beta}{\lambda_v}+2\beta^2\right)\frac{4C_f[1-(\kappa-1)\rho]}{(1-\rho)M},
\end{aligned} \tag{40}$$

where (i) follows from the fact that $\beta \leq \frac{\lambda_v}{8C_\phi^4}$ and $M \geq (\frac{1}{\lambda_v}+2\beta)\frac{96C_\phi^4[1-(\kappa-1)\rho]}{\lambda_v(1-\rho)}$, and here we define $C_f = 6[C_\phi^2(r_{\max}+2C_v)^2 + C_\phi^4 R_w]$. By Young's inequality, we have

$$\begin{aligned}
&\mathbb{E}[\|w_t - w(\theta_t)\|_2^2] \\
&\leq \left(1+\frac{1}{2(2/(\lambda_v\beta)-1)}\right)\mathbb{E}[\|w_t-w(\theta_{t-1})\|_2^2] + (1+2(2/(\lambda_v\beta)-1))\mathbb{E}[\|w(\theta_{t-1})-w(\theta_t)\|_2^2] \\
&\overset{(i)}{\leq} \left(\frac{4/(\lambda_v\beta)-1}{4/(\lambda_v\beta)-2}\right)\mathbb{E}[\|w_{t-1}-w(\theta_{t-1})\|_2^2] + \frac{4}{\lambda_v\beta}\mathbb{E}[\|w(\theta_{t-1})-w(\theta_t)\|_2^2] \\
&\quad + \left(\frac{4/(\lambda_v\beta)-1}{4/(\lambda_v\beta)-2}\right)\left(\frac{\beta}{\lambda_v}+2\beta^2\right)\frac{4C_f^2[1+(\kappa-1)\rho]}{M(1-\rho)} \\
&\overset{(ii)}{\leq} \left(1-\frac{\lambda_v}{4}\beta\right)\mathbb{E}[\|w_{t-1}-w(\theta_{t-1})\|_2^2] + \frac{4L_w^2}{\lambda_v\beta}\mathbb{E}[\|\theta_{t-1}-\theta_t\|_2^2] \\
&\quad + \left(\frac{\beta}{\lambda_v}+2\beta^2\right)\frac{4C_f^2[1+(\kappa-1)\rho]}{M(1-\rho)} \\
&\leq \left(1-\frac{\lambda_v}{4}\beta\right)\mathbb{E}[\|w_{t-1}-w(\theta_{t-1})\|_2^2] + \frac{2L_w^2\alpha^2}{\lambda_v\beta}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2] \\
&\quad + \frac{8L_w^2\alpha^2}{\lambda_v\beta}\mathbb{E}\left[\left\|g(\theta_{t-1},w_{t-1},\mathcal{B}_{t-1}) - \frac{1}{2}\nabla J(\theta_{t-1})\right\|_2^2\right] + \left(\frac{\beta}{\lambda_v}+2\beta^2\right)\frac{4C_f^2[1+(\kappa-1)\rho]}{M(1-\rho)},
\end{aligned} \tag{41}$$

where (i) follows from eq. (37) and (ii) follows from Lemma 4. We next bound the third term on the right hand side of eq. (41) as follows:

$$\mathbb{E}\left[\left\|g(\theta_{t-1},w_{t-1},\mathcal{B}_{t-1}) - \frac{1}{2}\nabla J(\theta_{t-1})\right\|_2^2\right]$$

$$\leq 2\mathbb{E}\left[\|g(\theta_{t-1}, w_{t-1}, \mathcal{B}_{t-1}) - g(\theta_{t-1}, w(\theta_{t-1}), \mathcal{B}_{t-1})\|_2^2\right] + 2\mathbb{E}\left[\left\|g(\theta_{t-1}, w(\theta_{t-1}), \mathcal{B}_{t-1}) - \frac{1}{2}\nabla J(\theta_{t-1})\right\|_2^2\right]$$
$$\overset{(i)}{\leq} 2L_e^2\mathbb{E}\left[\|w_{t-1} - w(\theta_{t-1})\|_2^2\right] + \frac{16C_g^2[1+(\kappa-1)\rho]}{M(1-\rho)}. \tag{42}$$

Substituting eq. (42) into eq. (41) yields

$$\begin{aligned}
&\mathbb{E}[\|w_t - w(\theta_t)\|_2^2] \\
&\leq \left(1 - \frac{\lambda_v}{4}\beta + \frac{16L_w^2L_e^2\alpha^2}{\lambda_v\beta}\right)\mathbb{E}[\|w_{t-1} - w(\theta_{t-1})\|_2^2] + \frac{2L_w^2\alpha^2}{\lambda_v\beta}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2] \\
&\quad + \left[\frac{128L_w^2C_g^2\alpha^2}{\lambda_v\beta} + 4C_f^2\left(\frac{\beta}{\lambda_v} + 2\beta^2\right)\right]\frac{[1+(\kappa-1)\rho]}{M(1-\rho)} \\
&\overset{(i)}{\leq} \left(1 - \frac{\lambda_v}{8}\beta\right)\mathbb{E}[\|w_{t-1} - w(\theta_{t-1})\|_2^2] + \frac{2L_w^2\alpha^2}{\lambda_v\beta}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2] + \frac{D_1[1+(\kappa-1)\rho]}{M(1-\rho)},
\end{aligned} \tag{43}$$

where (i) follows from the fact that $\alpha \leq \frac{\lambda_v}{8\sqrt{2}L_wL_e}\beta$ and we define $D_1 = \frac{128L_w^2C_g^2\alpha^2}{\lambda_v\beta} + 4C_f^2(\frac{\beta}{\lambda_v} + 2\beta^2)$. □

B.2 Proof of Theorem 2

Since $J(\theta)$ is L_J-gradient Lipschitz, we have

$$\begin{aligned}
&\mathbb{E}[J(\theta_{t+1})] \\
&\leq \mathbb{E}[J(\theta_t)] + \mathbb{E}[\langle\nabla J(\theta_t), \theta_{t+1} - \theta_t\rangle] + \frac{L_J}{2}\mathbb{E}[\|\theta_{t+1} - \theta_t\|_2^2] \\
&= \mathbb{E}[J(\theta_t)] - \frac{\alpha}{2}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] - \alpha\mathbb{E}[\langle\nabla J(\theta_t), g(\theta_t, w_t, \mathcal{B}_t) - \frac{1}{2}\nabla J(\theta_t)\rangle] + \frac{L_J\alpha^2}{2}\mathbb{E}[\|g(\theta_t, w_t, \mathcal{B}_t)\|_2^2] \\
&\leq \mathbb{E}[J(\theta_t)] - \left(\frac{\alpha}{4} - \frac{L_J\alpha^2}{8}\right)\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] + (\alpha + L_J\alpha^2)\mathbb{E}\left[\left\|g(\theta_t, w_t, \mathcal{B}_t) - \frac{1}{2}\nabla J(\theta_t)\right\|_2^2\right] \\
&\leq \mathbb{E}[J(\theta_t)] - \left(\frac{\alpha}{4} - \frac{L_J\alpha^2}{8}\right)\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] + 2(\alpha + L_J\alpha^2)\mathbb{E}[\|g(\theta_t, w_t, \mathcal{B}_t) - g(\theta_t, w(\theta_t), \mathcal{B}_t)\|_2^2] \\
&\quad + 2(\alpha + L_J\alpha^2)\mathbb{E}\left[\left\|g(\theta_t, w(\theta_t), \mathcal{B}_t) - \frac{1}{2}\nabla J(\theta_t)\right\|_2^2\right] \\
&\overset{(i)}{\leq} \mathbb{E}[J(\theta_t)] - \left(\frac{\alpha}{4} - \frac{L_J\alpha^2}{8}\right)\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] + 2(\alpha + L_J\alpha^2)L_e^2\mathbb{E}[\|w_t - w(\theta_t)\|_2^2] \\
&\quad + 2(\alpha + L_J\alpha^2)\frac{4C_g^2[1+(\kappa-1)\rho]}{M(1-\rho)},
\end{aligned} \tag{44}$$

where (i) follows from Assumption 5 and Lemma 2. Rearranging the above inequality and summing from $t = 0$ to $T - 1$ yield

$$\begin{aligned}
\left(\frac{\alpha}{4} - \frac{L_J\alpha^2}{8}\right)\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] &\leq J(\theta_0) - J(\theta_T) + 2(\alpha + L_J\alpha^2)T\frac{4C_g^2[1+(\kappa-1)\rho]}{M(1-\rho)} \\
&\quad + 2(\alpha + L_J\alpha^2)L_e^2\sum_{t=0}^{T-1}\mathbb{E}\|w_t - w(\theta_t)\|_2^2.
\end{aligned} \tag{45}$$

Now we upper bound the term $\sum_{t=0}^{T-1}\mathbb{E}\|w_t - w(\theta_t)\|_2^2$. Applying the inequality in Lemma 5 recursively yields

$$\mathbb{E}[\|w_t - w(\theta_t)\|_2^2] \leq \left(1 - \frac{\lambda_v}{8}\beta\right)^t\|w_0 - w(\theta_0)\|_2^2 + \frac{2L_w^2\alpha^2}{\lambda_v\beta}\sum_{i=0}^{t-1}\left(1 - \frac{\lambda_v}{8}\beta\right)^{t-1-i}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2]$$

$$+\frac{D_1[1+(\kappa-1)\rho]}{M(1-\rho)}\sum_{i=0}^{t-1}\Big(1-\frac{\lambda_v}{8}\beta\Big)^{t-1-i},$$

which implies

$$\begin{aligned}\sum_{t=0}^{T-1}\mathbb{E}\left\|w_t-w(\theta_t)\right\|_2^2 &\le \left\|w_0-w(\theta_0)\right\|_2^2\sum_{t=0}^{T-1}\Big(1-\frac{\lambda_v}{8}\beta\Big)^t+\frac{2L_w^2\alpha^2}{\lambda_v\beta}\sum_{t=0}^{T-1}\sum_{i=0}^{t-1}\Big(1-\frac{\lambda_v}{8}\beta\Big)^{t-1-i}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2]\\&\quad+\frac{D_1[1+(\kappa-1)\rho]}{M(1-\rho)}\sum_{t=0}^{T-1}\sum_{i=0}^{t-1}\Big(1-\frac{\lambda_v}{8}\beta\Big)^{t-1-i}\\&\le\frac{8\left\|w_0-w(\theta_0)\right\|_2^2}{\lambda_v\beta}+\frac{16L_w^2\alpha^2}{\lambda_v^2\beta^2}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_{t-1})\|_2^2]+\frac{8D_1T}{\lambda_v\beta}\frac{1+(\kappa-1)\rho}{M(1-\rho)}.\end{aligned}\tag{46}$$

Substituting eq. (46) into eq. (45) yields

$$\begin{aligned}&\Big(\frac{\alpha}{4}-\frac{L_J\alpha^2}{8}-\frac{32L_w^2L_e^2\alpha^3(1+L_J\alpha)}{\lambda_v^2\beta^2}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]\\&\le J(\theta_0)-\mathbb{E}[J(\theta_T)]+\frac{16(\alpha+L_J\alpha^2)L_e^2}{\lambda_v\beta}\left\|w_0-w(\theta_0)\right\|_2^2+2(\alpha+L_J\alpha^2)T\frac{4C_g^2[1+(\kappa-1)\rho]}{M(1-\rho)}\\&\quad+\frac{16D_1(\alpha+L_J\alpha^2)L_e^2T}{\lambda_v\beta}\frac{1+(\kappa-1)\rho}{M(1-\rho)}.\end{aligned}\tag{47}$$

Dividing both sides of eq. (47) by T and using the fact that $\frac{\alpha}{4}-\frac{L_J\alpha^2}{8}-\frac{32L_w^2L_e^2\alpha^3(1+L_J\alpha)}{\lambda_v^2\beta^2}\ge\frac{\alpha}{8}$, we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]\\&\le\frac{8(J(\theta_0)-\mathbb{E}[J(\theta_T)])}{\alpha T}+\frac{64(1+L_J\alpha)L_e^2}{\lambda_v\beta}\frac{\left\|w_0-w(\theta_0)\right\|_2^2}{T}+64(1+L_J\alpha)\Big(C_g^2+\frac{2D_1L_e^2}{\lambda_v\beta}\Big)\frac{1+(\kappa-1)\rho}{M(1-\rho)}.\end{aligned}\tag{48}$$

C Convergence Analysis of Two Time-scale Greedy-GQ

We make the following definitions. For a given θ, we define matrices $A_\theta=\mathbb{E}_{\mu_{\pi_b}}[(\gamma\mathbb{E}_{\pi_\theta}[\phi(s')|s]-\phi(s))\phi(s)^\top]$, $B_\theta=\mathbb{E}_{\mu_{\pi_b}}[\mathbb{E}_{\pi_\theta}[\phi(s')|s]\phi(s)^\top]$, $C=-\mathbb{E}_{\mu_{\pi_b}}[\phi(s)\phi(s)^\top]$ and vectors $b_\theta=\mathbb{E}_{\mu_{\pi_b}}[\mathbb{E}_{\pi_\theta}[r(s',s)|s]\phi(s)]$, $w^*(\theta)=C^{-1}(A_\theta\theta+b_\theta)$, $\theta^*=-A_\theta^{-1}b_\theta$. We also define the stochastic matrices $A_t=\frac{1}{|\mathcal{B}_t|}\sum_{j\in\mathcal{B}_t}\gamma\rho_{\theta_t}(s_j,a_j)\phi(s_{j+1})\phi(s_j)^\top-\phi(s_j)\phi(s_j)^\top$, $B_t=\frac{1}{|\mathcal{B}_t|}\sum_{j\in\mathcal{B}_t}\rho_{\theta_t}(s_j,a_j)\phi(s_{j+1})\phi(s_j)^\top$, $C_t=\frac{1}{|\mathcal{B}_t|}\sum_{j\in\mathcal{B}_t}\phi(s_j)\phi(s_j)^\top$ and stochastic vector $b_t=\frac{1}{|\mathcal{B}_t|}\sum_{j\in\mathcal{B}_t}\rho_{\theta_t}(s_j,a_j)r(s_{j+1},s_j)\phi(s_j)$.

We also define the full (semi)-gradient as follows:

$$-\frac{1}{2}\nabla J(\theta)=g(\theta)=(A_\theta-B_\theta C^{-1}A_\theta)\theta+(b_\theta-B_\theta C^{-1}b_\theta),\tag{49}$$

$$f(w)=C(w-w^*(\theta)),\tag{50}$$

and stochastic (semi)-gradient at step t as follows:

$$g_t(\theta_t)=(A_t-B_tC^{-1}A_{\theta_t})\theta_t+(b_t-B_tC^{-1}b_{\theta_t}),\tag{51}$$

$$f_t(w_t)=C_t(w_t-w^*(\theta_t)),\tag{52}$$

$$h_t(\theta_t)=(A_t-C_tC^{-1}A_{\theta_t})\theta_t+(b_t-C_tC^{-1}b_{\theta_t}).\tag{53}$$

We first consider the induction relationship for the fast time-scale variable w_t. Following similar steps from eq. (22) to eq. (23), letting $M\ge128\left(\rho_{\max}^2+\frac{1}{\lambda_2^2}\right)\frac{1+(\kappa-1)\rho}{1-\rho}\max\{1,\frac{\lambda_2^2\beta}{4\alpha^2}(\frac{2\beta}{\lambda_2}+2\beta^2)\}$ and $\beta\le\frac{\lambda_2}{4}$, we obtain

$$\mathbb{E}[\|w_{t+1}-w^*(\theta_{t+1})\|_2^2]$$

$$
\begin{aligned}
&\le \left(1-\frac{\lambda_2\beta}{4}+\frac{16\rho_{\max}^2\alpha^2}{\lambda_2^2\beta}\right)\mathbb{E}[\|w_t-w^*(\theta_t)\|_2^2]+\frac{100\alpha^2}{\lambda_2^2\beta}\mathbb{E}[\|\theta_t-\theta^*\|_2^2]\\
&\quad+32(4R_\theta^2\rho_{\max}^2+r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta}+\frac{2\beta}{\lambda_2}+2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\\
&\overset{(i)}{\le}\left(1-\frac{\lambda_2\beta}{8}\right)\mathbb{E}[\|w_t-w^*(\theta_t)\|_2^2]+\frac{100\lambda_1^2\alpha^2}{\lambda_2^2\beta}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]\\
&\quad+32(4R_\theta^2\rho_{\max}^2+r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta}+\frac{2\beta}{\lambda_2}+2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M},
\end{aligned}
\tag{54}
$$

where (i) follows from the fact that $\alpha\le\frac{\lambda_2\sqrt{\lambda_2}}{8\sqrt{2}\rho_{\max}}\beta$ and $\|\theta_t-\theta^*\|_2\le\lambda_1\|\nabla J(\theta_t)\|_2$ according to the definition of $\nabla J(\theta)$ in 49. We next consider the induction relationship for the slow time-scale variable θ_t. Since $J(\theta)$ is L_J-gradient Lipschitz, we have

$$
\begin{aligned}
&\mathbb{E}[J(\theta_{t+1})]\\
&\le\mathbb{E}[J(\theta_t)]+\mathbb{E}[\langle\nabla J(\theta_t),\theta_{t+1}-\theta_t\rangle]+\frac{L_J}{2}\mathbb{E}[\|\theta_{t+1}-\theta_t\|_2^2]\\
&=\mathbb{E}[J(\theta_t)]-\frac{\alpha}{2}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]-\alpha\mathbb{E}[\langle\nabla J(\theta_t),-g_t(\theta_t)-\frac{1}{2}\nabla J(\theta_t)\rangle]\\
&\quad+\alpha\mathbb{E}[\langle\nabla J(\theta_t),B_t(w_t-w^*(\theta_t))\rangle]+\frac{L_J\alpha^2}{2}\mathbb{E}[\|g_t(\theta_t)+B_t(w_t-w^*(\theta_t))\|_2^2]\\
&\overset{(i)}{\le}\mathbb{E}[J(\theta_t)]-\frac{\alpha}{4}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]+2\alpha\mathbb{E}\left[\left\|-g_t(\theta_t)-\frac{1}{2}\nabla J(\theta_t)\right\|_2^2\right]\\
&\quad+2\alpha\mathbb{E}[\|B_t(w_t-w^*(\theta_t))\|_2^2]+L_J\alpha^2\mathbb{E}[\|g_t(\theta_t)\|_2^2]+L_J\alpha^2\mathbb{E}[\|B_t(w_t-w^*(\theta_t))\|_2^2]\\
&\overset{(ii)}{\le}\mathbb{E}[J(\theta_t)]-\left(\frac{\alpha}{4}-\frac{L_J\alpha^2}{2}\right)\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]+2(\alpha+L_J\alpha^2)\mathbb{E}\left[\left\|-g_t(\theta_t)-\frac{1}{2}\nabla J(\theta_t)\right\|_2^2\right]\\
&\quad+(2\alpha+L_J\alpha^2)\rho_{\max}^2\mathbb{E}[\|w_t-w^*(\theta_t)\|_2^2],
\end{aligned}
\tag{55}
$$

where (i) follows from Young's inequality and (ii) follows from the fact that $\|g_t(\theta_t)\|_2^2\le\frac{1}{2}\|\nabla J(\theta_t)\|_2^2+2\left\|-g_t(\theta_t)-\frac{1}{2}\nabla J(\theta_t)\right\|_2^2$ and $\|B_t\|_2\le\rho_{\max}$. Then, we upper bound the term $\mathbb{E}[\left\|-g_t(\theta_t)-\frac{1}{2}\nabla J(\theta_t)\right\|_2^2]$ as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\left\|-g_t(\theta_t)-\frac{1}{2}\nabla J(\theta_t)\right\|_2^2\right]\\
&=\mathbb{E}\left[\left\|\left[(A_t-A_{\theta_t})-(B_t-B_{\theta_t})C_{\theta_t}^{-1}A_{\theta_t}\right]\theta_t+\left[(b_t-b_{\theta_t})-(B_t-B_{\theta_t})C_{\theta_t}^{-1}b_{\theta_t}\right]\right\|_2^2\right]\\
&\le4\mathbb{E}\left[\|(A_t-A_{\theta_t})\theta_t\|_2^2\right]+4\mathbb{E}\left[\left\|(B_t-B_{\theta_t})C_{\theta_t}^{-1}A_{\theta_t}\theta_t\right\|_2^2\right]+4\mathbb{E}\left[\|b_t-b_{\theta_t}\|_2^2\right]\\
&\quad+4\mathbb{E}\left[\left\|(B_t-B_{\theta_t})C_{\theta_t}^{-1}b_{\theta_t}\right\|_2^2\right]\\
&\le4\mathbb{E}\left[\|A_t-A_{\theta_t}\|_2^2\|\theta_t\|_2^2\right]+4\mathbb{E}\left[\|B_t-B_{\theta_t}\|_2^2\left\|C_{\theta_t}^{-1}\right\|_2^2\|A_{\theta_t}\|_2^2\|\theta_t\|_2^2\right]+4\mathbb{E}\left[\|b_t-b_{\theta_t}\|_2^2\right]\\
&\quad+4\mathbb{E}\left[\|B_t-B_{\theta_t}\|_2^2\left\|C_{\theta_t}^{-1}\right\|_2^2\|b_{\theta_t}\|_2^2\right]\\
&=4\mathbb{E}\left[\mathbb{E}[\|A_t-A_{\theta_t}\|_2^2|\mathcal{F}_t]\|\theta_t\|_2^2\right]+4\mathbb{E}\left[\mathbb{E}[\|B_t-B_{\theta_t}\|_2^2|\mathcal{F}_t]\left\|C_{\theta_t}^{-1}\right\|_2^2\|A_{\theta_t}\|_2^2\|\theta_t\|_2^2\right]+4\mathbb{E}\left[\|b_t-b_{\theta_t}\|_2^2\right]\\
&\quad+4\mathbb{E}\left[\mathbb{E}[\|B_t-B_{\theta_t}\|_2^2|\mathcal{F}_t]\left\|C_{\theta_t}^{-1}\right\|_2^2\|b_{\theta_t}\|_2^2\right]\\
&\le\frac{32(\rho_{\max}+1)^2[1+(\kappa-1)\rho]}{(1-\rho)M}\mathbb{E}\left[\|\theta_t\|_2^2\right]+\frac{32(\rho_{\max}+1)^2\rho_{\max}^2[1+(\kappa-1)\rho]}{(1-\rho)\lambda_2^2M}\mathbb{E}\left[\|\theta_t\|_2^2\right]\\
&\quad+\frac{32r_{\max}^2\rho_{\max}^2[1+(\kappa-1)\rho]}{(1-\rho)M}+\frac{32\rho_{\max}^2[1+(\kappa-1)\rho]}{(1-\rho)\lambda_2^2M}\\
&\le\frac{32(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{(1-\rho)M}\mathbb{E}\left[\|\theta_t\|_2^2\right]+\frac{32(r_{\max}^2+1)\rho_{\max}^2[1+(\kappa-1)\rho]}{(1-\rho)M}
\end{aligned}
$$

$$\le \frac{64(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{(1-\rho)M}\mathbb{E}\left[\|\theta_t^*\|_2^2\right] + \frac{64(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{(1-\rho)M}\mathbb{E}\left[\|\theta_t-\theta_t^*\|_2^2\right]$$
$$+ \frac{32(r_{\max}^2+1)\rho_{\max}^2[1+(\kappa-1)\rho]}{(1-\rho)M}$$
$$\le \frac{C_1[1+(\kappa-1)\rho]}{(1-\rho)M} + \frac{64\lambda_1^2(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{(1-\rho)M}\mathbb{E}\left[\|\nabla J(\theta_t)\|_2^2\right], \tag{56}$$

where $C_2 = 32[2(\rho_{\max}+1)^4R_\theta^2 + (r_{\max}^2+1)\rho_{\max}^2]$. Substituting eq. (56) into eq. (55), rearranging the terms and summing from $t=0$ to $T-1$ yield

$$\Big(\frac{\alpha}{4} - \frac{L_J\alpha^2}{2}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]$$
$$\le J(\theta_0) - \mathbb{E}[J(\theta_T)] + 2(\alpha+L_J\alpha^2)T\frac{C_2[1+(\kappa-1)\rho]}{M(1-\rho)} + (2\alpha+L_J\alpha^2)\rho_{\max}^2\sum_{t=0}^{T-1}\mathbb{E}\|w_t-w(\theta_t)\|_2^2$$
$$+ 2(\alpha+L_J\alpha^2)\frac{64\lambda_1^2(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{(1-\rho)M}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]. \tag{57}$$

Then, we bound the term $\sum_{t=0}^{T-1}\mathbb{E}\|w_t-w(\theta_t)\|_2^2$. Applying eq. (54) iteratively yields:

$$\mathbb{E}[\|w_t-w^*(\theta_t)\|_2^2]$$
$$\le \left(1-\frac{\lambda_2\beta}{8}\right)^t\|w_0-w^*(\theta_0)\|_2^2 + \frac{100\lambda_1^2\alpha^2}{\lambda_2^2\beta}\sum_{i=0}^{t-1}\left(1-\frac{\lambda_2\beta}{8}\right)^i\mathbb{E}[\|\nabla J(\theta_i)\|_2^2]$$
$$+ 32(4R_\theta^2\rho_{\max}^2+r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta}+\frac{2\beta}{\lambda_2}+2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}\sum_{i=0}^{t-1}\left(1-\frac{\lambda_2\beta}{8}\right)^i$$
$$\le \left(1-\frac{\lambda_2\beta}{8}\right)^t\|w_0-w^*(\theta_0)\|_2^2 + \frac{100\lambda_1^2\alpha^2}{\lambda_2^2\beta}\sum_{i=0}^{t-1}\left(1-\frac{\lambda_2\beta}{8}\right)^i\mathbb{E}[\|\nabla J(\theta_i)\|_2^2]$$
$$+ \frac{256}{\lambda_2\beta}(4R_\theta^2\rho_{\max}^2+r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta}+\frac{2\beta}{\lambda_2}+2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}. \tag{58}$$

Summing eq. (58) from $t=0$ to $T-1$ yields

$$\sum_{t=0}^{T-1}\mathbb{E}[\|w_t-w^*(\theta_t)\|_2^2]$$
$$\le \|w_0-w^*(\theta_0)\|_2^2\sum_{t=0}^{T-1}\left(1-\frac{\lambda_2\beta}{8}\right)^t + \frac{100\lambda_1^2\alpha^2}{\lambda_2^2\beta}\sum_{t=0}^{T-1}\sum_{i=0}^{t-1}\left(1-\frac{\lambda_2\beta}{8}\right)^{t-1-i}\mathbb{E}[\|\nabla J(\theta_i)\|_2^2]$$
$$+ \frac{256T}{\lambda_2\beta}(4R_\theta^2\rho_{\max}^2+r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta}+\frac{2\beta}{\lambda_2}+2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}$$
$$\le \frac{8}{\lambda_2\beta}\|w_0-w^*(\theta_0)\|_2^2 + \frac{800\lambda_1^2\alpha^2}{\lambda_2^3\beta^2}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]$$
$$+ \frac{256T}{\lambda_2\beta}(4R_\theta^2\rho_{\max}^2+r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta}+\frac{2\beta}{\lambda_2}+2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M}. \tag{59}$$

Substituting eq. (59) into eq. (57) yields

$$\Big(\frac{\alpha}{4} - \frac{L_J\alpha^2}{2}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2]$$
$$\le J(\theta_0) - \mathbb{E}[J(\theta_T)] + (2\alpha+L_J\alpha^2)\frac{8\rho_{\max}^2}{\lambda_2\beta}\|w_0-w^*(\theta_0)\|_2^2 + 2(\alpha+L_J\alpha^2)T\frac{C_2[1+(\kappa-1)\rho]}{M(1-\rho)}$$

$$
\begin{aligned}
&+ (2\alpha + L_J\alpha^2)\rho_{\max}^2 \frac{800\lambda_1^2\alpha^2}{\lambda_2^3\beta^2} \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla J(\theta_t)\|_2^2] \\
&+ 2(\alpha + L_J\alpha^2) \frac{64\lambda_1^2(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{(1-\rho)M} \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla J(\theta_t)\|_2^2] \\
&+ (2\alpha + L_J\alpha^2)\rho_{\max}^2 \frac{256T}{\lambda_2\beta}(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2\right)\frac{1+(\kappa-1)\rho}{(1-\rho)M} \\
&\overset{(i)}{\leq} J(\theta_0) - \mathbb{E}[J(\theta_T)] + \frac{24\alpha\rho_{\max}^2}{\lambda_2\beta}\|w_0 - w^*(\theta_0)\|_2^2 + 4\alpha T\frac{C_1[1+(\kappa-1)\rho]}{M(1-\rho)} \\
&+ \frac{2656\rho_{\max}^2\lambda_1^2\alpha^3}{\lambda_2^3\beta^2}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2], \qquad (60)
\end{aligned}
$$

where in (i) we let $\alpha \leq \frac{1}{L_J}$ and $M \geq \frac{\beta^2\lambda_2^3(\rho_{\max}+1)^4[1+(\kappa-1)\rho]}{\rho_{\max}^2\alpha^2(1-\rho)}$, and define $C_1 = C_2 + \frac{192\rho_{\max}^2}{\lambda_2\beta}(4R_\theta^2\rho_{\max}^2 + r_{\max}^2)\left(\frac{32\alpha^2}{\lambda_2^2\beta} + \frac{2\beta}{\lambda_2} + 2\beta^2\right)$. Rearranging eq. (60) yields

$$
\begin{aligned}
&\left(\frac{\alpha}{4} - \frac{L_J\alpha^2}{2} - \frac{2656\rho_{\max}^2\lambda_1^2\alpha^3}{\lambda_2^3\beta^2}\right)\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] \\
&\leq J(\theta_0) - \mathbb{E}[J(\theta_T)] + \frac{24\alpha\rho_{\max}^2}{\lambda_2\beta}\|w_0 - w^*(\theta_0)\|_2^2 + 4\alpha T\frac{C_1[1+(\kappa-1)\rho]}{M(1-\rho)}.
\end{aligned}
$$

Letting $\alpha \leq \min\{\frac{1}{8L_J}, \frac{L_J\lambda_2^3\beta^2}{5312\rho_{\max}^2\lambda_1^2}\}$, we obtain

$$
\begin{aligned}
&\frac{\alpha}{8}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] \\
&\leq J(\theta_0) - \mathbb{E}[J(\theta_T)] + \frac{24\alpha\rho_{\max}^2}{\lambda_2\beta}\|w_0 - w^*(\theta_0)\|_2^2 + 4\alpha T\frac{C_1[1+(\kappa-1)\rho]}{M(1-\rho)}.
\end{aligned}
$$

Dividing both sides of the above inequality by $\frac{\alpha T}{8}$ yields

$$
\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla J(\theta_t)\|_2^2] \leq \frac{8(J(\theta_0) - \mathbb{E}[J(\theta_T)])}{\alpha T} + \frac{192\rho_{\max}^2}{\lambda_2\beta}\frac{\|w_0 - w^*(\theta_0)\|_2^2}{T} + \frac{32C_1[1+(\kappa-1)\rho]}{M(1-\rho)}.
$$