SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Multi-agent robust policy evaluation for reinforcement learning via primal-dual online time-averaging

Gang CHEN[1*], Changli PU[1], Yaoyao ZHOU[1], Xiumin LI[1] & Huimiao CHEN[2]

[1]*College of Automation, Chongqing University, Chongqing 400044, China;*
[2]*Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China*

1 Appendix A

In this paper, a sequence of digraphs is considered, which is formally stated as follows.

Assumption 1. The sequence of graphs $\{G(t)\}_{t\geqslant 1}$ are uniformly jointly strongly connected and weight-balanced, i.e., there is a finite integer $B>0$ such that the graph $G(t)\cup G(t+1)\cup\ldots\cup G(t+B-1)$ is strongly connected and weight-balanced. Moreover, the edge weight satisfies $a_{ij}\geqslant o$ for some positive constant o as $(i,j)\in E$.

In light of the distributed policy evaluation algorithm, we have

$$\theta_{j,t+1}=\sum_{i=1}^{N}\phi_{ji}(t,1)\theta_{i,1}-\sum_{i=1}^{N}\sum_{s=1}^{t-1}\phi_{ji}(t,s+1)\eta_s g_{\theta_i,s}-\eta_t g_{\theta_j,t} \tag{1}$$

where the state transition matrix $\phi(t,l)=[\phi_{ji}(t,l)]$ with elements $\phi_{ji}(t,l)$ denoted by

$$\phi(t,l)=\hat{\mathbf{P}}(t)\hat{\mathbf{P}}(t-1)\ldots\hat{\mathbf{P}}(l),\ t\geqslant l$$

with $\hat{P}(t)=I_N-\sigma L_t$.

Under Assumption 1, the state transition matrix $\phi(t,l)$ has the following property.

Lemma 1. While Assumption 1 holds, one has

$$|\phi_{ji}(t,s)-\frac{1}{N}|\leqslant \nu r^{t-s}\ \textit{for all}\ t\geqslant s,$$

with $\nu=2\frac{1+o^{-B}}{1-o^{B}}$, $r=(1-o^{B})^{\frac{1}{B}}$.

We first show that the consensus constraints are satisfied for the distributed policy evaluation algorithm, and the proof of Theorem 1 proceeds as follows.

1.1 The proof of Theorem 1

Taking the summation on both sides of (1), we have

$$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t+1}=\frac{1}{N}\sum_{i=1}^{N}\theta_{i,1}-\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\eta_s g_{\theta_i,s}-\frac{1}{N}\sum_{i=1}^{N}\eta_t g_{\theta_i,t} \tag{2}$$

According to (1) and (2), one has

$$\theta_{j,t+1}-\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t+1}=\sum_{i=1}^{N}(\phi_{ji}(t,1)-\frac{1}{N})\theta_{i,1}$$

* Corresponding author (email: chengang@cqu.edu.cn)

$$-\sum_{i=1}^{N}\sum_{s=1}^{t-1}(\phi_{ji}(t,s+1)-\frac{1}{N})\eta_s g_{\theta_{i,s}}-\eta_t g_{\theta_{j,t}}+\frac{1}{N}\sum_{i=1}^{N}\eta_t g_{\theta_{i,t}} \tag{3}$$

Moreover, it yields from (3) that

$$\begin{aligned}\|\theta^a_{j,T+1}-\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_{i,t+1}\| \leqslant&\frac{1}{T}\sum_{t=1}^{T}\|\theta_{j,t+1}-\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t+1}\|\\ \leqslant&\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}|\phi_{ji}(t,1)-\frac{1}{N}|\cdot\|\theta_{i,1}\|\\ &+\frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\eta_s|\phi_{ji}(t,s+1)-\frac{1}{N}|\cdot\|g_{\theta_{i,s}}\|\\ &+\frac{1}{T}\sum_{t=1}^{T}(\eta_t(\|g_{\theta_{j,t}}\|+\frac{1}{N}\sum_{i=1}^{N}\eta_t\|g_{\theta_{i,t}}\|)\end{aligned} \tag{4}$$

Let $\|\theta_{i,1}\|\leqslant\theta_0$ and $\|g_{\theta_{j,t}}\|\leqslant g_{\theta_0}$ with $\theta_0>0, g_{\theta_0}>0$. By Lemma 1, we further get that

$$\begin{aligned}\|\theta^a_{j,T+1}-\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_{i,t+1}\| \leqslant&\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_0\nu r^{t-1}\\ &+\frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{i=1}^{N}g_{\theta_0}\nu\eta_s r^{t-s-1}+\frac{1}{T}\sum_{i=1}^{T}2g_{\theta_0}\eta_t\\ \leqslant&\frac{N\nu\theta_0}{(1-r)T}+\frac{\nu N g_{\theta_0}}{(1-r)T}\sum_{t=1}^{T-1}\eta_t+\frac{2g_{\theta_0}}{T}\sum_{t=1}^{T}\eta_t\end{aligned}$$

While $\eta_t=\frac{\varepsilon}{T^{\triangle}}$, we have

$$\|\theta^a_{j,T+1}-\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_{i,t+1}\|\leqslant\frac{N\nu\theta_0}{(1-r)T}+\frac{2g_{\theta_0}\varepsilon}{T^{\triangle}}+\frac{\nu N g_{\theta_0}\varepsilon}{(1-r)T^{\triangle}} \tag{5}$$

While $\eta_t=\frac{\varepsilon}{\sqrt{t}}$, by noting $\sum_{t=1}^{T}\frac{\varepsilon}{\sqrt{t}}\leqslant\varepsilon(1+\int_{t=1}^{T}\frac{1}{\sqrt{t}}dt)\leqslant 2\varepsilon\sqrt{T}$, we get that

$$\|\theta^a_{j,T+1}-\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_{i,t+1}\|\leqslant\frac{N\nu\theta_0}{(1-r)T}+\frac{4g_{\theta_0}\varepsilon}{\sqrt{T}}+\frac{2\nu N g_{\theta_0}\varepsilon}{(1-r)\sqrt{T}} \tag{6}$$

From (5) and (6), we see that the consensus constraints for θ_i, $i=1,\cdots,N$ are satisfied as $T\to\infty$. The similar proof and results are also applicable for the dual variable ω_i, $i=1,\cdots,N$.

The following lemma is useful in the following analyses.

Lemma 2. Under Assumption 1, the following bound holds.

$$\sum_{t=1}^{T}\|\boldsymbol{L}\boldsymbol{\theta}_t\|\leqslant 2N\theta_0+\frac{N^2\theta_0\nu}{1-r}+2Ng_{\theta_0}\sum_{t=1}^{T-1}\eta_t+N^2 g_{\theta_0}\nu\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}\eta_s r^{t-s-1}$$

Proof. Note that

$$\begin{aligned}\|\boldsymbol{L}\boldsymbol{\theta}_{t+1}\|=&\|\boldsymbol{\theta}_{t+1}-1_d\otimes(\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t+1})\|\\ \leqslant&\sum_{i=1}^{N}\|\theta_{i,t+1}-\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t+1}\|\end{aligned}$$

From (3) and Lemma 1, we further get that

$$\|\boldsymbol{L}\boldsymbol{\theta}_{t+1}\| \leqslant \sum_{i=1}^{N}(N\nu\theta_0 r^{t-1} + \sum_{s=1}^{t-1} N\nu g_{\theta_0}\eta_s r^{t-s-1} + 2g_{\theta_0}\eta_t)$$

Noting that

$$\begin{aligned}\sum_{t=1}^{T}\|\boldsymbol{L}\boldsymbol{\theta}_t\| &= \sum_{t=0}^{T-1}\|\boldsymbol{L}\boldsymbol{\theta}_{t+1}\| \\ &= \|\boldsymbol{L}\boldsymbol{\theta}_1\| + \sum_{t=1}^{T-1}\|\boldsymbol{L}\boldsymbol{\theta}_{t+1}\|\end{aligned}$$

and applying the inequalities

$$\|\boldsymbol{L}\boldsymbol{\theta}_1\| \leqslant 2N\theta_0$$

and

$$\sum_{t=1}^{T-1}\|\boldsymbol{L}\boldsymbol{\theta}_{t+1}\| \leqslant \frac{N^2\nu\theta_0}{1-r} + 2Ng_{\theta_0}\sum_{t=1}^{T-1}\eta_t + N^2\nu g_{\theta_0}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}\eta_s r^{t-s-1}$$

we get the result Lemma 2.

Then, we construct the evaluation error, and find the cumulative error bounds corresponding to the primal and dual variables, respectively.

Lemma 3. Let $\{(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)\}_{t\geqslant 1}$ be a sequence of the iterative process, and $(\boldsymbol{\theta}_p, \boldsymbol{\omega}_p)$ be the variable at any time p, $1 \leqslant p \leqslant t$. The bound of primal variable evaluation error is as follows:

$$\begin{aligned}2(J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t) - J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_t)) \leqslant & \frac{1}{\eta_t}(\|\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{\theta}_p\|^2 - \|\boldsymbol{M}\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_p\|^2) \\ & + 2\|g_{\boldsymbol{\theta}_t}\|(\|\boldsymbol{L}\boldsymbol{\theta}_t\| + \|\boldsymbol{L}\boldsymbol{\theta}_p\|) + \eta_t\|g_{\boldsymbol{\theta}_t}\|^2. \end{aligned} \tag{7}$$

Similarly, for the dual variable, we have

$$\begin{aligned}2(J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t) - J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_p)) \geqslant & -\frac{1}{\eta_t}(\|\boldsymbol{M}\boldsymbol{\omega}_t - \boldsymbol{\omega}_p\|^2 - \|\boldsymbol{M}\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_p\|^2) \\ & - 2\|g_{\boldsymbol{\omega}_t}\|(\|\boldsymbol{L}\boldsymbol{\omega}_t\| + \|\boldsymbol{L}\boldsymbol{\omega}_p\|) - \eta_t\|g_{\boldsymbol{\omega}_t}\|^2. \end{aligned} \tag{8}$$

Proof. Multiplying $\boldsymbol{M}$ on both sides of the first equation in the distributed policy evaluation algorithm and noting that $\boldsymbol{M}\boldsymbol{L}_t = 0$, we can derive the following equation

$$\boldsymbol{M}\boldsymbol{\theta}_{t+1} = \boldsymbol{M}\boldsymbol{\theta}_t - \eta_t\boldsymbol{M}g_{\boldsymbol{\theta}_t} \tag{9}$$

By subtracting $\boldsymbol{\theta}_p$ on both sides of (9) and applying the relationships $\boldsymbol{M}^T = \boldsymbol{M}, \boldsymbol{M}^2 = \boldsymbol{M}$, we have

$$\|\boldsymbol{M}\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_p\|^2 = \|\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{\theta}_p\|^2 + \|\eta_t\boldsymbol{M}g_{\boldsymbol{\theta}_t}\|^2 - 2\eta_t g_{\boldsymbol{\theta}_t}^\top(\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{M}\boldsymbol{\theta}_p). \tag{10}$$

Using the convexity of $J(\boldsymbol{\theta}, \boldsymbol{\omega})$ with respect to $\boldsymbol{\theta}$, the last term of (10) can be bounded as follows

$$\begin{aligned}-g_{\boldsymbol{\theta}_t}^\top(\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{M}\boldsymbol{\theta}_p) &= -g_{\boldsymbol{\theta}_t}^\top(\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{\theta}_t) - g_{\boldsymbol{\theta}_t}^\top(\boldsymbol{\theta}_p - \boldsymbol{M}\boldsymbol{\theta}_p) - g_{\boldsymbol{\theta}_t}^\top(\boldsymbol{\theta}_t - \boldsymbol{\theta}_p) \\ &\leqslant g_{\boldsymbol{\theta}_t}^\top\boldsymbol{L}\boldsymbol{\theta}_t - g_{\boldsymbol{\theta}_t}^\top\boldsymbol{L}\boldsymbol{\theta}_p + J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_t) - J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t). \end{aligned} \tag{11}$$

Combining (10) with (11), we have

$$\begin{aligned}J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t) - J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_t) \leqslant & \frac{1}{2\eta_t}(\|\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{\theta}_p\|^2 - \|\boldsymbol{M}\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_p\|^2) \\ & + g_{\boldsymbol{\theta}_t}^\top\boldsymbol{L}\boldsymbol{\theta}_t - g_{\boldsymbol{\theta}_t}^\top\boldsymbol{L}\boldsymbol{\theta}_p + \frac{1}{2\eta_t}\|\eta_t\boldsymbol{M}g_{\boldsymbol{\theta}_t}\|^2. \end{aligned} \tag{12}$$

According to the Cauchy-Schwarz inequality,

$$g_{\boldsymbol{\theta}_t}^{\top} \boldsymbol{L}\boldsymbol{\theta}_t - g_{\boldsymbol{\theta}_t}^{\top} \boldsymbol{L}\boldsymbol{\theta}_p \leqslant \|g_{\boldsymbol{\theta}_t}\|(\|\boldsymbol{L}\boldsymbol{\theta}_t\| + \|\boldsymbol{L}\boldsymbol{\theta}_p\|),$$

and the facts that $\|\boldsymbol{M}\| \leqslant 1$ and

$$\|\eta_t \boldsymbol{M} g_{\boldsymbol{\theta}_t}\| \leqslant \eta_t \|\boldsymbol{M}\| \|g_{\boldsymbol{\theta}_t}\| \leqslant \eta_t \|g_{\boldsymbol{\theta}_t}\|,$$

we further get that

$$\begin{aligned} J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t) - J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_t) \leqslant & \frac{1}{2\eta_t}(\|\boldsymbol{M}\boldsymbol{\theta}_t - \boldsymbol{\theta}_p\|^2 - \|\boldsymbol{M}\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_p\|^2) \\ & + \|g_{\boldsymbol{\theta}_t}\|(\|\boldsymbol{L}\boldsymbol{\theta}_t\| + \|\boldsymbol{L}\boldsymbol{\theta}_p\|) + \frac{\eta_t}{2}\|g_{\boldsymbol{\theta}_t}\|^2. \end{aligned} \tag{13}$$

Similarly, for any dual variable ω_p, we can derive that

$$\begin{aligned} J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t) - J(\boldsymbol{\theta}_t, \boldsymbol{\omega}_p) \geqslant & -\frac{1}{2\eta_t}(\|\boldsymbol{M}\boldsymbol{\omega}_t - \boldsymbol{\omega}_p\|^2 - \|\boldsymbol{M}\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_p\|^2) \\ & - \|g_{\boldsymbol{\omega}_t}\|(\|\boldsymbol{L}\boldsymbol{\omega}_t\| + \|\boldsymbol{L}\boldsymbol{\omega}_p\|) - \frac{\eta_t}{2}\|g_{\boldsymbol{\omega}_t}\|^2. \end{aligned} \tag{14}$$

Based on Lemma 3, we will provide the cumulative estimation error corresponding to the online time-averages.

Lemma 4. For any $(\boldsymbol{\theta}_p, \boldsymbol{\omega}_p)$, $1 \leqslant p \leqslant t$, we have the following cumulative error corresponding to the dual averaging

$$\sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}_p, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \leqslant \frac{u(t, \boldsymbol{\theta}_p)}{2} \tag{15}$$

where

$$\begin{aligned} u(t, \boldsymbol{\theta}_p) = & \sum_{s=2}^{t} (\|\boldsymbol{M}\boldsymbol{\theta}_s - \boldsymbol{\theta}_p\|^2)(\frac{1}{\eta_s} - \frac{1}{\eta_{s-1}}) + \frac{2}{\eta_1}(\|\boldsymbol{\theta}_1\|^2 + \|\boldsymbol{\theta}_p\|^2) \\ & + \sum_{s=1}^{t} \eta_s \|g_{\boldsymbol{\theta}_s}\|^2 + 2\sum_{s=1}^{t} \|g_{\boldsymbol{\theta}_s}\| \|\boldsymbol{L}\boldsymbol{\theta}_s\| + 2\|\boldsymbol{L}\boldsymbol{\theta}_p\| \sum_{s=1}^{t} \|g_{\boldsymbol{\theta}_s}\|. \end{aligned} \tag{16}$$

Correspondingly, the cumulative error with respect to the primal averaging is

$$-\frac{u(t, \boldsymbol{\omega}_p)}{2} \leqslant \sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}_{t+1}^{\mathrm{a}} \boldsymbol{\omega}_p), \tag{17}$$

Proof. Based on Lemma 3, making the summation over $s = 1, \cdots, t$ on both sides of (7), we derive that

$$\begin{aligned} 2\sum_{s=1}^{t}(J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_s)) \leqslant & \sum_{s=2}^{t}(\|\boldsymbol{M}\boldsymbol{\theta}_s - \boldsymbol{\theta}_p\|^2)(\frac{1}{\eta_s} - \frac{1}{\eta_{s-1}}) \\ & + \frac{1}{\eta_1}\|\boldsymbol{M}\boldsymbol{\theta}_1 - \boldsymbol{\theta}_p\|^2 + \sum_{s=1}^{t} \eta_s \|g_{\boldsymbol{\theta}_s}\|^2 \\ & + 2\sum_{s=1}^{t} \|g_{\boldsymbol{\theta}_s}\| \|\boldsymbol{L}\boldsymbol{\theta}_s\| + 2\|\boldsymbol{L}\boldsymbol{\theta}_p\| \sum_{s=1}^{t} \|g_{\boldsymbol{\theta}_s}\|. \end{aligned} \tag{18}$$

Noting that $\|\boldsymbol{M}\boldsymbol{\theta}_1 - \boldsymbol{\theta}_p\|^2 \leqslant 2\|\boldsymbol{\theta}_1\|^2 + 2\|\boldsymbol{\theta}_p\|^2$, we have

$$2\sum_{s=1}^{t}(J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_s)) \leqslant u(t, \boldsymbol{\theta}_p) \tag{19}$$

Since the function $J(\boldsymbol{\theta}, \boldsymbol{\omega})$ is concave with respective to $\boldsymbol{\omega}$, one has

$$J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \geqslant \frac{1}{t}\sum_{s=1}^{t} J(\boldsymbol{\theta}_p, \boldsymbol{\omega}_s) \tag{20}$$

Based on (19) and (20), we get (15). Similarly, the lower bound (17) can be derived by making the summation over $s = 1, \cdots, t$ on (14) and utilizing the convexity of $J(\boldsymbol{\theta}, \boldsymbol{\omega})$ with respective to $\boldsymbol{\theta}$, i.e.,$J(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}_p) \leqslant \frac{1}{t}\sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_p)$.

In Lemma 4, we have established the upper and lower bounds of the cumulative errors. Let $(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*)$ denote the saddle point, which satisfies

$$J(\boldsymbol{\theta}^*, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \leqslant J(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*) \leqslant J(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}^*) \tag{21}$$

We further get the following fundamental results on the online time-average evaluations.

Lemma 5. Let $(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*)$ be the saddle point of $J(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*)$. The following inequalities hold:

$$\begin{aligned} -u(t, \boldsymbol{\omega}^*) - u(t, \boldsymbol{\theta}_{t+1}^{\mathrm{a}}) \leqslant & 2t(J(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) - J(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*)) \\ \leqslant & u(t, \boldsymbol{\theta}^*) + u(t, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \end{aligned}$$

Proof. Replacing $\boldsymbol{\theta}_p$ with $\boldsymbol{\theta}^*$ in (15) and applying (21), we have

$$\sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*) \leqslant \sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}^*, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \leqslant \frac{u(t, \boldsymbol{\theta}^*)}{2}. \tag{22}$$

Replacing $\boldsymbol{\omega}_p$ with $\boldsymbol{\omega}^*$ in (17) and using (21), we get that

$$\sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*) \geqslant \sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}^*) \geqslant -\frac{u(t, \boldsymbol{\omega}^*)}{2}. \tag{23}$$

By (22) and (23), we have

$$-\frac{u(t, \boldsymbol{\omega}^*)}{2} \leqslant \sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*) \leqslant \frac{u(t, \boldsymbol{\theta}^*)}{2}. \tag{24}$$

Letting $\boldsymbol{\theta}_p = \boldsymbol{\theta}_{t+1}^{\mathrm{a}}$ in (15) yields

$$\sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \leqslant \frac{u(t, \boldsymbol{\theta}_{t+1}^{\mathrm{a}})}{2} \tag{25}$$

Similarly, letting $\boldsymbol{\omega}_p = \boldsymbol{\omega}_{t+1}^{a}$ in (17) yields

$$-\frac{u(t, \boldsymbol{\omega}_{t+1}^{\mathrm{a}})}{2} \leqslant \sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) - tJ(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) \tag{26}$$

Combining (25) and (26), we get that

$$-\frac{u(t, \boldsymbol{\theta}_{t+1}^{\mathrm{a}})}{2} \leqslant tJ(\boldsymbol{\theta}_{t+1}^{\mathrm{a}}, \boldsymbol{\omega}_{t+1}^{\mathrm{a}}) - \sum_{s=1}^{t} J(\boldsymbol{\theta}_s, \boldsymbol{\omega}_s) \leqslant \frac{u(t, \boldsymbol{\omega}_{t+1}^{\mathrm{a}})}{2}. \tag{27}$$

According to (24) and (27), we get Lemma 5.

1.2 The proof of Theorem 2

From Lemma 5, we have

$$J(\boldsymbol{\theta}_{T+1}^{\mathrm{a}},\boldsymbol{\omega}_{T+1}^{\mathrm{a}})-J(\boldsymbol{\theta}^*,\boldsymbol{\omega}^*)\leqslant\frac{u(T,\boldsymbol{\theta}^*)+u(T,\boldsymbol{\omega}_{T+1}^{\mathrm{a}})}{2T}\tag{28}$$

Next, we analyze the terms $u(T,\boldsymbol{\theta}^*)$ and $u(T,\boldsymbol{\omega}_{T+1}^{\mathrm{a}})$ one by one. Noting that $\eta_t=\frac{\varepsilon}{T^{\triangle}}$ and $\boldsymbol{L}\boldsymbol{\theta}^*=0$, we get from (16) that

$$\begin{aligned}u(T,\boldsymbol{\theta}^*)&=\frac{2T^{\triangle}}{\varepsilon}(\|\boldsymbol{\theta}_1\|^2+\|\boldsymbol{\theta}^*\|^2)+\frac{\varepsilon}{T^{\triangle}}\sum_{s=1}^{T}\|g_{\theta_s}\|^2+2\sum_{s=1}^{T}\|g_{\boldsymbol{\theta}_s}\|\|\boldsymbol{L}\boldsymbol{\theta}_s\|\\&\leqslant\frac{4T^{\triangle}\theta_0^2}{\varepsilon}+\varepsilon g_{\theta_0}^2T^{1-\triangle}+2g_{\theta_0}\sum_{s=1}^{T}\|\boldsymbol{L}\boldsymbol{\theta}_s\|\end{aligned}\tag{29}$$

According to Lemma 2, we have

$$\begin{aligned}\sum_{s=1}^{T}\|\boldsymbol{L}\boldsymbol{\theta}_s\|&\leqslant2N\theta_0+\frac{N^2\theta_0\nu}{1-r}+2Ng_{\theta_0}\varepsilon T^{1-\triangle}+N^2g_{\theta_0}\nu\varepsilon T^{-\triangle}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}r^{t-s-1}\\&\leqslant2N\theta_0+\frac{N^2\theta_0\nu}{1-r}+(2Ng_{\theta_0}\varepsilon+\frac{N^2g_{\boldsymbol{\theta}_0}\nu\varepsilon}{1-r})T^{1-\triangle}\end{aligned}\tag{30}$$

Substituting (30) into (29) and making some reorganization yield

$$\begin{aligned}u(T,\boldsymbol{\theta}^*)\leqslant&\frac{T^{\triangle}}{\varepsilon}(4\theta_0^2+4N\theta_0g_{\theta_0}+\frac{2N^2\theta_0\nu g_{\theta_0}}{1-r}\\&+(g_{\theta_0}^2\varepsilon^2+4Ng_{\theta_0}^2\varepsilon^2+\frac{2N^2g_{\theta_0}^2\nu\varepsilon^2}{1-r})T^{1-2\triangle})\end{aligned}\tag{31}$$

For the term $u(T,\boldsymbol{\omega}_{T+1}^{\mathrm{a}})$, we get from (16) that

$$u(T,\boldsymbol{\omega}_{T+1}^{\mathrm{a}})\leqslant\frac{4T^{\triangle}\omega_0^2}{\varepsilon}+\varepsilon g_{\omega_0}^2T^{1-\triangle}+2g_{\omega_0}\sum_{s=1}^{T}\|\boldsymbol{L}\boldsymbol{\omega}_s\|+2g_{\omega_0}\|\boldsymbol{L}\boldsymbol{\omega}_{T+1}^{\mathrm{a}}\|T$$

In light of $\|\boldsymbol{L}\boldsymbol{\omega}_{T+1}^{\mathrm{a}}\|\leqslant\frac{1}{T}\sum_{s=1}^{T}\|\boldsymbol{L}\boldsymbol{\omega}_s\|$, we further get that

$$\begin{aligned}u(T,\boldsymbol{\omega}_{T+1}^{\mathrm{a}})\leqslant&\frac{4T^{\triangle}\omega_0^2}{\varepsilon}+\varepsilon g_{\omega_0}^2T^{1-\triangle}+4g_{\omega_0}\sum_{s=1}^{T}\|\boldsymbol{L}\boldsymbol{\omega}_s\|\\\leqslant&\frac{T^{\triangle}}{\varepsilon}(4\omega_0^2+8N\omega_0g_{\omega_0}+\frac{4N^2\omega_0\nu g_{\omega_0}}{1-r}\\&+(g_{\omega_0}^2\varepsilon^2+8Ng_{\omega_0}^2\varepsilon^2+\frac{4N^2g_{\omega_0}^2\nu\varepsilon^2}{1-r})T^{1-2\triangle})\end{aligned}\tag{32}$$

Substituting (31) and (32) into (28), and according to Lemma 5, we have

$$J(\boldsymbol{\theta}^*,\boldsymbol{\omega}^*)-J(\boldsymbol{\theta}_{T+1}^{\mathrm{a}},\boldsymbol{\omega}_{T+1}^{\mathrm{a}})\leqslant\frac{u(T,\boldsymbol{\omega}^*)+u(T,\boldsymbol{\theta}_{T+1}^{\mathrm{a}})}{2T}\tag{33}$$

Applying the similar dirivation process as (28)-(33), we can get Theorem 2.

1.3 The proof of Theorem 3

From Lemma 5, we have

$$J(\boldsymbol{\theta}_{t+1}^{\mathrm{a}},\boldsymbol{\omega}_{t+1}^{\mathrm{a}})-J(\boldsymbol{\theta}^*,\boldsymbol{\omega}^*)\leqslant\frac{u(t,\boldsymbol{\theta}^*)+u(t,\boldsymbol{\omega}_{t+1}^{\mathrm{a}})}{2t}\tag{34}$$

For the term $u(t,\boldsymbol{\theta}^*)$, we get from (16) that

$$\begin{aligned}u(t,\boldsymbol{\theta}^*) &\leqslant 4\theta_0^2\varepsilon\sum_{s=2}^{t}\frac{1}{\sqrt{s}}+\frac{4\theta_0^2}{\varepsilon}+g_{\theta_0}^2\varepsilon\sum_{s=1}^{t}\frac{1}{\sqrt{s}}+2g_{\theta_0}\sum_{s=1}^{t}\|\boldsymbol{L}\boldsymbol{\theta}_s\|\\&\leqslant\frac{4\theta_0^2}{\varepsilon}+(8\theta_0^2\varepsilon+2g_{\theta_0}^2\varepsilon)\sqrt{t}+2g_{\theta_0}\sum_{s=1}^{t}\|\boldsymbol{L}\boldsymbol{\theta}_s\| \end{aligned}\tag{35}$$

From Lemma 2, we have

$$\begin{aligned}\sum_{s=1}^{t}\|\boldsymbol{L}\boldsymbol{\theta}_s\| \leqslant &2N\theta_0+\frac{N^2\theta_0\nu}{1-r}+2Ng_{\theta_0}\varepsilon\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}}\\&+N^2g_{\theta_0}\nu\varepsilon\sum_{l=2}^{t-1}\sum_{s=1}^{l-1}\frac{r^{t-s-1}}{\sqrt{s}}\\\leqslant&2N\theta_0+\frac{N^2\theta_0\nu}{1-r}+(4Ng_{\theta_0}\varepsilon+\frac{2N^2g_{\theta_0}\nu\varepsilon}{1-r})\sqrt{t}\end{aligned}\tag{36}$$

Substituting (36) into (35), we have

$$u(t,\boldsymbol{\theta}^*)\leqslant 2\alpha'_{\theta_1}+2\alpha'_{\theta_2}\sqrt{t}\tag{37}$$

For the term $u(t,\boldsymbol{\omega}_{t+1}^{\mathrm{a}})$, we get from (16) that

$$\begin{aligned}u(t,\boldsymbol{\omega}_{t+1}^{\mathrm{a}})\leqslant&4\omega_0^2\sum_{s=1}^{t}\frac{\varepsilon}{\sqrt{s}}+\frac{4\omega_0^2}{\varepsilon}+g_{\omega_0}^2\sum_{s=1}^{t}\frac{\varepsilon}{\sqrt{s}}\\&+2g_{\omega_0}\sum_{s=1}^{t}\|\boldsymbol{L}\boldsymbol{\omega}_s\|+2g_{\omega_0}\|\boldsymbol{L}\boldsymbol{\omega}_{t+1}^{\mathrm{a}}\|t\\\leqslant&\frac{4\omega_0^2}{\varepsilon}+(8\omega_0^2\varepsilon+2g_{\omega_0}^2\varepsilon)\sqrt{t}+4g_{\omega_0}\sum_{s=1}^{t}\|\boldsymbol{L}\boldsymbol{\omega}_s\|\end{aligned}\tag{38}$$

Since the term $\sum_{s=1}^{t}\|\boldsymbol{L}\boldsymbol{\omega}_s\|$ satisfies the similar relationship as that of (36), we further get that

$$u(t,\boldsymbol{\omega}_{t+1}^{\mathrm{a}})\leqslant 2\beta'_{\omega_1}+2\beta'_{\omega_2}\sqrt{t}\tag{39}$$

By (34), (37) and (39), we get the inequality on the right side of Theorem 3. From Lemma 5, we have

$$J(\boldsymbol{\theta}^*,\boldsymbol{\omega}^*)-J(\boldsymbol{\theta}_{t+1}^{\mathrm{a}},\boldsymbol{\omega}_{t+1}^{\mathrm{a}})\leqslant\frac{u(t,\boldsymbol{\omega}^*)+u(t,\boldsymbol{\theta}_{t+1}^{\mathrm{a}})}{2t}\tag{40}$$

Using the similar derivation process as that of (35)-(39), we can get Theorem 3.