Supplementary materials for
“Finite-Sample Analysis of Decentralized Temporal-Difference Learning with Linear Function Approximation”

A Proof of Theorem 1

Proof. From the definition of $G(\Theta)$ in (24), we have that

$$G(\Theta(k), \xi_k) = \begin{bmatrix} \theta_1^T(k)[\gamma \phi(s(k+1)) - \phi(s(k))]\phi^T(s(k)) \\ \theta_2^T(k)[\gamma \phi(s(k+1)) - \phi(s(k))]\phi^T(s(k)) \\ \vdots \\ \theta_M^T(k)[\gamma \phi(s(k+1)) - \phi(s(k))]\phi^T(s(k)) \end{bmatrix} + \begin{bmatrix} r_1(k)\phi^T(s(k)) \\ r_2(k)\phi^T(s(k)) \\ \vdots \\ r_M(k)\phi^T(s(k)) \end{bmatrix} = \Theta(k)[\gamma \phi(s(k+1)) - \phi(s(k))]\phi^T(s(k)) + r(k)\phi^T(s(k)) = \Theta(k)H^T(\xi_k) + r(k)\phi^T(s(k))$$

where we have used the definitions that $r(k) = [r_1(k) r_2(k) \cdots r_M(k)]^T$ and $H(\xi_k) := \phi(s(k))[\gamma \phi(s(k+1)) - \phi^T(s(k))]$. Using standard norm inequalities, it follows that

$$\|\Delta G(\Theta(k), \xi_k)\|_F \leq \|[\gamma \phi(s(k+1)) - \phi(s(k))]\phi^T(s(k))\|_F \cdot \|\Delta \Theta(k)\|_F + \|r(k)\phi^T(s(k))\|_F$$

$$\leq 1 + \gamma \|\Delta \Theta(k)\|_F + \|\phi(s(k))\|_F \cdot \|r(k)\|_F \cdot \|\phi(s(k))\|_F$$

$$\leq (1 + \gamma)\|\Delta \Theta(k)\|_F + \sqrt{Mr_{\max}}$$

(44)

where $1 + \gamma \leq 2$ for the discounting factor $0 \leq \gamma < 1$, and the last inequality holds since feature vectors $\|\phi(s)\| \leq 1$, rewards $r(k) \leq r_{\max}$, and the Frobenious norm of rank-one matrices is equivalent to the $\ell_2$-norm of vectors. For future reference, notice from the above inequality that $\lambda_{\max}(H(\xi_k)) \leq \|H(\xi_k)\|_F = \|[\gamma \phi(s(k+1)) - \phi(s(k))]\phi^T(s(k))\| \leq 1 + \gamma \leq 2$, for all $k \in \mathbb{N}^+$.

It follows from (28) that

$$\|\Delta \Theta(k+1)\|_F \leq \|W \Delta \Theta(k)\|_F + \alpha \|\Delta G(\Theta(k))\|_F$$

$$\leq [\lambda_2^W + 2\alpha] \|\Delta \Theta(k)\|_F + \alpha \sqrt{Mr_{\max}}$$

(45)

where the second inequality is obtained after using (44), and the following inequality [Nedić et al., 2018, Ma et al., 2019]

$$\|W \Delta \Theta(k)\|_F = \left\|W \left(I - \frac{1}{M} 11^T\right) \Theta(k)\right\| \leq \lambda_2^W \|\Delta \Theta(k)\|_F.$$  

(46)

Then applying (45) recursively from iteration $k$ to 0 gives rise to

$$\|\Delta \Theta(k)\|_F \leq (\lambda_2^W + 2\alpha)^k \|\Delta \Theta(0)\|_F + \alpha \sqrt{Mr_{\max}} \sum_{i=0}^{k-1} (\lambda_2^W + 2\alpha)^i$$

$$\leq (\lambda_2^W + 2\alpha)^k \|\Delta \Theta(0)\|_F + \frac{\alpha \sqrt{Mr_{\max}}}{1 - \lambda_2^W - 2\alpha}$$

$$\leq (\lambda_2^W + 2\alpha)^k \|\Delta \Theta(0)\|_F + \alpha \cdot \frac{2\sqrt{Mr_{\max}}}{1 - \lambda_2^W}$$

(47)

where the last inequality is a consequence of using the fact that $0 < \alpha < \frac{1}{2} \cdot \frac{1 - \lambda_2^W}{2}$. This concludes the proof of Theorem 1. \qed
B Proof of Lemma 1

Proof. Recalling the definitions of $H(\xi_k)$ ($\bar{H}$) and $b(\xi_k)$ ($\bar{b}$), it is not difficult to verify that in the stationary distribution $\pi$ of the Markov chain, the expectations of $H(\xi_k)$ and $b(\xi_k)$ obey

$$E_\pi[H(\xi_k)] = \bar{H}$$

and

$$E_\pi[b_\bar{\xi}(\xi_k)] = \bar{b}.$$  
Thus,

$$E_\pi \left[ \frac{1}{M} G^T(\Theta(k), \xi_k)1 | F(k) \right] = E_\pi \left[ H(\xi_k)\bar{\theta}(k) + b_\bar{\xi}(\xi_k) | F(k) \right] = \bar{H}\bar{\theta}(k) + \bar{b}$$

and its variance satisfies

$$E_\pi \left[ \left\| \frac{1}{M} G^T(\Theta(k), \xi_k)1 - \bar{g}(\bar{\theta}(k)) \right\|^2 | F(k) \right] = E_\pi \left[ \left\| (H(\xi_k) - \bar{H})\bar{\theta}(k) + b_\bar{\xi}(\xi_k) - \bar{b} \right\|^2 | F(k) \right]$$

$$\leq E_\pi [2\|H(\xi_k) - \bar{H}\|\bar{\theta}(k)||^2 + 2\|b_\bar{\xi}(\xi_k) - \bar{b}\|^2 | F(k) \right]$$

$$\leq 2\beta^2\|\bar{\theta}(k) - \theta^*\|^2 + 2\|b_\bar{\xi}(\xi_k) - \bar{b}\|^2$$

$$\leq 2\beta^2\|\bar{\theta}(k) - \theta^*\|^2 + 4\beta^2\|\theta^*\|^2 + 8r^2_{\max}$$

where $\beta$ denotes the largest absolute value of eigenvalues of $H(\xi_k) - \bar{H}$, for any $k \in \mathbb{N}^+$. \hfill \Box

C Proof of Theorem 2

Proof. Clearly, it holds that

$$E_\pi[\left\| \bar{\theta}(k+1) - \theta^* \right\|^2 | F(k) ] = E_\pi \left[ \left\| \bar{\theta}(k) - \theta^* + \frac{1}{M} G^T(\Theta, \xi_k)1 \right\|^2 | F(k) \right]$$

$$\leq \left\| \bar{\theta}(k) - \theta^* \right\|^2 + 2\alpha \left\langle \bar{\theta}(k) - \theta^*, E_\pi \left[ \frac{1}{M} G(\Theta(k), \xi_k)^T1 | F(k) \right] \right\rangle$$

$$+ \alpha^2 E_\pi \left[ \left\| \frac{1}{M} G(\Theta(k), \xi_k)^T1 - \bar{g}(\bar{\theta}(k)) + \bar{g}(\bar{\theta}(k)) \right\|^2 | F(k) \right]$$

$$\leq \left\| \bar{\theta}(k) - \theta^* \right\|^2 + 2\alpha \left\langle \bar{\theta}(k) - \theta^*, \bar{H}(\bar{\theta}(k) - \theta^*) \right\rangle$$

$$+ 2\alpha^2(2\beta^2\|\bar{\theta}(k) - \theta^*\|^2 + 4\alpha^2\|\bar{\theta}(k) - \theta^*\|^2 + 8r^2_{\max}) + 2\alpha^2\|\bar{H}(\bar{\theta}(k) - \theta^*)\|^2$$

$$\leq \left[ 1 + 2\alpha^2\lambda_{\max}^H + 8\alpha^2\beta^2 + 2\alpha^2(\lambda_{\min}^H)^2 \right] \left\| \bar{\theta}(k) - \theta^* \right\|^2$$

$$+ (8\alpha^2\beta^2\|\theta^*\|^2 + 16\alpha^2r^2_{\max}).$$

(52)

where $\lambda_{\max}^H$ and $\lambda_{\min}^H$ are the largest and the smallest eigenvalues of $\bar{H}$, respectively. Because $\bar{H}$ is a negative definite matrix, then it follows that $\lambda_{\min}^H < \lambda_{\max}^H < 0$.

Defining constants $c_1 := 1 + 2\alpha\lambda_{\max}^H + 8\alpha^2\beta^2 + 2\alpha^2(\lambda_{\min}^H)^2$, and choosing any constant stepsize $\alpha$ obeying $0 < \alpha \leq \frac{1}{2} \cdot \frac{\lambda_{\max}^H}{4\alpha^2(\lambda_{\min}^H)^2}$, then we have $c_1 < 1$ and $\frac{1}{1-c_1} \leq -\frac{\lambda_{\max}^H}{\alpha\lambda_{\min}^H}$. Now, taking expectation with respect to $F(k)$ in (52) gives rise to

$$E[\left\| \bar{\theta}(k+1) - \theta^* \right\|^2] \leq c_1 E[\left\| \bar{\theta}(k) - \theta^* \right\|^2] + (8\alpha^2\beta^2\|\theta^*\|^2 + 16\alpha^2r^2_{\max}).$$

(53)

Applying the above recursion from iteration $k$ to iteration 0 yields

$$E[\left\| \bar{\theta}(k) - \theta^* \right\|^2] \leq c_1^n \left\| \bar{\theta}(0) - \theta^* \right\|^2 + \frac{1-c_1}{1-c_1} (8\alpha^2\beta^2\|\theta^*\|^2 + 16\alpha^2r^2_{\max})$$

(54)
For notational brevity, let
\[ c_2 := \frac{8\beta^2\|\theta^*\|^2 + 16\alpha r_{\max}^2}{-\lambda_{\text{max}}} \]
where this concludes the proof.

\[ \Box \]

**D Proof of Proposition 1**

**Proof.** We have that
\[
E[\|\theta_m(k) - \theta^*\|^2] = E[\|\theta_m(k) - \bar{\theta}(k) + \bar{\theta}(k) - \theta^*\|^2] \\
\leq 2E[\|\theta_m(k) - \bar{\theta}(k)\|^2] + 2E[\|\bar{\theta}(k) - \theta^*\|^2] \\
\leq 2E[\|\Delta \Theta(k)\|^2_F] + 2E[\|\bar{\theta}(k) - \theta^*\|^2] \\
\leq 2\left(\lambda_2^W + 2\alpha\right)^k \|\Delta \Theta(0)\|_F \leq 2\left[\frac{\alpha \sqrt{M r_{\max}}}{1 - \lambda_2^W} + \frac{8\alpha^2 M r_{\max}^2}{(1 - \lambda_2^W)^2} + 2c_1^k \|\bar{\theta}(0) - \theta^*\|^2 + 2\alpha c_2 \right] \\
\leq 4\left(\lambda_2^W + 2\alpha\right)^{2k} \|\Delta \Theta(0)\|_F \leq 2\left[\frac{\alpha \sqrt{M r_{\max}}}{1 - \lambda_2^W} + \frac{8\alpha^2 M r_{\max}^2}{(1 - \lambda_2^W)^2} + 2c_1^k \|\bar{\theta}(0) - \theta^*\|^2 + 2\alpha c_2 \right].
\]

where the third inequality follows from using (29) and (54). Letting \( c_3 := \max\{\lambda_2^W + 2\alpha\}^2, V_0 := 2\max\{4\|\Delta \Theta(0)\|_F^2, 2\|\bar{\theta}(0) - \theta^*\|^2\}, \) and \( c_4 := \alpha \cdot \frac{8Mr_{\max}^2}{(1 - \lambda_2^W)^2} + \frac{16\beta^2\|\theta^*\|^2 + 32r_{\max}^2}{-\lambda_{\text{max}}} \), then it is straightforward from (55) that our desired result follows; that is,
\[
E[\|\theta_m(k) - \theta^*\|^2] \leq c_3^k V_0 + c_4 \alpha
\]
which concludes the proof.

\[ \Box \]

**E Proof of Lemma 2**

**Proof.** For notational brevity, let \( r_g(k) := (1/M) \sum_{m \in M} r_m(k) \) for each \( k \in \mathbb{N}^+ \). It then follows that
\[
\| \frac{1}{K^M} \sum_{j=k}^{k+K-1} E[G^T(\Theta, \xi_j)1|F(k)] - \bar{g}(\theta) \| \\
= \| \frac{1}{K} \sum_{j=k}^{k+K-1} E[\phi(s(k))[\gamma \phi(s(k+1)) - \phi(s(k))]^T \bar{\theta} + \frac{1}{M} \phi(s(k))r^T(k)1] - E[\bar{\theta}(\theta)] \| \\
= \| \frac{1}{K} \sum_{j=k}^{k+K-1} \sum_{s \in S} \Pr[s(j) = s|F(k)] - \pi(s) \left[ \phi(s)(\gamma P(s, s') \phi(s') - \phi(s))^T (\bar{\theta} + \theta^*) + r_g(s)\phi(s) \right] \| \\
\leq \max_{s, s'} \left\| \phi(s)[\gamma P(s, s') \phi(s') - \phi(s)]^T (\bar{\theta} + \theta^*) + r_g(s)\phi(s) \right\| \\
\times \frac{1}{K} \sum_{j=k}^{k+K-1} \sum_{s \in S} \Pr[s(j) = s|F(k)] - \pi(s) \\
\leq (1 + \gamma)(\|\bar{\theta} - \theta^*\| + 2\|\theta^*\| + r_{\max}) \times \frac{1}{K} \sum_{j=k}^{k+K-1} \nu_0 \rho^k \cdot \rho^{j-k} \\
\leq \frac{(1 + \gamma)\nu_0 \rho^k}{(1 - \rho)K} (\|\bar{\theta} - \theta^*\| + 2\|\theta^*\| + r_{\max}) \\
\leq \sigma_k(K)(\|\bar{\theta} - \theta^*\| + 1)
\]

(57)
where \( \sigma_k(K) = (1+\gamma)^{\sigma_0} \times \max \{2\|\theta^*\|+r_{\max},1\} \), and the second inequality arises from the fact that any finite-state, irreducible, and aperiodic Markov chains converges geometrically fast (with some initial constant \( \nu_0 > 0 \) and rate \( 0 < \rho < 1 \)) to its unique stationary distribution [Levin and Peres, 2017, Thm. 4.9]. Thus, we conclude that Lemma 2 holds true with monotonically decreasing function \( \sigma(K) \) of \( K \in \mathbb{N}^+ \) as defined above.

### F Proof of Lemma 3

**Proof.** Recalling the definition of our multi-step Lyapunov function, we obtain that

\[
\mathbb{E} [\mathcal{V}(k+1) - \mathcal{V}(k) | \mathcal{F}(k)] = \mathbb{E} [\|\hat{\theta}(k+1) - \theta^*\|^2 - \|\hat{\theta}(k) - \theta^*\|^2 | \mathcal{F}(k)].
\] (58)

Thus, we should next derive the bound of the right hand side of above equation. Following from iterate (27), we can write

\[
\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{\alpha}{M} \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1.
\] (59)

As a consequence (without particular statement, the expectation in the rest of this proof is taken with respect to the \( \xi_k \) to \( \xi_{k+\rho-1} \) conditioned on \( \xi_0 \) to \( \xi_{k-1} \)),

\[
\mathbb{E} [\|\hat{\theta}(k+1) - \theta^*\|^2 | \mathcal{F}(k)] = \mathbb{E} [\|\hat{\theta}(k) - \theta^* + \frac{\alpha}{M} \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1\|^2 | \mathcal{F}(k)]
\]

\[
= \mathbb{E} [\|\hat{\theta}(k) - \theta^* + \frac{\alpha}{M} \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1\|^2 | \mathcal{F}(k)]
\]

\[
= \mathbb{E} [\|\hat{\theta}(k) - \theta^*\|^2 + 2\alpha \mathbb{E} \left[ \left( \hat{\theta}(k) - \theta^*, \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1 - G^T(\Theta(k), \xi_j) 1 + G^T(\Theta(k), \xi_j) 1 \right) \mathcal{F}(k) \right]
\]

\[
+ \alpha^2 \mathbb{E} \left[ \left( \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1 - G^T(\Theta(k), \xi_j) 1 + G^T(\Theta(k), \xi_j) 1 \right) \mathcal{F}(k) \right]
\]

\[
= \|\hat{\theta}(k) - \theta^*\|^2 + 2\alpha \mathbb{E} \left[ \left( \hat{\theta}(k) - \theta^*, \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1 - G^T(\Theta(k), \xi_j) 1 + G^T(\Theta(k), \xi_j) 1 \right) \mathcal{F}(k) \right]
\]

\[
+ \alpha^2 \mathbb{E} \left[ \left( \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1 - G^T(\Theta(k), \xi_j) 1 + G^T(\Theta(k), \xi_j) 1 \right) \mathcal{F}(k) \right]
\]

\[
(60)
\]

where the second and the third equality result from adding and subtracting the same terms and the last equality holds since \( \bar{g}(\theta^*) = 0 \). In the following, we will bound the four terms in the above equality.

#### 1) Bounding the second term.

As a direct result of the definition of \( \bar{g}(\theta) \), we have that \( \bar{g}(\theta) - \bar{g}(\theta^*) = H(\theta - \theta^*) \). Therefore, it holds that

\[
2\alpha \mathbb{E} \left[ \left( \hat{\theta}(k) - \theta^*, \sum_{j=k}^{k+\rho-1} G^T(\Theta(j), \xi_j) 1 - G^T(\Theta(k), \xi_j) 1 + G^T(\Theta(k), \xi_j) 1 \right) \mathcal{F}(k) \right] = 2\alpha K \mathbb{E} \left[ (\hat{\theta}(k) - \theta^*)^\top H(\theta(k) - \theta^*) | \mathcal{F}(k) \right]
\]
where $\lambda_{\text{max}}^H$ is the largest eigenvalue of $H$. Because $H$ is a negative definite matrix, it holds that $\lambda_{\text{max}}^H < 0$.

2) **Bounding the third term.** Defining first $p(k, \Theta(k), K) := \sum_{j=k}^{k+K-1} \frac{1}{M} \left[ G^\top(\Theta(j), \xi_j)1 - G^\top(\Theta(k), \xi_k)1 \right]$, then it follows that

\[
p(k, \Theta(k), K) = \frac{1}{M} \sum_{j=k}^{k+K-2} \left[ G^\top(\Theta(j), \xi_j)1 - G^\top(\Theta(k), \xi_k)1 \right]
+ \frac{1}{M} \left[ G^\top(\Theta(k + K - 1), \xi_{k+K-1})1 - G^\top(\Theta(k), \xi_{k+K-1})1 \right]
= p(k, \Theta(k), K - 1) + \frac{1}{M} \left[ G^\top(\Theta(k + K - 1), \xi_{k+K-1})1 - G^\top(\Theta(k), \xi_{k+K-1})1 \right]
= p(k, \Theta(k), K - 1) + H(k + K - 1) \bar{\theta}(k + K - 1) - \bar{\theta}(k).
\]

Recalling that $2$ is the largest absolute value of eigenvalues of $H(k)$ for any $k \in \mathbb{N}^+$ (which clearly exists and is bounded due to the bounded feature vectors $\phi(s)$ for any $s \in S$), the norm of $p(k, \Theta(k), K)$ can be bounded as follows

\[
\| p(k, \Theta(k), K) \| \leq \| p(k, \Theta(k), K - 1) \| + 2 \| \bar{\theta}(k + K - 1) - \bar{\theta}(k) \|
= \| p(k, \Theta(k), K - 1) \| + 2 \alpha \sum_{j=k}^{k+K-2} \frac{1}{M} \left[ G^\top(\Theta(j), \xi_j)1 - G^\top(\Theta(k), \xi_k)1 \right]
+ \sum_{j=k}^{k+K-2} \frac{1}{M} G^\top(\Theta(k), \xi_j)1
\leq (1 + 2\alpha)\| p(k, \Theta(k), K - 1) \| + 2 \sum_{j=k}^{k+K-2} \alpha \| H(j)\bar{\theta}(k) + b_G \|
\leq (1 + 2\alpha)\| p(k, \Theta(k), K - 1) \| + 4\alpha \left( \sum_{j=k}^{k+K-2} \| \bar{\theta}(k) \| + \frac{r_{\text{max}}}{2} \right)
\]

where the last inequality follows from $\| H(j)\bar{\theta}(k) \| \leq 2\| \bar{\theta}(k) \|$ for any $j \geq 0$. Following the above recursion, we can write

\[
\| p(k, \Theta(k), K) \| \leq (1 + 2\alpha)^K \| p(k, \Theta(k), 0) \| + 4\alpha K \| \bar{\theta}(k) \| \sum_{j=0}^{K-1} (1 + 2\alpha)^2 (K - 1 - j)
\leq 4\alpha (\| \bar{\theta}(k) \| + \frac{r_{\text{max}}}{2}) \sum_{j=0}^{K-1} (1 + 2\alpha)^2 (K - 1 - j)
\]

where the second inequality because $\| p(k, \Theta(k), 0) \| = 0$.

For any positive constant $x \neq 1$ and $K \in \mathbb{N}^+$, the following equality holds

\[
\sum_{j=0}^{K-1} x^j (K - 1 - j) = \frac{x^K - Kx + K - 1}{(1 - x)^2}.
\]

Substituting $x = 1 + 2\alpha$ into (63) along with plugging the result into (62) yields

\[
\| p(k, \Theta(k), K) \| \leq \frac{(1 + 2\alpha)^K - 2K\alpha - 1}{\alpha} K \| \bar{\theta}(k) \|.
\]

According to the mid-value theorem, there exists some suitable constant $\delta \in [0, 1]$ such that the following holds true

\[
(1 + 2\alpha)^K = 1 + 2K\alpha + \frac{1}{2} K(K - 1)(1 + \delta(2\alpha)^{K-2}(2\alpha)^2)
\]
Thus, it is clear that
\[
\frac{(1 + 2\alpha)^{K} - 2K\alpha - 1}{\alpha} \leq 2aK^{2}(1 + 2\alpha)^{K-2}.
\] (66)

Upon plugging (66) into (64), it follows that
\[
\|p(k, \Theta(k), K)\| \leq 2aK^{2}(1 + 2\alpha)^{K-2}((\|\bar{\theta}(k)\| + \frac{r_{\max}}{2})
\leq 2aK^{2}(1 + 2\alpha)^{K-2}((\|\bar{\theta}(k) - \theta^{*}\| + \|\theta^{*}\| + \frac{r_{\max}}{2})).
\] (67)

Now, we turn to the third term in (60)
\[
2\alpha\mathbb{E}\left[\langle \bar{\theta}(k) - \theta^{*}, \sum_{j=k}^{k+K-1} \frac{1}{M} G^{T}(\Theta(j), \xi_{j})1 - G^{T}(\Theta(k), \xi_{j})1 \rangle \mid \mathcal{F}(k)\right]
\leq 2\alpha\mathbb{E}\left[\|\bar{\theta}(k) - \theta^{*}\| \cdot \|p(k, \Theta(k), K)\| \mid \mathcal{F}(k)\right]
\leq 4a^{2}K^{2}(1 + 2\alpha)^{K-2}(\|\bar{\theta}(k) - \theta^{*}\| \cdot (\|\bar{\theta}(k) - \theta^{*}\| + \|\theta^{*}\| + \frac{r_{\max}}{2} + \frac{r_{\max}}{8})
\leq 4a^{2}K^{2}(1 + 2\alpha)^{K-2}(2\|\bar{\theta}(k) - \theta^{*}\|^{2} + \frac{1}{4} \|\theta^{*}\|^{2} + \frac{r_{\max}}{8}).
\] (68)

where the second inequality is obtained by plugging in (67), and the last one follows from the inequality $a(a + b) \leq 2a^{2} + (1/4)b^{2}$.

3) **Bounding the fourth term.** It follows that
\[
2\alpha\mathbb{E}\left[\langle \bar{\theta}(k) - \theta^{*}, \sum_{j=k}^{k+K-1} \frac{1}{M} G^{T}(\Theta(k), \xi_{j})1 - K\tilde{g}(\bar{\theta}(k)) \rangle \mid \mathcal{F}(k)\right]
\leq 2\alpha\mathbb{E}\left[\|\bar{\theta}(k) - \theta^{*}\| \cdot \mathbb{E}\left[\sum_{j=k}^{k+K-1} \frac{1}{M} G^{T}(\Theta(k), \xi_{j}) \mid \mathcal{F}(k)\right]\right]
\leq 2aK\sigma(K)\|\bar{\theta}(k) - \theta^{*}\| \cdot (\|\bar{\theta}(k) - \theta^{*}\| + 1)
\leq 2aK\sigma(K)\left(2\|\bar{\theta}(k) - \theta^{*}\|^{2} + \frac{1}{4}\right).
\] (69)

4) **Bounding the last term.** Evidently, we have that
\[
\left\|\frac{1}{M} \sum_{j=k}^{k+K-1} G^{T}(\Theta(j), \xi_{j})1 - G^{T}(\Theta(k), \xi_{j})1 + G^{T}(\Theta(k), \xi_{j})1\right\|^{2}
\leq 2 \|p(k, \Theta(k), K)\|^{2} + 2 \left\|\sum_{j=k}^{k+K-1} \frac{1}{M} G^{T}(\Theta(k), \xi_{j})1\right\|^{2}
\leq 2 \|p(k, \Theta(k), K)\|^{2} + 2 \left\|\sum_{j=k}^{k+K-1} H(j)\bar{\theta}(k) + \frac{1}{M} r^{T}(j)1\phi(j)\right\|^{2}
\leq 16a^{2}K^{4}(1 + 2\alpha)^{2K-4}\|\bar{\theta}(k)\|^{2} + 16K\|\bar{\theta}(k)\|^{2} + \left[a^{2}K^{4}(1 + 2\alpha)^{2K-4} + 4K\right]r_{\max}^{2}.
Proof.

G Proof of Lemma 4

From the definition of our multi-step Lyapunov function, we obtain that

\[ \alpha \text{in} \]

where the last inequality is due to the fact that functions \( \Gamma \) and the second is obtained by plugging in (67). Hence, upon taking expectation of both sides of (70) conditioning on \( \mathcal{F}(k) \), we arrive at

\[
\alpha^2 \mathbb{E} \left[ \left\| \frac{k^{K-1}}{M} \sum_{j=k}^{k+K-1} \left[ G^\top \Theta(j, \xi_j) 1 - G^\top \Theta(k, \xi_j) 1 + G^\top \Theta(k, \xi_j) 1 \right] \right\|^2 | \mathcal{F}(k) \right] \\
\leq 32 \alpha^4 K^6 (1 + 2\alpha)^{2K-4} + 32 K \alpha^2 \left( \| \tilde{\theta}(k) - \theta^* \|^2 + \| \theta^* \|^2 \right) + \alpha^2 \left[ \alpha^2 K^4 (1 + 2\alpha)^{2K-4} + 4 K \right] r_{\text{max}}^2.
\]

We have successfully bounded each of the four terms in (60). Putting now together the bounds in (61), (68), (69), and (71) into (60), we finally arrive at

\[
\mathbb{E} \left[ \| \tilde{\theta}(k + K) - \theta^* \|^2 | \mathcal{F}(k) \right] \leq \left[ 1 + 2 \alpha T \lambda^H_{\text{max}} + \alpha \Gamma_1(\alpha, K) \right] \| \tilde{\theta}(k) - \theta^* \|^2 + \alpha \Gamma_2(\alpha, K)
\]

where

\[
\Gamma_1(\alpha, K) = 32 \alpha^3 K^4 (1 + 2\alpha)^{2K-4} + 32 K \alpha + 8 \alpha K^2 (1 + 2\alpha)^K - 4 K \sigma(K)
\]

\[
\Gamma_2(\alpha, K) = \left[ 32 \alpha^3 K^4 (1 + 2\alpha)^{2K-4} + 32 K \alpha + \alpha K^2 (1 + 2\alpha)^K \right] \| \theta^* \|^2 \\
+ \left[ 4 \alpha^3 K^4 (1 + 2\alpha)^{2K-4} + \frac{1}{2} \alpha K^2 (1 + 2\alpha)^K - 4 K \right] r_{\text{max}}^2 + \frac{1}{2} K \sigma(K)
\]

From the definition of our multi-step Lyapunov function, we obtain that

\[
\mathbb{E} \left[ V(k + 1) - V(k) | \mathcal{F}(k) \right] = \mathbb{E} \left[ \| \tilde{\theta}(k + K) - \theta^* \|^2 | \mathcal{F}(k) \right] - \| \tilde{\theta}(k) - \theta^* \|^2 \\
\leq \alpha \left[ 2 K \lambda^H_{\text{max}} + \Gamma_1(\alpha, K) \right] \| \tilde{\theta}(k) - \theta^* \|^2 + \alpha \Gamma_2(\alpha, K)
\]

\[
\leq \alpha \left[ 2 K \lambda^H_{\text{max}} + \Gamma_2(\alpha, K) \right] \| \tilde{\theta}(k) - \theta^* \|^2 + \alpha \Gamma_2(\alpha, K)
\]

where the last inequality is due to the fact that functions \( \Gamma_1(\alpha, K) \) and \( \Gamma_2(\alpha, K) \) are monotonically increasing in \( \alpha \). This concludes the proof.

\[ \square \]

G Proof of Lemma 4

Proof. It is straightforward to check that

\[
\| \tilde{\theta}(k + i) - \theta^* \|^2 = \| \tilde{\theta}(k + i - 1) - \theta^* + \frac{\alpha}{M} G^\top (\Theta(k + i - 1), \xi_{k+i-1}) 1 \\
- \frac{\alpha}{M} G^\top (1(\theta^* - \theta^*)^\top, \xi_{k+i-1}) 1 + \frac{\alpha}{M} G^\top (1(\theta^* - \theta^*)^\top, \xi_{k+i-1}) 1 \|^2 \\
\leq \| \tilde{\theta}(k + i - 1) - \theta^* \|^2 + 3 \alpha^2 \| H(k)(\theta^* - \theta^* \|^2 \\
+ 3 \alpha^2 \| H(k)(\theta^* - \theta^* \|^2 + \frac{1}{M} \phi(s(k)) r^\top(k) 1 \|^2 \\
\leq (3 + 12 \alpha^2) \| \tilde{\theta}(k + i - 1) - \theta^* \|^2 + 6 \alpha^2 \max(3 + 12 \alpha^2) \| \theta^* \|^2 + r_{\text{max}}^2 \\
\leq (3 + 12 \alpha^2) \| \tilde{\theta}(k) - \theta^* \|^2 + 6 \alpha^2 \max(3 + 12 \alpha^2) \| \theta^* \|^2 + \sum_{j=0}^{i-1} (3 + 12 \alpha^2)^j.
\]

(76)
As a result, $\mathcal{V}(k)$ can be bounded as

$$
\mathcal{V}(k) = \sum_{i=0}^{K_2-1} \|\bar{\theta}(k+i) - \theta^*\|^2
$$

$$
\leq \sum_{i=0}^{K_2-1} (3 + 12\alpha^2)^i \|\bar{\theta}(k) - \theta^*\|^2 + 6\alpha^2(4\|\theta^*\|^2 + r_{\max}) \sum_{j=0}^{K_2-1} \sum_{i=1}^{j} (3 + 12\alpha^2)^j
$$

$$
= \frac{(3 + 12\alpha^2)^{K_2-1} - 1}{2 + 12\alpha^2} \|\bar{\theta}(k) - \theta^*\|^2
$$

$$
+ \alpha^2 \frac{6(3 + 12\alpha^2)[(3 + 12\alpha^2)^{K_2-1} - 1] - 6K_\varphi + 6(4\|\theta^*\|^2 + r_{\max}^2)}{(2 + 12\alpha^2)^2}
$$

With $c_5 := \frac{(3+12\alpha_{\max}^2)^{K_\varphi-1}}{2+3\alpha_{\max}^2}$ and $c_6 := \frac{6(3+12\alpha_{\max}^2)[(3+12\alpha_{\max}^2)^{K_\varphi-1} - 1] - 6K_\varphi + 6}{2+12\alpha_{\max}^2}$, we conclude that

$$
\mathcal{V}(k) \leq c_5 \|\bar{\theta}(k) - \theta^*\|^2 + \alpha^2 c_6.
$$

(78)

### H Proof of Theorem 3

**Proof.** The convergence of $\mathbb{E}[\|\bar{\theta}(k) - \theta^*\|^2]$ is separately addressed in two phases:

1) The time instant $k < k_\alpha$, with $k_\alpha = \max\{k|\|\theta^k\| \geq \alpha\}$, namely, it holds that $\alpha\sigma(K) \leq \sigma_k(K) \leq \sigma(K)$ for any $k < k_\alpha$;

2) The time instant $k \geq k_\alpha$, i.e., it holds that $\sigma_k(K) \leq \alpha\sigma(K)$ for any $k \geq k_\alpha$.

**Convergence of the first phase**

From Lemma 4, we have

$$
-\|\bar{\theta}(k) - \theta^*\|^2 \leq -\frac{1}{c_5} \mathcal{V}(k) + \frac{\alpha^2 c_6}{c_5}.
$$

(79)

Substituting (79) into (75), and rearranging the terms give the recursion of Lyapunov function as follows

$$
\mathbb{E}[\mathcal{V}(k+1)|\mathcal{F}(k)] \leq \left\{ 1 + \frac{1}{c_5} \left[ 2\alpha K_\varphi \lambda^H_{\max} + \alpha \Gamma_1(\alpha_{\max}, K_\varphi) \right] \right\} \mathbb{E}[\mathcal{V}(k)|\mathcal{F}(k)]
$$

$$
+ \alpha \left\{ \alpha \Gamma_2(\alpha, K_\varphi) - \frac{\alpha^2 c_6}{c_5} \left[ 2K_\varphi \lambda^H_{\max} + \Gamma_1(\alpha_{\max}, K_\varphi) \right] \right\}
$$

$$
\leq c_7 \mathbb{E}[\mathcal{V}(k)|\mathcal{F}(k)] + \alpha c_8
$$

(80)

where $c_7 := 1 + \frac{1}{2c_5} \alpha_{\max} K_\varphi \lambda^H_{\max} \in (0, 1)$; constant $c_8 := \Gamma_2(\alpha_{\max}, K_\varphi) - \frac{\alpha_{\max}^2 c_6}{c_5} K_\varphi \lambda^H_{\max} > 0$, and the last inequality holds true because of (41).

Deducing from (80), we obtain that

$$
\mathbb{E}[\mathcal{V}(k)] \leq c_7^k \mathbb{E}[\mathcal{V}(0)] + \alpha c_8 \frac{1 - c_7^k}{1 - c_7}
$$

$$
\leq c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + \alpha^2 c_6 c_7^k + \alpha c_8 \frac{1 - c_7^k}{1 - c_7}
$$

$$
\leq c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + \alpha^2 c_6 + \frac{\alpha c_8}{1 - c_7}
$$

(81)

$$
= c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + \alpha^2 c_6 - \frac{2c_5 c_8}{K_\varphi \lambda^H_{\max}}
$$

(82)

Recalling the definition of Lyapunov function, it is obvious that

$$
\mathbb{E}[\|\bar{\theta}(k) - \theta^*\|^2] \leq \mathbb{E}[\mathcal{V}(k)] \leq c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + \alpha^2 c_6 - \frac{2c_5 c_8}{K_\varphi \lambda^H_{\max}}
$$

(83)
which finishes the proof of the first phase.

**Convergence of the second phase**

Without repeating similar derivation, we directly have that the following holds for $\sigma_k(K) \leq \alpha_2(K)$:

\[
\Gamma_1(\alpha, K) := 32\alpha^3 K^4 (1 + 2\alpha)^{2K-4} + 32 K \alpha + 8 \alpha K^2 (1 + 2\alpha)^{K-2} + 4 K \alpha \sigma(K) \\
\Gamma_2(\alpha, K) := [32\alpha^3 K^4 (1 + 2\alpha)^{2K-4} + 32 K \alpha + \alpha K^2 (1 + 2\alpha)^{K-2}] \|\theta^*\|^2 \\
\quad + [4\alpha^3 K^4 (1 + 2\alpha)^{2K-4} + \frac{1}{2} \alpha K^2 (1 + 2\alpha)^{K-2} + 4 \alpha K] r_{\text{max}} + \frac{1}{2} K \alpha \sigma(K).
\]

(84)

Subsequently, we have the following recursion of $V(k)$ that is similar to but slightly different from (80).

\[
E[V(k+1)|F(k)] \leq c_7 E[V(k)|F(k)] + \alpha^2 c_8', \quad \forall k \geq k_\alpha
\]

(86)

where $c_8' := \frac{[16 \alpha_{\text{max}}^2 K_0^3 (1 + 2\alpha_{\text{max}})^{2K-4} + 32 K_0 + 2 K_0^3 (1 + 2\alpha_{\text{max}})^{K-2}] \|\theta^*\|^2 + 4 K_0 r_{\text{max}}^2 - \frac{1}{8} K_0 \lambda_{\text{max}} - \frac{\alpha_{\text{max}} c_5 K_0 H_{\text{max}}}{c_5} \frac{\alpha_{\text{max}} c_5 K_0 H_{\text{max}}}{c_5}$. It is easy to check that $c_8' \geq c_8$ due to the fact that $\alpha_{\text{max}} < 1$ in our case. Repeatedly applying the above recursion from $k = k_\alpha$ to any $k > k_\alpha$ yields

\[
E[V(k)] \leq c_7^{k-k_\alpha} E[V(k_\alpha)] + \alpha^2 c_8 \frac{1 - c_7^{k-k_\alpha}}{1 - c_7}
\]

\[
\leq c_7^{k-k_\alpha} \left( c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + \alpha^2 c_6 - \frac{2 c_5 c_8}{K_0 \lambda_{\text{max}}} \right) - \alpha \frac{2 c_5 c_8'}{K_0 \lambda_{\text{max}}}
\]

\[
\leq c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + c_7^{k-k_\alpha} \alpha^2 c_6 - (c_7^{k-k_\alpha} + \alpha) \frac{2 c_5 c_8'}{K_0 \lambda_{\text{max}}}
\]

(87)

where we have used $c_8 \leq c_8'$ for simplicity. Again, using the definition of the Lyapunov function and (87), it follows that

\[
E[\|\bar{\theta}(k) - \theta^*\|^2] \leq c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 + c_7^{k-k_\alpha} \alpha^2 c_6 - (c_7^{k-k_\alpha} + \alpha) \frac{2 c_5 c_8'}{K_0 \lambda_{\text{max}}}, \quad \forall k \geq k_\alpha
\]

(88)

Combining the results in the above two phases, we conclude that the following bound holds for any $k \in \mathbb{N}^+$

\[
E[\|\bar{\theta}(k) - \theta^*\|^2] \leq c_5 c_7^k \|\bar{\theta}(0) - \theta^*\|^2 - \frac{2 c_5 c_8'}{K_0 \lambda_{\text{max}}} \alpha + \min\{1, c_7^{k-k_\alpha}\} \times \left( \alpha^2 c_6 - \frac{2 c_5 c_8'}{K_0 \lambda_{\text{max}}} \right).
\]

(89)

\[\square\]

![Figure 1: Consensus and convergence of decentralized TD(0) learning](image)

(a) Average parameter norm  
(b) Local parameters’ norm  
(c) Local parameters
I SIMULATIONS

In order to verify our analytical results, we carried out experiments on a multi-agent networked system. The details of our experimental setup are as follows: the number of agents $M = 30$, the state space size $|\mathcal{S}| = 100$ with each state $s$ being a vector of length $|s| = 20$, the dimension of learning parameter $\theta$ is $p = 10$, the reward upper bound $r_{\text{max}} = 10$, and the stepsize $\alpha = 0.01$. The feature vectors are cosine functions, that is, $\phi(s) = \cos(As)$, where $A \in \mathbb{R}^{p \times |s|}$ is a randomly generated matrix. The communication weight matrix $W$ depicting the neighborhood of the agents including the topology and the weights was generated randomly, with each agent being associated with 5 neighbors on average. As illustrated in Fig. 1(a), the parameter average $\bar{\theta}$ converges to a small neighborhood of the optimum at a linear rate. To demonstrate the consensus among agents, convergence of the parameter norms $\|\theta_m\|$ for $m = 1, 2, 3, 4$ is presented in Fig. 1(b), while that of their first elements $|\theta_{m,1}|$ is depicted in Fig. 1(c). The simulation results corroborate our theoretical analysis.