Decentralized Online Optimization

December 20, 2018

Abstract

ddd

1 Notations and assumptions

$$\mathcal{R}_T^{DOG} = \sum_{i=1}^{n}\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\xi_{i,t}\sim D_{i,t}} f_{i,t}(\mathbf{x}_{i,t};\xi_{i,t}) - \mathop{\mathbb{E}}_{\xi_{i,t}\sim D_{i,t}} f_{i,t}(\mathbf{x}_t^*;\xi_{i,t}),$$

where, for any $\mathbf{x}$,

$$f_{i,t}(\mathbf{x};\xi_{i,t}) := \beta g_{i,t}(\mathbf{x}) + (1-\beta)h_t(\mathbf{x};\xi_{i,t})$$

with $0<\beta<1$

Define the dynamic regret as

$$\mathcal{R}_T^{DOG} = \sum_{i=1}^{n}\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\xi} f_{i,t}(\mathbf{x}_{i,t};\xi) - \mathop{\mathbb{E}}_{\xi} f_{i,t}(\mathbf{x}_t^*;\xi),$$

where, for any $\mathbf{x}$,

$$f_{i,t}(\mathbf{x};\xi) := \beta g_{i,t}(\mathbf{x}) + (1-\beta)f(\mathbf{x};\xi)$$

with $0<\beta<1$, and ξ is a random variable drawn from an unknown distribution. $g_{i,t}$ is an adversary loss function, and f is a given loss function.

The budget of the dynamics is defined as

$$\sum_{t=1}^{T} \left\| \mathbf{x}_{t+1}^* - \mathbf{x}_t^* \right\| \le D.$$

Assumption 1. *We make the following assumptions.*

- *For any $i\in[n]$ and $t\in[T]$, we assume* $\max\left\{\left\|\nabla g_{i,t}(\mathbf{x})\right\|^2, \left\|\nabla f(\mathbf{x};\xi)\right\|^2, \left\|\nabla f_{i,t}(\mathbf{x};\xi)\right\|^2\right\} \le G$, *and* $\left\|\nabla f(\mathbf{x};\xi) - \nabla f(\mathbf{x})\right\|^2 \le \sigma^2$.
- *For any $\mathbf{x}$ and $\mathbf{y}$, we assume $\left\|\mathbf{x}-\mathbf{y}\right\|^2 \le R$.*
- *For any $i\in[n]$ and $t\in[T]$, we assume the function $f_{i,t}(\mathbf{x};\xi)$ is convex and differentiable.*

Algorithm 1 DOG: Decentralized Online Gradient.

Require: The learning rate η, number of iterations T, and the confusion matrix $\mathbf{W}$.

for $t = 1, 2, ..., T$ **do**

  For the i-th node with $i \in [n]$:

    Predict $\mathbf{x}_{i,t}$.

    Observe the loss function $f_{i,t}$,

    and suffer loss $f_{i,t}(\mathbf{x}_{i,t}; \xi)$.

  Update:

    Query the gradient $\nabla f_{i,t}(\mathbf{x}_{i,t}; \xi)$.

    $\mathbf{x}_{i,t+1} = \sum_{j=1}^{n} \mathbf{W}_{i,j} \mathbf{x}_{j,t} - \eta \nabla f_{i,t}(\mathbf{x}_{i,t}; \xi)$.

2 Algorithm

Theorem 1. *Using Assumptions 1, and choosing $\eta > 0$ in Algorithm 1, we have*

$$\begin{aligned}
\mathcal{R}_T \leq & \sum_{t=1}^{T} \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\xi} f_{i,t}(\mathbf{x}_{i,t}; \xi) - \mathop{\mathbb{E}}_{\xi} f_{i,t}(\mathbf{x}_t^*; \xi) \\
\leq & \left(1 + 4\beta \left(\frac{1}{\sqrt{\beta^2 + \frac{1}{n}}} + 4\right)\right) nT\beta G\eta + \left(L + \frac{2\eta L^2 (1-\beta)^2}{\sqrt{\beta^2 + \frac{1}{n}}} + \frac{\sqrt{\beta^2 + \frac{1}{n}}}{2\eta} + 10\eta L^2 + \frac{\beta}{2\eta}\right) \frac{nT\eta^2 G}{(1-\rho)^2} \\
& + n \left(\frac{1}{\sqrt{\beta^2 + \frac{1}{n}}} + 4\right) (f(\bar{\mathbf{x}}_1) - f(\bar{\mathbf{x}}_{T+1})) + T\eta\sigma^2(1-\beta) \left(\frac{2}{\sqrt{\beta^2 + \frac{1}{n}}} + 9\right) + \frac{n}{2\eta} \left(4\sqrt{R}D + R\right).
\end{aligned}$$

Corollary 1. *Using Assumptions 1, and choosing*

$$\eta = \sqrt{\frac{nD}{T\left(n\beta + n\sqrt{\frac{n\beta^2+1}{n}} + 2\sqrt{\frac{n}{n\beta^2+1}} + 9\right)}}$$

in Algorithm 1, we have

$$\mathcal{R}_T^{DOG} \leq \mathcal{O}\left(\sqrt{n^2\beta DT + n\sqrt{n}DT}\right).$$

Furthermore, when

$$\beta = n^{-a}, \text{ and } a \geq 0,$$

we have

$$\mathcal{R}_T^{DOG} \leq \mathcal{O}\left(n^{\max\{\frac{3}{4}, \frac{2-a}{2}\}} \sqrt{DT}\right).$$

Appendix

Proof to Theorem 1:

Proof.

$$\mathop{\mathbb{E}}_{\xi} \frac{1}{n} \sum_{i=1}^{n} f_{i,t}(\mathbf{x}_{i,t}; \xi) - f_t(\mathbf{x}_t^*; \xi)$$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{i=1}^{n}\beta\left(g_{i,t}(\mathbf{x}_{i,t})-g_{i,t}(\mathbf{x}_t^*)\right)+(1-\beta)\mathop{\mathbb{E}}_{\xi}\left(f(\mathbf{x}_{i,t};\xi)-f(\mathbf{x}_t^*;\xi)\right)\\
\le&\frac{1}{n}\sum_{i=1}^{n}\beta\left\langle\nabla g_{i,t}(\mathbf{x}_{i,t}),\mathbf{x}_{i,t}-\mathbf{x}_t^*\right\rangle+(1-\beta)\mathop{\mathbb{E}}_{\xi}\left\langle\nabla f(\mathbf{x}_{i,t};\xi),\mathbf{x}_{i,t}-\mathbf{x}_t^*\right\rangle\\
=&\frac{1}{n}\sum_{i=1}^{n}\beta\left(\left\langle\nabla g_{i,t}(\mathbf{x}_{i,t}),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle+\left\langle\nabla g_{i,t}(\mathbf{x}_{i,t}),\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\rangle+\left\langle\nabla g_{i,t}(\mathbf{x}_{i,t}),\bar{\mathbf{x}}_{t+1}-\mathbf{x}_t^*\right\rangle\right)\\
&+\frac{1}{n}\sum_{i=1}^{n}(1-\beta)\left(\mathop{\mathbb{E}}_{\xi}\left\langle\nabla f(\mathbf{x}_{i,t};\xi),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle+\mathop{\mathbb{E}}_{\xi}\left\langle\nabla f(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\rangle+\mathop{\mathbb{E}}_{\xi}\left\langle\nabla f(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_{t+1}-\mathbf{x}_t^*\right\rangle\right)\\
=&\underbrace{\frac{1}{n}\sum_{i=1}^{n}\beta\left(\left\langle\nabla g_{i,t}(\mathbf{x}_{i,t}),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle+\left\langle\nabla g_{i,t}(\mathbf{x}_{i,t}),\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\rangle\right)}_{I_1(t)}\\
&+\underbrace{\frac{1}{n}\sum_{i=1}^{n}(1-\beta)\left(\left\langle\nabla f(\mathbf{x}_{i,t}),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle+\mathop{\mathbb{E}}_{\xi}\left\langle\nabla f(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\rangle\right)}_{I_2(t)}\\
&+\underbrace{\mathop{\mathbb{E}}_{\xi}\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_{t+1}-\mathbf{x}_t^*\right\rangle}_{I_3(t)}
\end{aligned}
$$

Now, we begin to bound $I_1(t)$.

$$
\begin{aligned}
I_1(t)\le&\frac{\beta}{n}\sum_{i=1}^{n}\left(\frac{\eta}{2}\left\|\nabla g_{i,t}(\mathbf{x}_{i,t})\right\|^2+\frac{1}{2\eta}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\frac{\eta}{2}\left\|\nabla f_{i,t}(\mathbf{x}_{i,t})\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2\right)\\
\le&\beta G\eta+\frac{\beta}{2n\eta}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\frac{\beta}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2.
\end{aligned}
$$

Now, we begin to bound $I_2(t)$.

$$
I_2(t)=(1-\beta)\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left\langle\nabla f(\mathbf{x}_{i,t}),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle}_{J_1(t)}+\underbrace{\mathop{\mathbb{E}}_{\xi}\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\rangle}_{J_2(t)}\right).
$$

For $J_1(t)$ and $\nu>0$, we have

$$
\begin{aligned}
J_1(t)=&\frac{1}{n}\sum_{i=1}^{n}\left\langle\nabla f(\mathbf{x}_{i,t}),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\langle\nabla f(\mathbf{x}_{i,t})-\nabla f(\bar{\mathbf{x}}_t),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle+\frac{1}{n}\sum_{i=1}^{n}\left\langle\nabla f(\bar{\mathbf{x}}_t),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle\\
\le&\frac{L}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\frac{1}{n}\sum_{i=1}^{n}\left\langle\nabla f(\bar{\mathbf{x}}_t),\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\rangle\\
\le&\frac{L}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\eta}{2\nu}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\frac{\nu}{2\eta}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2\right)
\end{aligned}
$$

$$\leq \frac{L}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\frac{\eta}{2\nu}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\frac{\nu}{2\eta n}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2. \tag{1}$$

According to Lemma 1, we have

$$\begin{aligned}
&\frac{\eta}{2}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2\\
\leq&\frac{\eta}{2}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\left(\frac{\eta}{2}-\frac{L\eta^2}{2}\right)\left\|\frac{1}{n}\sum_{i=1}^{n}f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\\
\leq&f(\bar{\mathbf{x}}_t)-f(\bar{\mathbf{x}}_{t+1})+4G\eta\beta^2+\frac{2\eta L^2(1-\beta)^2}{n}\sum_{i=1}^{n}\left\|\bar{\mathbf{x}}_t-\mathbf{x}_{i,t}\right\|^2+\frac{2\eta\sigma^2(1-\beta)^2}{n}.
\end{aligned}\tag{2}$$

Substituting (2) into (1), we obtain

$$\begin{aligned}
&J_1(t)\\
\leq&\frac{L}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\left(\frac{1}{\nu}\left(f(\bar{\mathbf{x}}_t)-f(\bar{\mathbf{x}}_{t+1})\right)+\frac{4G\eta\beta^2}{\nu}+\frac{2\eta L^2(1-\beta)^2}{n\nu}\sum_{i=1}^{n}\left\|\bar{\mathbf{x}}_t-\mathbf{x}_{i,t}\right\|^2+\frac{2\eta\sigma^2(1-\beta)^2}{n\nu}\right)\\
&+\frac{\nu}{2n\eta}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2\\
=&\left(\frac{L}{n}+\frac{2\eta L^2(1-\beta)^2}{n\nu}+\frac{\nu}{2n\eta}\right)\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+\frac{1}{\nu}\left(f(\bar{\mathbf{x}}_t)-f(\bar{\mathbf{x}}_{t+1})\right)+\frac{4G\eta\beta^2}{\nu}+\frac{2\eta\sigma^2(1-\beta)^2}{n\nu}.
\end{aligned}$$

For $J_2(t)$, we have

$$\begin{aligned}
J_2(t)=&\mathop{\mathbb{E}}_{\xi}\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\rangle\\
\leq&\left(\frac{\eta}{2}\mathop{\mathbb{E}}_{\xi}\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi)\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2\right)\\
\leq&\left(\frac{\eta}{2}\mathop{\mathbb{E}}_{\xi}\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi)-\nabla f(\mathbf{x}_{i,t})+\nabla f(\mathbf{x}_{i,t})\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2\right)\\
\leq&\left(\eta\mathop{\mathbb{E}}_{\xi}\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi)-\nabla f(\mathbf{x}_{i,t})\right\|^2+\eta\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t})\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2\right)\\
\leq&\frac{\eta}{n}\sigma^2+\eta\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t})-\nabla f(\bar{\mathbf{x}}_t)+\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2\\
\leq&\frac{\eta}{n}\sigma^2+2\eta\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t})-\nabla f(\bar{\mathbf{x}}_t)\right\|^2+2\eta\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2\\
\leq&\frac{\eta}{n}\sigma^2+\frac{2\eta L^2}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2+2\eta\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\frac{1}{2\eta}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2
\end{aligned}$$

Recall Lemma 1, and we have

$J_2(t)$

$$
\begin{aligned}
&\le \frac{\eta}{n}\sigma^2 + \frac{2\eta L^2}{n}\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 + 4\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) + 16G\eta\beta^2 + \frac{8\eta L^2(1-\beta)^2}{n}\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 \\
&\quad + \frac{8\eta\sigma^2(1-\beta)^2}{n} + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2 \\
&\le \frac{\eta\sigma^2(1-8(1-\beta)^2)}{n}\sigma^2 + \left(2\eta L^2 + 8\eta L^2(1-\beta)^2\right)\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 + 4\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) + 16G\eta\beta^2 + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2 \\
&\le \frac{9\eta\sigma^2}{n}\sigma^2 + \left(2\eta L^2 + 8\eta L^2(1-\beta)^2\right)\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 + 4\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) + 16G\eta\beta^2 + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
&I_2(t) \\
=&(1-\beta)(J_1(t) + J_2(t)) \\
\le&(1-\beta)\left(\left(\frac{L}{n} + \frac{2\eta L^2(1-\beta)^2}{n\nu} + \frac{\nu}{2n\eta} + \frac{\eta L^2(2+8(1-\beta)^2)}{n}\right)\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 + \left(\frac{1}{\nu} + 4\right)\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right)\right) \\
&+ (1-\beta)\left(4G\eta\beta^2\left(\frac{1}{\nu} + 4\right) + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2\right) + \frac{2\eta\sigma^2(1-\beta)^3}{n\nu} + \frac{9\eta\sigma^2(1-\beta)}{n} \\
\le&(1-\beta)\left(\left(\frac{L}{n} + \frac{2\eta L^2(1-\beta)^2}{n\nu} + \frac{\nu}{2n\eta} + \frac{10\eta L^2}{n}\right)\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2\right) + \left(\frac{1}{\nu} + 4\right)\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) \\
&+ (1-\beta)\left(4G\eta\beta^2\left(\frac{1}{\nu} + 4\right) + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2\right) + \frac{2\eta\sigma^2(1-\beta)^3}{n\nu} + \frac{9\eta\sigma^2(1-\beta)}{n}
\end{aligned}
$$

Combine those bounds of $I_1(t)$ and $I_2(t)$. We thus have

$$
\begin{aligned}
&I_1(t) + I_2(t) \\
\le&\beta G\eta + \frac{\beta}{2n\eta}\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 + \frac{\beta}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2 \\
&+ (1-\beta)\left(\left(\frac{L}{n} + \frac{2\eta L^2(1-\beta)^2}{n\nu} + \frac{\nu}{2n\eta} + \frac{10\eta L^2}{n}\right)\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2\right) + \left(\frac{1}{\nu} + 4\right)\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) \\
&+ (1-\beta)\left(4G\eta\beta^2\left(\frac{1}{\nu} + 4\right) + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2\right) + \frac{2\eta\sigma^2(1-\beta)^3}{n\nu} + \frac{\eta\sigma^2(1-\beta)}{n} \\
=&\left(1 + 4\beta(1-\beta)\left(\frac{1}{\nu} + 4\right)\right)\beta G\eta + \left((1-\beta)\left(\frac{L}{n} + \frac{2\eta L^2(1-\beta)^2}{n\nu} + \frac{\nu}{2n\eta} + \frac{10\eta L^2}{n}\right) + \frac{\beta}{2n\eta}\right)\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 \\
&+ \left(\frac{1}{\nu} + 4\right)\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2 + \frac{\eta\sigma^2(1-\beta)}{n}\left(\frac{2(1-\beta)^2}{\nu} + 9\right) \\
\le&\left(1 + 4\beta\left(\frac{1}{\nu} + 4\right)\right)\beta G\eta + \left(\frac{L}{n} + \frac{2\eta L^2(1-\beta)^2}{n\nu} + \frac{\nu}{2n\eta} + \frac{10\eta L^2}{n} + \frac{\beta}{2n\eta}\right)\sum_{i=1}^n \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 \\
&+ \left(\frac{1}{\nu} + 4\right)\left(f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1})\right) + \frac{1}{2\eta}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\|^2 + \frac{\eta\sigma^2(1-\beta)}{n}\left(\frac{2}{\nu} + 9\right).
\end{aligned}
$$

According to Lemma 2, we have

$$
\frac{1}{n}\sum_{i=1}^n\sum_{t=1}^T \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2 \le \frac{T\eta^2 G}{(1-\rho)^2}.
$$

Therefore,

$$
\begin{aligned}
&\sum_{t=1}^{T}(I_1(t)+I_2(t))\\
\le&\left(1+4\beta\left(\frac{1}{\nu}+4\right)\right)T\beta G\eta+\left(\frac{L}{n}+\frac{2\eta L^2(1-\beta)^2}{n\nu}+\frac{\nu}{2n\eta}+\frac{10\eta L^2}{n}+\frac{\beta}{2n\eta}\right)\sum_{t=1}^{T}\sum_{i=1}^{n}\left\|\mathbf{x}_{i,t}-\bar{\mathbf{x}}_t\right\|^2\\
&+\left(\frac{1}{\nu}+4\right)\left(f(\bar{\mathbf{x}}_1)-f(\bar{\mathbf{x}}_{T+1})\right)+\frac{1}{2\eta}\sum_{t=1}^{T}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2+\frac{T\eta\sigma^2(1-\beta)}{n}\left(\frac{2}{\nu}+9\right)\\
\le&\left(1+4\beta\left(\frac{1}{\nu}+4\right)\right)T\beta G\eta+\left(L+\frac{2\eta L^2(1-\beta)^2}{\nu}+\frac{\nu}{2\eta}+10\eta L^2+\frac{\beta}{2\eta}\right)\frac{T\eta^2 G}{(1-\rho)^2}\\
&+\left(\frac{1}{\nu}+4\right)\left(f(\bar{\mathbf{x}}_1)-f(\bar{\mathbf{x}}_{T+1})\right)+\frac{1}{2\eta}\sum_{t=1}^{T}\left\|\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_{t+1}\right\|^2+\frac{T\eta\sigma^2(1-\beta)}{n}\left(\frac{2}{\nu}+9\right).
\end{aligned}
$$

Now, we begin to bound $I_3(t)$. Recall that the update rule is

$$
\mathbf{x}_{i,t+1}=\sum_{j=1}^{n}\mathbf{W}_{ij}\mathbf{x}_{j,t}-\eta\nabla f_{i,t}(\mathbf{x}_{i,t};\xi).
$$

By taking average over $i\in[n]$ on both sides, we have

$$
\bar{\mathbf{x}}_{t+1}=\bar{\mathbf{x}}_t-\eta\left(\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right). \tag{3}
$$

Denote a new auxiliary function $h(\mathbf{z})$ as

$$
h(\mathbf{z})=\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi),\mathbf{z}\right\rangle+\frac{1}{2\eta}\left\|\mathbf{z}-\bar{\mathbf{x}}_t\right\|^2.
$$

Note that (3) is equivalent to

$$
\begin{aligned}
\bar{\mathbf{x}}_{t+1}=&\operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^d}h(\mathbf{z})\\
=&\operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^d}\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi),\mathbf{z}\right\rangle+\frac{1}{2\eta}\left\|\mathbf{z}-\bar{\mathbf{x}}_t\right\|^2.
\end{aligned}
$$

Furthermore, denote a new auxiliary variable $\bar{\mathbf{x}}_\tau$ as

$$
\bar{\mathbf{x}}_\tau=\bar{\mathbf{x}}_{t+1}+\tau\left(\mathbf{x}_t^*-\bar{\mathbf{x}}_{t+1}\right),
$$

where $0\le\tau\le1$. According to the optimality of $\bar{\mathbf{x}}_{t+1}$, we have

$$
\begin{aligned}
0\le&h(\bar{\mathbf{x}}_\tau)-h(\bar{\mathbf{x}}_{t+1})\\
=&\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi),\bar{\mathbf{x}}_\tau-\bar{\mathbf{x}}_{t+1}\right\rangle+\frac{1}{2\eta}\left(\left\|\bar{\mathbf{x}}_\tau-\bar{\mathbf{x}}_t\right\|^2-\left\|\bar{\mathbf{x}}_{t+1}-\bar{\mathbf{x}}_t\right\|^2\right)\\
=&\left\langle\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi),\tau\left(\mathbf{x}_t^*-\bar{\mathbf{x}}_{t+1}\right)\right\rangle+\frac{1}{2\eta}\left(\left\|\bar{\mathbf{x}}_{t+1}+\tau\left(\mathbf{x}_t^*-\bar{\mathbf{x}}_{t+1}\right)-\bar{\mathbf{x}}_t\right\|^2-\left\|\bar{\mathbf{x}}_{t+1}-\bar{\mathbf{x}}_t\right\|^2\right)
\end{aligned}
$$

$$=\left\langle \frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t};\xi), \tau\left(\mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}\right)\right\rangle + \frac{1}{2\eta}\left(\left\|\tau\left(\mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}\right)\right\|^2 + 2\left\langle \tau\left(\mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}\right), \bar{\mathbf{x}}_{t+1} - \bar{\mathbf{x}}_t\right\rangle\right).$$

Dividing τ on both sides, and letting τ be close to 0, we have

$$\begin{aligned}
I_3(t) =& \left\langle \frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t};\xi), \bar{\mathbf{x}}_{t+1} - \mathbf{x}_t^*\right\rangle \\
\le& \frac{1}{2\eta}\left(\lim_{\tau\to 0}\tau\left\|\left(\mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}\right)\right\|^2 + 2\left\langle \mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}, \bar{\mathbf{x}}_{t+1} - \bar{\mathbf{x}}_t\right\rangle\right) \\
=& \frac{1}{\eta}\left\langle \mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}, \bar{\mathbf{x}}_{t+1} - \bar{\mathbf{x}}_t\right\rangle \\
=& \frac{1}{2\eta}\left(\left\|\mathbf{x}_t^* - \bar{\mathbf{x}}_t\right\|^2 - \left\|\mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}\right\|^2 - \left\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\right\|^2\right).
\end{aligned} \tag{4}$$

Besides, we have

$$\begin{aligned}
& \left\|\mathbf{x}_{t+1}^* - \bar{\mathbf{x}}_{t+1}\right\|^2 - \left\|\mathbf{x}_t^* - \bar{\mathbf{x}}_{t+1}\right\|^2 \\
=& \left\|\mathbf{x}_{t+1}^*\right\|^2 - \left\|\mathbf{x}_t^*\right\|^2 - 2\left\langle \bar{\mathbf{x}}_{t+1}, -\mathbf{x}_t^* + \mathbf{x}_{t+1}^*\right\rangle \\
=& \left(\left\|\mathbf{x}_{t+1}^*\right\| - \left\|\mathbf{x}_t^*\right\|\right)\left(\left\|\mathbf{x}_{t+1}^*\right\| + \left\|\mathbf{x}_t^*\right\|\right) - 2\left\langle \bar{\mathbf{x}}_{t+1}, -\mathbf{x}_t^* + \mathbf{x}_{t+1}^*\right\rangle \\
\le& \left\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\right\|\left(\left\|\mathbf{x}_{t+1}^*\right\| + \left\|\mathbf{x}_t^*\right\|\right) - 2\left\|\bar{\mathbf{x}}_{t+1}\right\|\left\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\right\| \\
\le& 4\sqrt{R}\left\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\right\|. \quad (\text{due to } \left\|\mathbf{x}-\mathbf{y}\right\|^2 \le R, \forall \mathbf{x},\mathbf{y}\in\mathcal{X})
\end{aligned}$$

Thus, telescoping $I_3(t)$ over $t\in[T]$, we have

$$\begin{aligned}
\sum_{t=1}^T I_3(t) \le& \frac{1}{2\eta}\left(4\sqrt{R}\sum_{t=1}^T\left\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\right\| + \left\|\bar{\mathbf{x}}_1^* - \bar{\mathbf{x}}_1\right\|^2 - \left\|\bar{\mathbf{x}}_T^* - \bar{\mathbf{x}}_{T+1}\right\|^2\right) - \frac{1}{2\eta}\sum_{t=1}^T\left\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\right\|^2 \\
\le& \frac{1}{2\eta}\left(4\sqrt{R}\sum_{t=1}^T D + R\right) - \frac{1}{2\eta}\sum_{t=1}^T\left\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_{t+1}\right\|^2.
\end{aligned}$$

Combining those bounds of $I_1(t)$, $I_2(t)$ and $I_3(t)$ together, we finally obtain

$$\begin{aligned}
& \mathop{\mathbb{E}}_{\xi}\sum_{t=1}^T\sum_{i=1}^n f_{i,t}(\mathbf{x}_{i,t};\xi) - f_t(\mathbf{x}_t^*;\xi) \\
\le& n\sum_{t=1}^T\left(I_1(t) + I_2(t) + I_3(t)\right) \\
\le& \left(1 + 4\beta\left(\frac{1}{\nu} + 4\right)\right)nT\beta G\eta + \left(L + \frac{2\eta L^2(1-\beta)^2}{\nu} + \frac{\nu}{2\eta} + 10\eta L^2 + \frac{\beta}{2\eta}\right)\frac{nT\eta^2 G}{(1-\rho)^2} \\
& + n\left(\frac{1}{\nu} + 4\right)\left(f(\bar{\mathbf{x}}_1) - f(\bar{\mathbf{x}}_{T+1})\right) + T\eta\sigma^2(1-\beta)\left(\frac{2}{\nu} + 9\right) + \frac{n}{2\eta}\left(4\sqrt{R}D + R\right).
\end{aligned}$$

Let $\nu = \sqrt{\beta^2 + \frac{1}{n}}$, and thus we have

$$\begin{aligned}
& \sum_{t=1}^T\sum_{i=1}^n \mathop{\mathbb{E}}_{\xi} f_{i,t}(\mathbf{x}_{i,t};\xi) - \mathop{\mathbb{E}}_{\xi} f_t(\mathbf{x}_t^*;\xi) \\
\le& \left(1 + 4\beta\left(\frac{1}{\sqrt{\beta^2+\frac{1}{n}}} + 4\right)\right)nT\beta G\eta + \left(L + \frac{2\eta L^2(1-\beta)^2}{\sqrt{\beta^2+\frac{1}{n}}} + \frac{\sqrt{\beta^2+\frac{1}{n}}}{2\eta} + 10\eta L^2 + \frac{\beta}{2\eta}\right)\frac{nT\eta^2 G}{(1-\rho)^2}
\end{aligned}$$

$$
+ n\left(\frac{1}{\sqrt{\beta^2+\frac{1}{n}}}+4\right)(f(\bar{\mathbf{x}}_1)-f(\bar{\mathbf{x}}_{T+1}))+T\eta\sigma^2(1-\beta)\left(\frac{2}{\sqrt{\beta^2+\frac{1}{n}}}+9\right)+\frac{n}{2\eta}\left(4\sqrt{R}D+R\right).
$$

It completes the proof.

□

Lemma 1.

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\xi}\frac{\eta}{2}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+\left(\frac{\eta}{2}-\frac{L\eta^2}{2}\right)\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\\
\le& f(\bar{\mathbf{x}}_t)-f(\bar{\mathbf{x}}_{t+1})+4G\eta\beta^2+\frac{2\eta L^2(1-\beta)^2}{n}\sum_{i=1}^{n}\left\|\bar{\mathbf{x}}_t-\mathbf{x}_{i,t}\right\|^2+\frac{2\eta\sigma^2(1-\beta)^2}{n}.
\end{aligned}
$$

Proof.

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\xi} f(\bar{\mathbf{x}}_{t+1})\\
\le& f(\bar{\mathbf{x}}_t)+\mathop{\mathbb{E}}_{\xi}\langle\nabla f(\bar{\mathbf{x}}_t),\bar{\mathbf{x}}_{t+1}-\bar{\mathbf{x}}_t\rangle+\frac{L}{2}\mathop{\mathbb{E}}_{\xi}\left\|\bar{\mathbf{x}}_{t+1}-\bar{\mathbf{x}}_t\right\|^2\\
=& f(\bar{\mathbf{x}}_t)+\mathop{\mathbb{E}}_{\xi}\left\langle\nabla f(\bar{\mathbf{x}}_t),-\frac{\eta}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\rangle+\frac{L}{2}\mathop{\mathbb{E}}_{\xi}\left\|\frac{\eta}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\\
=& f(\bar{\mathbf{x}}_t)+\frac{\eta}{2}\left(\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2-\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2-\mathop{\mathbb{E}}_{\xi}\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\right)\\
&+\frac{L}{2}\mathop{\mathbb{E}}_{\xi}\left\|\frac{\eta}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\\
=& f(\bar{\mathbf{x}}_t)+\frac{\eta}{2}\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2-\frac{\eta}{2}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2-\left(\frac{\eta}{2}-\frac{L\eta^2}{2}\right)\mathop{\mathbb{E}}_{\xi}\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\\
\le& f(\bar{\mathbf{x}}_t)+\frac{\eta}{2}\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2-\frac{\eta}{2}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2-\left(\frac{\eta}{2}-\frac{L\eta^2}{2}\right)\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t})\right\|^2. \qquad (3)
\end{aligned}
$$

Additionally, we have

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right\|^2\\
=&\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\left(\beta\nabla g_{i,t}(\mathbf{x}_{i,t})+(1-\beta)\nabla f(\mathbf{x}_{i,t};\xi)\right)\right\|^2\\
\le& 2\beta^2\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla g_{i,t}(\mathbf{x}_{i,t})\right\|^2+2(1-\beta)^2\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi)\right\|^2\\
\le& 2\beta^2\left(2\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2+2\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla g_{i,t}(\mathbf{x}_{i,t})\right\|^2\right)+2(1-\beta)^2\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi)\right\|^2\\
\le& 8G\beta^2+2(1-\beta)^2\mathop{\mathbb{E}}_{\xi}\left\|\nabla f(\bar{\mathbf{x}}_t)-\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t})+\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t})-\frac{1}{n}\sum_{i=1}^{n}\nabla f(\mathbf{x}_{i,t};\xi)\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq 8G\beta^2 + 4(1-\beta)^2 \left\| \nabla f(\bar{\mathbf{x}}_t) - \frac{1}{n}\sum_{i=1}^n \nabla f(\mathbf{x}_{i,t}) \right\|^2 + 4(1-\beta)^2 \mathop{\mathbb{E}}_{\xi} \left\| \frac{1}{n}\sum_{i=1}^n \nabla f(\mathbf{x}_{i,t}) - \frac{1}{n}\sum_{i=1}^n \nabla f(\mathbf{x}_{i,t};\xi) \right\|^2 \\
&\leq 8G\beta^2 + \frac{4(1-\beta)^2}{n}\sum_{i=1}^n \left\| \nabla f(\bar{\mathbf{x}}_t) - \nabla f(\mathbf{x}_{i,t}) \right\|^2 + \frac{4(1-\beta)^2}{n^2}\sum_{i=1}^n \mathop{\mathbb{E}}_{\xi} \left\| \nabla f(\mathbf{x}_{i,t}) - \nabla f(\mathbf{x}_{i,t};\xi) \right\|^2 \\
&\leq 8G\beta^2 + \frac{4L^2(1-\beta)^2}{n}\sum_{i=1}^n \left\| \bar{\mathbf{x}}_t - \mathbf{x}_{i,t} \right\|^2 + \frac{4\sigma^2(1-\beta)^2}{n}. \qquad (4)
\end{aligned}
$$

Substituting (4) into (3), we obtain

$$
\begin{aligned}
& f(\bar{\mathbf{x}}_{t+1}) \\
\leq & f(\bar{\mathbf{x}}_t) + \frac{\eta}{2}\left(8G\beta^2 + \frac{4L^2(1-\beta)^2}{n}\sum_{i=1}^n \left\| \bar{\mathbf{x}}_t - \mathbf{x}_{i,t} \right\|^2 + \frac{4\sigma^2(1-\beta)^2}{n} \right) \\
& - \frac{\eta}{2}\left\| \nabla f(\bar{\mathbf{x}}_t) \right\|^2 - \left(\frac{\eta}{2} - \frac{L\eta^2}{2} \right) \left\| \frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t}) \right\|^2.
\end{aligned}
$$

Equivalently, we obtain

$$
\begin{aligned}
& \frac{\eta}{2}\left\| \nabla f(\bar{\mathbf{x}}_t) \right\|^2 + \left(\frac{\eta}{2} - \frac{L\eta^2}{2} \right) \left\| \frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t}) \right\|^2 \\
\leq & f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_{t+1}) + 4G\eta\beta^2 + \frac{2\eta L^2(1-\beta)^2}{n}\sum_{i=1}^n \left\| \bar{\mathbf{x}}_t - \mathbf{x}_{i,t} \right\|^2 + \frac{2\eta\sigma^2(1-\beta)^2}{n}.
\end{aligned}
$$

It completes the proof.

□

Lemma 2.

$$
\frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \left\| \mathbf{x}_{i,t} - \bar{\mathbf{x}}_t \right\|^2 \leq \frac{1}{(1-\rho)^2}\sum_{t=1}^T \eta^2 G.
$$

Proof. Recall that

$$
\mathbf{x}_{i,t+1} = \sum_{j=1}^n \mathbf{W}_{ij}\mathbf{x}_{j,t} - \eta\nabla f_{i,t}(\mathbf{x}_{i,t};\xi),
$$

and

$$
\bar{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_t - \eta\left(\frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t};\xi) \right).
$$

Denote

$$
\begin{aligned}
\mathbf{X}_t = & [\mathbf{x}_{1,t}, \mathbf{x}_{2,t}, ..., \mathbf{x}_{n,t}] \in \mathbb{R}^{d\times n}, \\
\mathbf{G}_t = & [\nabla f_{1,t}(\mathbf{x}_{1,t};\xi), \nabla f_{2,t}(\mathbf{x}_{2,t};\xi), ..., \nabla f_{n,t}(\mathbf{x}_{n,t};\xi)] \in \mathbb{R}^{d\times n}.
\end{aligned}
$$

By letting $\mathbf{x}_{i,1} = \mathbf{0}$ for any $i \in [n]$, the update rule is re-formulated as

$$
\mathbf{X}_{t+1} = \mathbf{X}_t\mathbf{W} - \eta\mathbf{G}_t = -\sum_{s=1}^t \eta\mathbf{G}_s\mathbf{W}^{t-s}.
$$

Similarly, denote $\bar{\mathbf{G}}_t = \frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t};\xi)$, and we have

$$\bar{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_t - \eta\left(\frac{1}{n}\sum_{i=1}^n \nabla f_{i,t}(\mathbf{x}_{i,t};\xi)\right) = -\sum_{s=1}^t \eta\bar{\mathbf{G}}_s. \tag{5}$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^n \left\|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\right\|^2 \\
\overset{①}{=}&\sum_{i=1}^n \left\|\sum_{s=1}^{t-1} \eta\bar{\mathbf{G}}_s - \eta\mathbf{G}_s\mathbf{W}^{t-s-1}\mathbf{e}_i\right\|^2 \\
\overset{②}{=}&\left\|\sum_{s=1}^{t-1} \eta\mathbf{G}_s\mathbf{v}_1\mathbf{v}_1^{\mathrm{T}} - \eta\mathbf{G}_s\mathbf{W}^{t-s-1}\right\|_F^2 \\
\overset{③}{\le}&\left(\eta\rho^{t-s-1}\left\|\sum_{s=1}^{t-1}\mathbf{G}_s\right\|_F\right)^2 \\
\le&\left(\sum_{s=1}^{t-1}\eta\rho^{t-s-1}\left\|\mathbf{G}_s\right\|_F\right)^2.
\end{aligned}$$

① holds due to $\mathbf{e}_i$ is a unit basis vector, whose i-th element is 1 and other elements are 0s. ② holds due to $\mathbf{v}_1 = \frac{\mathbf{1}_n}{\sqrt{n}}$. ③ holds due to Lemma 3.

According to Lemma 4, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n\sum_{t=1}^T \left\|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\right\|^2 \le& \frac{1}{n(1-\rho)^2}\sum_{t=1}^T \eta^2\left\|\mathbf{G}_s\right\|_F^2 \\
\le& \frac{1}{(1-\rho)^2}\sum_{t=1}^T \eta^2 G.
\end{aligned}$$

It completes the proof.

□

Lemma 3 (Appeared in Lemma 5 in [Tang et al., 2018]). *For any matrix $\mathbf{X}_t \in \mathbb{R}^{d\times n}$, decompose the confusion matrix $\mathbf{W}$ as $\mathbf{W} = \sum_{i=1}^n \lambda_i\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathrm{T}}$, where $\mathbf{P} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n] \in \mathbb{R}^{n\times n}$, $\mathbf{v}_i$ is the normalized eigenvector of λ_i. $\mathbf{\Lambda}$ is a diagonal matrix, and λ_i be its i-th element. We have*

$$\left\|\mathbf{X}_t\mathbf{W}^t - \mathbf{X}_t\mathbf{v}_1\mathbf{v}_1^{\mathrm{T}}\right\|_F^2 \le \left\|\rho^t\mathbf{X}_t\right\|_F^2,$$

where $\rho = \max\{|\lambda_2(\mathbf{W})|, |\lambda_n(\mathbf{W})|\}$.

Lemma 4 (Appeared in Lemma 6 in [Tang et al., 2018]). *Given two non-negative sequences $\{a_t\}_{t=1}^\infty$ and $\{b_t\}_{t=1}^\infty$ that satisfying*

$$a_t = \sum_{s=1}^t \rho^{t-s}b_s,$$

with $\rho \in [0, 1)$, we have

$$\sum_{t=1}^k a_t^2 \le \frac{1}{(1-\rho)^2}\sum_{t=1}^k b_s^2.$$

References

H. Tang, S. Gan, C. Zhang, T. Zhang, and J. Liu. Communication Compression for Decentralized Training. *arXiv.org*, Mar. 2018.