A Provably Communication-Efficient Asynchronous Distributed Inference Method for Convex and Nonconvex Problems (Supplementary Material)

Jineng Ren, *Student Member, IEEE,* and Jarvis Haupt, *Senior Member, IEEE*

Appendix C
Proof of Theorem III.1 with More Details

Proof of Theorem III.1. We begin by establishing the first conclusion of the theorem. The idea is to use the cumulative descent of the value of function $\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t)$ to bound the summation of the gaps between $\mathbf{x}^t$ and $\mathbf{x}^{t+1}$. Summing inequality (9) in Lemma III.2 over t yields

$$\begin{aligned}&\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T)-\mathbf{F}(\mathbf{x}^1,\mathbf{x}^0)\\&\quad\le\sum_{t=1}^{T}\Big(\frac{3L}{2}-\frac{\gamma(\rho)}{2}+L\delta\tau\Big)\big\|\Delta^{(t)}\big\|^2\\&\qquad+\sum_{t=1}^{T}\Big(\frac{L\tau}{\delta}-\frac{\rho}{2}\Big)\big\|\Delta^{(t-1)}\big\|^2.\end{aligned}$$

To simplify the right hand side of this inequality, now define

$$c:=\min\left\{\frac{\gamma(\rho)}{2}-\frac{3L}{2}-L\delta\tau,\frac{\rho}{2}-\frac{L\tau}{\delta}\right\},$$

by Assumption III.2 we have $\gamma(\rho)>3L+2L\delta\tau$ and $\rho>\frac{2L\tau}{\delta}$, therefore $c>0$. It holds that

$$\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T)-\mathbf{F}(\mathbf{x}^1,\mathbf{x}^0)\le -c\sum_{t=0}^{T}\big\|\Delta^{(t)}\big\|^2. \tag{42}$$

Note that by Lemma III.3 the LHS of (42) is bounded from below. By letting $T\to\infty$, it follows that

$$\big\|\Delta^{(t)}\big\|\to 0,\qquad t\to\infty.$$

Moreover, Lemma III.3 shows that $\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t)$ is bounded, but due to the coerciveness assumption

$$\lim_{\|\mathbf{x}\|\to\infty}\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x})+h(\mathbf{x})+\frac{\rho}{2}\big\|\mathbf{x}-\mathbf{x}^t\big\|^2=+\infty, \tag{43}$$

so we know $\{\mathbf{x}^{t+1}\}$ is bounded. Therefore the first conclusion is proved.

We now establish the second conclusion of the theorem, which claims that every limit point of the iterates generated by Algorithm 1 is a stationary point. From (3), we know that

$$\begin{aligned}\mathbf{x}^{t+1}=\mathbf{Prox}_h\Big[&\mathbf{x}^{t+1}-\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})\\&-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})+\rho\left(\mathbf{x}^{t+1}-\mathbf{x}^t\right)\Big)\Big],\end{aligned} \tag{44}$$

here we recall $\mathbf{Prox}_h$ is a proximal operator defined by $\mathbf{Prox}_h[\mathbf{z}]:=\operatorname*{argmin}_{\mathbf{x}} h(\mathbf{x})+\frac{1}{2}\|\mathbf{x}-\mathbf{z}\|^2$. Equation (44) further implies that

$$\begin{aligned}&\left\|\mathbf{x}^t-\mathbf{Prox}_h\left(\mathbf{x}^t-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right)\right\|\\&\le\Bigg\|\mathbf{x}^t-\mathbf{x}^{t+1}+\mathbf{x}^{t+1}\\&\quad-\mathbf{Prox}_h\left(\mathbf{x}^t-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right)\Bigg\|\\&\le\left\|\mathbf{x}^t-\mathbf{x}^{t+1}\right\|\\&+\Bigg\|\mathbf{Prox}_h\Bigg[\mathbf{x}^{t+1}-\Bigg(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})\\&-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})+\rho\big(\mathbf{x}^{t+1}-\mathbf{x}^t\big)\Bigg)\Bigg]\\&-\mathbf{Prox}_h\left(\mathbf{x}^t-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right)\Bigg\|\\&\overset{(a)}{\le}\Bigg\|(1+\rho)(\mathbf{x}^{t+1}-\mathbf{x}^t)+\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\\&-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})\big)\Bigg\|\\&+\big\|\Delta^{(t)}\big\|\\&\le(2+\rho)\big\|\Delta^{(t)}\big\|+2L\sum_{k=0}^{\tau}\big\|\Delta^{(t-k)}\big\|,\end{aligned} \tag{45}$$

which gives that $\big\|\mathbf{x}^t-\mathbf{Prox}_h(\mathbf{x}^t-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t))\big\|\to 0$, as $t\to\infty$. Note that here inequality (a) holds because of the nonexpansiveness of the operator $\mathbf{Prox}_h$. The last inequality follows from Assumption III.1.

Let $\mathbf{X}^*$ be the set of stationary points of problem (2), and define

$$\operatorname{dist}\left(\mathbf{x}^t,\mathbf{X}^*\right):=\min_{\widehat{x}\in\mathbf{X}^*}\|\mathbf{x}^t-\widehat{\mathbf{x}}\|$$

as the distance between $\mathbf{x}^t$ and the set $\mathbf{X}^*$. Now we prove that every limit point of $\{\mathbf{x}^t\}$ is a stationary point.

We establish this conclusion by showing that the distance between $\mathbf{X}^*$ and its subsequence converges to 0. Suppose on the contrary there exists a subsequence $\{\mathbf{x}^{t_k}\}$ of $\{\mathbf{x}^t\}$ such that $\mathbf{x}^{t_k} \to \widehat{\mathbf{x}},\ k \to \infty$ but

$$\lim_{k\to\infty} \mathrm{dist}(\mathbf{x}^{t_k}, \mathbf{X}^*) \geq \gamma > 0. \tag{46}$$

Then it is obvious that $\lim_{k\to\infty} \mathrm{dist}(\mathbf{x}^{t_k}, \widehat{\mathbf{x}}) = 0$. Therefore, there exists some $K(\gamma) > 0$, such that

$$\|\mathbf{x}^{t_k} - \widehat{\mathbf{x}}\| \leq \frac{\gamma}{2}, \quad k > K(\gamma). \tag{47}$$

On the other hand, from (45) and the lower semi-continuity of $h(\mathbf{x})$ we have $\widehat{\mathbf{x}} \in \mathbf{X}^*$, so by the definition of the distance function we have

$$\mathrm{dist}(\mathbf{x}^{t_k}, \mathbf{X}^*) \leq \mathrm{dist}(\mathbf{x}^{t_k}, \widehat{\mathbf{x}}). \tag{48}$$

Combining (47) and (48), we must have

$$\mathrm{dist}(\mathbf{x}^{t_k}, \mathbf{X}^*) \leq \frac{\gamma}{2}, \quad k > K(\gamma).$$

This contradicts to (46), so the second result is proved.

We finally prove the third conclusion of the theorem, which gives the sublinear convergence rate of the optimality gap. Summing (45) over t yields

$$\begin{aligned}
&\sum_{t=0}^{T} \left\| \mathbf{x}^t - \mathbf{Prox}_h \left(\mathbf{x}^t - \frac{1}{m} \sum_{j\in[m]} \nabla \mathbf{L}_j(\mathbf{x}^t) \right) \right\|^2 \\
&\leq \sum_{t=0}^{T} 2(2+\rho)^2 \big\|\Delta^{(t)}\big\|^2 + 2(2L)^2 \sum_{t=0}^{T} \sum_{k=0}^{\tau} \big\|\Delta^{(t-k)}\big\|^2 \\
&\leq \big(2(2+\rho)^2 + 8L^2\tau\big) \sum_{t=0}^{T} \big\|\Delta^{(t)}\big\|^2.
\end{aligned} \tag{49}$$

Combining (42) and (49) we can bound the optimality gap by the cumulative descent of the Lyapunov function:

$$\sum_{t=0}^{T} \left\| \widetilde{\nabla} \mathbf{L}(\mathbf{x}^t) \right\|^2 \leq \frac{\mu}{c} \left(\mathbf{F}(\mathbf{x}^1, \mathbf{x}^0) - \mathbf{F}(\mathbf{x}^{T+1}, \mathbf{x}^T) \right),$$

where $\mu := \big(2(2+\rho)^2 + 8L^2\tau\big)$.

Let $T(\epsilon) := \min\left\{ t \mid \left\| \widetilde{\nabla} \mathbf{L}(\mathbf{x}^t) \right\|^2 \leq \epsilon, t \geq 0 \right\}$. Then the above inequality implies

$$T(\epsilon)\epsilon \leq \frac{\mu}{c} \left(\mathbf{F}(\mathbf{x}^1, \mathbf{x}^0) - \mathbf{F}(\mathbf{x}^{T+1}, \mathbf{x}^T) \right).$$

Thus from Lemma III.3 it follows the third conclusion of Theorem III.1 that

$$\epsilon \leq \frac{C \cdot \big(\mathbf{F}(\mathbf{x}^1, \mathbf{x}^0) - \underline{\mathbf{F}}\big)}{T(\epsilon)},$$

where $C := \frac{\mu}{c} > 0$. Therefore the conclusions of Theorem III.1 are proved. □

Appendix D
Proof of Theorem III.2 with More Details

Proof of Theorem III.2. We begin by defining $\widetilde{\Delta}^{(t+1)} = \mathbf{F}(\mathbf{x}^{t+1}, \mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*, \mathbf{x}^*)$, the optimality gap in terms of the Lyapunov function. Note that $\mathbf{F}(\mathbf{x}, \mathbf{x}^t) = \frac{1}{m} \sum_{j\in[m]} \mathbf{L}_j(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^t\|^2 + h(\mathbf{x}) \geq \mathbf{F}(\mathbf{x}^*, \mathbf{x}^*) = \frac{1}{m} \sum_{j\in[m]} \mathbf{L}_j(\mathbf{x}^*) + h(\mathbf{x}^*)$. Therefore $\widetilde{\Delta}^{(t+1)}$ is nonnegative. Moreover, from Lemma III.2 it follows that

$$\begin{aligned}
\widetilde{\Delta}^{(t+1)} &\leq \widetilde{\Delta}^{(t)} + \Big(\frac{3L}{2} - \frac{\rho}{2} + L\delta\tau\Big)\big\|\Delta^{(t)}\big\|^2 \\
&\quad -\frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2 + \frac{L}{\delta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2.
\end{aligned} \tag{50}$$

Note that from (13) of Lemma III.4 we have

$$\begin{aligned}
\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\widetilde{\Delta}^{(t+1)} &\leq \frac{\delta_1 L + \frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\big\|\Delta^{(t)}\big\|^2 \\
&\quad + \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2 \sum_{j\in[m]}\|\mathbf{x}^{t_j} - \mathbf{x}^t\|^2.
\end{aligned} \tag{51}$$

Based on these two inequalities, the conclusion can be proved. Specifically, by summing (51) and (50), one has

$$\begin{aligned}
&\Big(1 + \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\Big)\widetilde{\Delta}^{(t+1)} \\
&\leq \widetilde{\Delta}^{(t)} + \Big[\frac{\delta_1 L + \frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1} + \frac{3L}{2} - \frac{\rho}{2} + L\delta\tau\Big]\big\|\Delta^{(t)}\big\|^2 \\
&- \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2 + \frac{L}{\delta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2 \\
&+ \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2 \sum_{j\in[m]}\|\mathbf{x}^{t_j} - \mathbf{x}^t\|^2.
\end{aligned} \tag{52}$$

Inequality (52) gives us an relation between $\widetilde{\Delta}^{(t+1)}$ and $\widetilde{\Delta}^{(t)}$. Let us define $\eta := 1 + \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}$ and

$$(P3) := \frac{\delta_1 L + \frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1} + \frac{3L}{2} - \frac{\rho}{2} + L\delta\tau.$$

Then by applying (52) recursively we have

$$\begin{aligned}
&\widetilde{\Delta}^{(t+1)} \\
&\leq \frac{1}{\eta}\widetilde{\Delta}^{(t)} + \frac{1}{\eta}(P3)\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2\eta}\big\|\Delta^{(t-1)}\big\|^2 \\
&+ \frac{L}{\delta\eta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2 + \frac{1}{\eta}\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1 L^2}{2m}\sum_{j\in[m]}\|\mathbf{x}^{t_j} - \mathbf{x}^t\|^2 \\
&\leq \frac{1}{\eta^2}\widetilde{\Delta}^{(t-1)} + \frac{1}{\eta}\left(\frac{1}{\eta}(P3)\big\|\Delta^{(t-1)}\big\|^2 - \frac{\rho}{2\eta}\big\|\Delta^{(t-2)}\big\|^2\right) \\
&+ \frac{1}{\eta}(P3)\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2\eta}\big\|\Delta^{(t-1)}\big\|^2 \\
&+ \left(\frac{L}{\delta\eta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2 + \frac{L}{\delta\eta^2}\sum_{k=1}^{\tau}\big\|\Delta^{(t-1-k)}\big\|^2\right)
\end{aligned}$$

$$+\frac{1}{\eta}\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2\tau\sum_{l=0}^{1}\frac{1}{\eta^{l+1}}\sum_{j\in[m]}\sum_{k=1}^{\tau}\left\|\Delta^{(t-l-k)}\right\|^2$$
$$\cdots$$
$$\leq\frac{1}{\eta^t}\widetilde{\Delta}^{(1)}+\frac{1}{\eta}(P3)\left\|\Delta^{(t)}\right\|^2+\Big(\frac{1}{\eta^2}(P3)-\frac{\rho}{2\eta}\Big)\left\|\Delta^{(t-1)}\right\|^2$$
$$+\big(\frac{1}{\eta^3}(P3)-\frac{\rho}{2\eta^2}\big)\left\|\Delta^{(t-2)}\right\|^2+\cdots$$
$$+\big(\frac{1}{\eta^{t+1}}(P3)-\frac{\rho}{2\eta^t}\big)\left\|\Delta^{(0)}\right\|^2+\Big(\frac{L}{\delta\eta}\sum_{l=0}^{t}\frac{1}{\eta^l}\sum_{k=1}^{\tau}\left\|\Delta^{(t-l-k)}\right\|^2\Big)$$
$$+\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2\tau\sum_{l=0}^{t}\frac{1}{\eta^{l+1}}\sum_{j\in[m]}\sum_{k=1}^{\tau}\left\|\Delta^{(t-l-k)}\right\|^2$$
$$\leq\frac{1}{\eta^t}\widetilde{\Delta}^{(1)}+\frac{1}{\eta}(P3)\left\|\Delta^{(t)}\right\|^2+\left(\frac{1}{\eta^2}(P3)-\frac{\rho}{2\eta}\right)\left\|\Delta^{(t-1)}\right\|^2$$
$$+\left(\frac{1}{\eta^3}(P3)-\frac{\rho}{2\eta^2}\right)\left\|\Delta^{(t-2)}\right\|^2+\cdots$$
$$+\left(\frac{1}{\eta^{t+1}}(P3)-\frac{\rho}{2\eta^t}\right)\left\|\Delta^{(0)}\right\|^2$$
$$+\left(\frac{L}{\delta}+\frac{\frac{\delta_1}{2}L^2\tau}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\right)\frac{\eta^\tau-1}{\eta-1}\sum_{l=1}^{t}\frac{1}{\eta^{l+1}}\left\|\Delta^{(t-l)}\right\|^2, \quad (53)$$

where the last inequality is resulted from combining the coefficients of the same items; moreover, by the definition of $\Delta^{(l)}$, we let $\Delta^{(l)}=0$ for $l<0$, and use the following fact that

$$\sum_{l=0}^{t}\frac{1}{\eta^{l+1}}\sum_{k=1}^{\tau}\left\|\Delta^{(t-l-k)}\right\|^2 = \eta^{-t-1}\sum_{l=0}^{t}\frac{\eta^t}{\eta^l}\sum_{k=1}^{\tau}\left\|\Delta^{(t-l-k)}\right\|^2 = \eta^{-t-1}\sum_{j=0}^{t}\eta^j\sum_{k=j-\tau}^{j-1}\left\|\Delta^{(k)}\right\|^2$$

which further implies that

$$\sum_{l=0}^{t}\frac{1}{\eta^{l+1}}\sum_{k=1}^{\tau}\left\|\Delta^{(t-l-k)}\right\|^2 \overset{(h)}{\leq}\eta^{-t-1}\sum_{j=0}^{t-1}\eta^j(1+\eta+\cdots+\eta^{\tau-1})\left\|\Delta^{(j)}\right\|^2 \leq\frac{\eta^\tau-1}{\eta-1}\sum_{j=0}^{t-1}\frac{1}{\eta^{t-j+1}}\left\|\Delta^{(j)}\right\|^2 \leq\frac{\eta^\tau-1}{\eta-1}\sum_{l=1}^{t}\frac{1}{\eta^{l+1}}\left\|\Delta^{(t-l)}\right\|^2,$$

where inequality (h) holds because the coefficient of $\left\|\Delta^{(j)}\right\|^2$ in the summation is less than $\eta^j(1+\eta+\cdots+\eta^{\tau-1})$. Therefore if $\rho>0$ satisfies that

$$(P3):=\frac{\delta_1 L+\frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1}+\frac{3L}{2}-\frac{\rho}{2}+L\delta\tau<0, \quad (54)$$

and

$$(P3)-\frac{\rho\eta}{2}+\Big(\frac{L}{\delta}+\frac{\frac{\delta_1}{2}L^2\tau}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\Big)\frac{\eta^\tau-1}{\eta-1} =\frac{\delta_1 L+\frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1}+\frac{3L}{2}-\frac{\rho}{2}+L\delta\tau -\frac{\rho\eta}{2}+\Big(\frac{L}{\delta}+\frac{\frac{\delta_1}{2}L^2\tau}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\Big)\frac{\eta^\tau-1}{\eta-1}<0, \quad (55)$$

which are the coefficients of the nonnegative terms in (53), then inputting (54) and (55) into (53) we have

$$0\leq\widetilde{\Delta}^{(t+1)}\leq\frac{1}{\eta^t}\widetilde{\Delta}^{(1)}.$$

The conclusion is proved. □

Appendix E
Proofs of Lemmata and Additional Results

A. Proof of Lemma III.1 with More Details

The conclusion can be proved by the optimality of $\mathbf{x}^{t+1}$ and the convexity assumption. Firstly, using the optimality of $\mathbf{x}^{t+1}$ in the update (3), we have

$$-\Big[\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j=1}^{m}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})\Big] \in\partial\Big[h(\mathbf{x})+\tfrac{\rho}{2}\left\|\mathbf{x}-\mathbf{x}^t\right\|^2\Big]. \quad (56)$$

Recall that in Assumption III.2 (I), we define the convex modulus of $\frac{\rho}{2}\|\mathbf{x}-\mathbf{x}^t\|^2+h(\mathbf{x})$ by $\gamma(\rho)$. It follows that

$$\tfrac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2+h(\mathbf{x}^{t+1})-\big(\tfrac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^t\|^2+h(\mathbf{x}^t)\big) \leq\Big\langle-\Big[\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j=1}^{m}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})\Big], \mathbf{x}^{t+1}-\mathbf{x}^t\Big\rangle-\frac{\gamma(\rho)}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2 =-\Big\langle\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j=1}^{m}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1}), \Delta^{(t)}\Big\rangle-\frac{\gamma(\rho)}{2}\left\|\Delta^{(t)}\right\|^2,$$

where we define $\Delta^{(t)}:=\mathbf{x}^{t+1}-\mathbf{x}^t$. Therefore

$$\frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2+h(\mathbf{x}^{t+1})-\frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^{t-1}\|^2-h(\mathbf{x}^t) =\frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2+h(\mathbf{x}^{t+1})-\frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^t\|^2-h(\mathbf{x}^t) +\frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^t\|^2+h(\mathbf{x}^t)-\frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^{t-1}\|^2-h(\mathbf{x}^t) \leq-\Big\langle\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j=1}^{m}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1}), \Delta^{(t)}\Big\rangle-\frac{\gamma(\rho)}{2}\left\|\Delta^{(t)}\right\|^2-\frac{\rho}{2}\left\|\Delta^{(t-1)}\right\|^2, \quad (57)$$

proving Lemma III.1.

B. Proof of Lemma III.2 with More Details

The proof of this lemma relies on the Lipschitz continuity of $\nabla \mathbf{L}_j(\mathbf{x})$ and Lemma III.1. It follows from Assumption III.1 that $\nabla \mathbf{L}_j(\mathbf{x})$ is Lipschitz continuous with constant L. Therefore we have

$$\begin{aligned}
&\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1}) - \frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t})\\
&\leq \left\langle \mathbf{x}^{t+1}-\mathbf{x}^t, \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right\rangle + \frac{L}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2\\
&= \left\langle \Delta^{(t)}, \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right\rangle + \frac{L}{2}\big\|\Delta^{(t)}\big\|^2. \qquad (58)
\end{aligned}$$

By the above definition of function $\mathbf{F}$, combining (57) and (58) results in

$$\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^{t},\mathbf{x}^{t-1})\\
&\overset{(b)}{\leq} -\Big\langle \nabla\mathbf{L}_1(\mathbf{x}^{t+1}) + \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) - \nabla\mathbf{L}_1(\mathbf{x}^{t_1}),\\
&\qquad \Delta^{(t)}\Big\rangle - \frac{\gamma(\rho)}{2}\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2\\
&+ \left\langle \Delta^{(t)}, \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right\rangle + \frac{L}{2}\big\|\Delta^{(t)}\big\|^2\\
&= -\Big\langle \nabla\mathbf{L}_1(\mathbf{x}^{t+1}) + \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t}) - \nabla\mathbf{L}_1(\mathbf{x}^{t}),\\
&\qquad \Delta^{(t)}\Big\rangle - \frac{\gamma(\rho)}{2}\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2\\
&+ \left\langle \Delta^{(t)}, \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right\rangle + \frac{L}{2}\big\|\Delta^{(t)}\big\|^2\\
&+ \left\langle \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}), \Delta^{(t)}\right\rangle\\
&+ \left\langle \nabla\mathbf{L}_1(\mathbf{x}^{t_1}) - \nabla\mathbf{L}_1(\mathbf{x}^{t}), \Delta^{(t)}\right\rangle\\
&\leq \left(\frac{3L}{2} - \frac{\gamma(\rho)}{2}\right)\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2\\
&\underbrace{+ \left\langle \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}), \Delta^{(t)}\right\rangle}_{(P1)}\\
&\underbrace{+ \left\langle \nabla\mathbf{L}_1(\mathbf{x}^{t_1}) - \nabla\mathbf{L}_1(\mathbf{x}^{t}), \Delta^{(t)}\right\rangle}_{(P1)}, \qquad (59)
\end{aligned}$$

where inequality (b) is due to Lemma III.1 and Assumption III.1. Note that

$$\begin{aligned}
&\nabla\mathbf{L}_1(\mathbf{x}^{t_1}) - \nabla\mathbf{L}_1(\mathbf{x}^{t})\\
&\qquad = \sum_{k=1}^{t-t_1}\left(\nabla\mathbf{L}_1(\mathbf{x}^{t-k}) - \nabla\mathbf{L}_1(\mathbf{x}^{t-k+1})\right),
\end{aligned}$$

which implies

$$\begin{aligned}
&\|\nabla\mathbf{L}_1(\mathbf{x}^{t_1}) - \nabla\mathbf{L}_1(\mathbf{x}^{t})\|\\
&\qquad \leq \sum_{k=1}^{t-t_1}\big\|\nabla\mathbf{L}_1(\mathbf{x}^{t-k}) - \nabla\mathbf{L}_1(\mathbf{x}^{t-k+1})\big\|\\
&\qquad \leq \sum_{k=1}^{t-t_1} L\,\big\|\mathbf{x}^{t-k} - \mathbf{x}^{t-k+1}\big\|\\
&\qquad \leq \sum_{k=1}^{\tau} L\,\big\|\mathbf{x}^{t-k} - \mathbf{x}^{t-k+1}\big\|\\
&\qquad = \sum_{k=1}^{\tau} L\big\|\Delta^{(t-k)}\big\|.
\end{aligned}$$

Similarly, we can see

$$\begin{aligned}
&\Big\|\tfrac{1}{m}\textstyle\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) - \tfrac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})\Big\|\\
&\qquad\qquad \leq \sum_{k=1}^{\tau} L\big\|\Delta^{(t-k)}\big\|.
\end{aligned}$$

These two inequalities result in

$$\begin{aligned}
(P1) &\leq 2L\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|\big\|\Delta^{(t)}\big\|\\
&\leq L\sum_{k=1}^{\tau}\left(\frac{1}{\delta}\big\|\Delta^{(t-k)}\big\|^2 + \delta\big\|\Delta^{(t)}\big\|^2\right)\\
&\leq \frac{L}{\delta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2 + L\delta\tau\big\|\Delta^{(t)}\big\|^2, \qquad (60)
\end{aligned}$$

where $(P1)$ is defined in (59) and in the second inequality we apply the fact that

$$a\cdot b \leq \frac{1}{2}\left(\frac{1}{\delta}a^2 + \delta b^2\right)$$

for any $a, b, \delta > 0$. By inserting (60) into (59) we have

$$\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^{t},\mathbf{x}^{t-1})\\
&\quad \leq \left(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau\right)\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2\\
&\qquad\qquad + \frac{L}{\delta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2,
\end{aligned}$$

proving the conclusion of Lemma III.2.

C. Proof of Lemma III.3 with More Details

In this lemma we prove the boundedness of $\mathbf{F}$. First of all, summing the above inequality (9) of Lemma III.2 over t yields

$$\begin{aligned}
&\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T) - \mathbf{F}(\mathbf{x}^{1},\mathbf{x}^{0})\\
&\qquad \leq \sum_{t=1}^{T}\left(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau\right)\big\|\Delta^{(t)}\big\|^2\\
&\qquad\qquad + \sum_{t=1}^{T}\left(\frac{L\tau}{\delta} - \frac{\rho}{2}\right)\big\|\Delta^{(t-1)}\big\|^2.
\end{aligned}$$

If ρ satisfies Assumption III.2, then it holds that

$$\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T) - \mathbf{F}(\mathbf{x}^{1},\mathbf{x}^{0}) < 0.$$

By taking $\mathbf{x}^T$ as the initial point, similarly we have

$$\mathbf{F}(\mathbf{x}^{2T+1},\mathbf{x}^{2T})-\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^{T})<0.$$

Continuing this process we get a decreasing subsequence $\left\{\mathbf{F}(\mathbf{x}^{kT+1},\mathbf{x}^{kT})\right\}_{k=0,1,\cdots}$. Therefore there exists a constant $\bar{F}_0$ such that

$$\mathbf{F}(\mathbf{x}^{kT+1},\mathbf{x}^{kT})<\bar{F}_0. \tag{61}$$

When starting with $\mathbf{x}^1,\cdots,\mathbf{x}^{T-1}$, with similar analysis we can prove that there exists constants $\bar{F}_1,\cdots,\bar{F}_{T-1}$ such that

$$\mathbf{F}(\mathbf{x}^{kT+l+1},\mathbf{x}^{kT+l})<\bar{F}_l, \tag{62}$$

for $l=1,2,\cdots,T-1$. Define $\bar{\mathbf{F}}:=\max\left\{\bar{F}_0,\cdots,\bar{F}_{T-1}\right\}$, then

$$\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^{t})<\bar{\mathbf{F}}<+\infty, \quad \forall t\in\mathbb{N}.$$

On the other hand, let $\underline{\mathbf{F}}:=\underline{L}$, then by the definition of $\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t)$ and Assumption III.2, we have

$$\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^{t})\\
&=\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1})+\frac{\rho}{2}\left\|\mathbf{x}^{t+1}-\mathbf{x}^t\right\|^2+h(\mathbf{x}^{t+1})\\
&\geq\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1})+h(\mathbf{x}^{t+1})\\
&=\mathbf{L}(\mathbf{x}^{t+1})>\underline{L}=\underline{\mathbf{F}}>-\infty,
\end{aligned}$$

for any $t\in\mathbb{N}$. Therefore the boundedness of function $\mathbf{F}$ in Lemma III.3 is proved.

D. Proof of Lemma III.4 with More Details

Lemma III.4 gives a bound of $\left(\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t)-\mathbf{F}(\mathbf{x}^*,\mathbf{x}^*)\right)$ with respect to the distance between $\mathbf{x}^t$ and $\mathbf{x}^{t+1}$ as well as that between $\mathbf{x}^t$ and $\mathbf{x}^{t_j}$. Note that $\left(\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t)-\mathbf{F}(\mathbf{x}^*,\mathbf{x}^*)\right)$ is nonnegative. The proof of this lemma relies on the strong convexity in Assumption III.3. Firstly, by the optimality of $\mathbf{x}^{t+1}$ in the update (3), we have

$$\begin{aligned}
&\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})\\
&+\partial h(\mathbf{x}^{t+1})+\rho(\mathbf{x}^{t+1}-\mathbf{x}^t)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x})\leq 0,
\end{aligned} \tag{63}$$

for all $\mathbf{x}\in\mathbb{R}^p$. Letting $\mathbf{x}=\mathbf{x}^*$ implies

$$\begin{aligned}
&\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})+\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})-\nabla\mathbf{L}_1(\mathbf{x}^{t_1})\\
&+\partial h(\mathbf{x}^{t+1})+\rho(\mathbf{x}^{t+1}-\mathbf{x}^t)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\leq 0.
\end{aligned} \tag{64}$$

Note that (64) further implies that

$$\begin{aligned}
&\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})-\nabla\mathbf{L}_1(\mathbf{x}^t)+\big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})\\
&-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\big)+\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\\
&+\partial h(\mathbf{x}^{t+1})+\rho(\mathbf{x}^{t+1}-\mathbf{x}^t)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\leq 0.
\end{aligned} \tag{65}$$

By the strong convexity of $\mathbf{L}_j$ one has

$$\begin{aligned}
\mathbf{L}_j(\mathbf{y})\geq\mathbf{L}_j(\mathbf{x})+\left(\nabla\mathbf{L}_j(\mathbf{x})\right)^\top(\mathbf{y}-\mathbf{x})+\frac{\sigma^2}{2}\|\mathbf{y}-\mathbf{x}\|^2,\\
\forall\mathbf{x},\ \mathbf{y}\in\mathbb{R}^p.
\end{aligned} \tag{66}$$

Setting $\mathbf{y}=\mathbf{x}^*$, $\mathbf{x}=\mathbf{x}^{t+1}$ in (66) we have

$$\begin{aligned}
\mathbf{L}_j(\mathbf{x}^*)\geq\mathbf{L}_j(\mathbf{x}^{t+1})&+\left(\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\right)^\top(\mathbf{x}^*-\mathbf{x}^{t+1})\\
&+\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2,
\end{aligned}$$

which further implies that

$$\begin{aligned}
&\left(\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\right)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\
&\quad\geq\mathbf{L}_j(\mathbf{x}^{t+1})-\mathbf{L}_j(\mathbf{x}^*)+\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2.
\end{aligned} \tag{67}$$

Summing (67) over $j\in[m]$ gives that

$$\begin{aligned}
\frac{1}{m}\sum_{j\in[m]}\left(\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\right)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\geq\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1})\\
-\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^*)+\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2.
\end{aligned} \tag{68}$$

Putting (68) into the above inequality (65) that resulted from the optimality, we have

$$\begin{aligned}
&\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})-\nabla\mathbf{L}_1(\mathbf{x}^t)+\Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})\\
&-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\
&+\Big(\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1})-\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^*)\Big)\\
&+\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2+\partial h(\mathbf{x}^{t+1})(\mathbf{x}^{t+1}-\mathbf{x}^*)\\
&+\rho(\mathbf{x}^{t+1}-\mathbf{x}^t)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\leq 0.
\end{aligned} \tag{69}$$

Since $h(\mathbf{x})$ is convex, we have

$$h(\mathbf{x}^{t+1})-h(\mathbf{x}^*)\leq\partial h(\mathbf{x}^{t+1})(\mathbf{x}^{t+1}-\mathbf{x}^*).$$

Putting it into (69), we have

$$\begin{aligned}
&\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1})-\nabla\mathbf{L}_1(\mathbf{x}^t)+\Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j})\\
&-\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\
&+\Big(\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1})-\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^*)\Big)\\
&+\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2+h(\mathbf{x}^{t+1})-h(\mathbf{x}^*)\\
&+\rho(\mathbf{x}^{t+1}-\mathbf{x}^t)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\leq 0.
\end{aligned} \tag{70}$$

Note that

$$\rho(\mathbf{x}^{t+1}-\mathbf{x}^t)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*) = \frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 + \frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2 - \frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^*\|^2. \quad (71)$$

Then putting (71) into (70), one obtains

$$\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^t) + \Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*) + \Big(\frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^{t+1}) - \frac{1}{m}\sum_{j\in[m]}\mathbf{L}_j(\mathbf{x}^*)\Big) + \frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2 + h(\mathbf{x}^{t+1}) - h(\mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2 + \frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 - \frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^*\|^2 \le 0,$$

which is equivalent to

$$\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*,\mathbf{x}^*) \le -\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2 - \Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^t) + \Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*) - \frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 + \frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^*\|^2. \quad (72)$$

Now, note that

$$\begin{aligned}&\Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^t) + \Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\ &= \Big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^t) + \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\ &\quad \underbrace{+\Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\Big)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)}_{(P5)}.\end{aligned} \quad (73)$$

Given a matrix $\mathbf{H}\in\mathbb{R}^{s\times s}$, we first define $\|\mathbf{v}\|_{\mathbf{H}}^2 := \mathbf{v}^\top\mathbf{H}\mathbf{v}$ for any vector $\mathbf{v}\in\mathbb{R}^s$. Then by using the Mean Value Theorem we have

$$\begin{aligned}&\left(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^t) + \Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\right)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\ &= (\mathbf{x}^{t+1}-\mathbf{x}^t)^\top\Big[\nabla^2\mathbf{L}_1(\xi) - \frac{1}{m}\sum_{j\in[m]}\nabla^2\mathbf{L}_j(\xi)\Big]\cdot(\mathbf{x}^{t+1}-\mathbf{x}^*)\\ &= \frac{1}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2 - \frac{1}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^{t+1}\|_\Sigma^2 + \frac{1}{2}\|\mathbf{x}^t-\mathbf{x}^{t+1}\|_\Sigma^2 - \frac{1}{2}\|\mathbf{x}^t-\mathbf{x}^*\|_\Sigma^2,\end{aligned}$$

where $\Sigma := \nabla^2\mathbf{L}_1(\xi) - \frac{1}{m}\sum_{j\in[m]}\nabla^2\mathbf{L}_j(\xi)$. It follows that

$$\begin{aligned}&\left(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^t) + \Big(\frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t+1})\Big)\right)^\top(\mathbf{x}^{t+1}-\mathbf{x}^*)\\ &= \frac{1}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2 + \frac{1}{2}\|\mathbf{x}^t-\mathbf{x}^{t+1}\|_\Sigma^2 - \frac{1}{2}\|\mathbf{x}^t-\mathbf{x}^{t+1}+\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2\\ &\ge \frac{1}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2 + \frac{1}{2}\|\mathbf{x}^t-\mathbf{x}^{t+1}\|_\Sigma^2 - \frac{1}{2}(1+\delta_1)\|\mathbf{x}^{t+1}-\mathbf{x}^t\|_\Sigma^2 - \frac{1}{2}(1+1/\delta_1)\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2\\ &\ge -\frac{1}{2}\delta_1\|\mathbf{x}^t-\mathbf{x}^{t+1}\|_\Sigma^2 - \frac{1}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2,\end{aligned} \quad (74)$$

for $\delta_1 > 0$. Note that

$$\frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^*\|^2 = \frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^{t+1}+\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 \le \frac{\rho}{2}(1+\delta_1)\|\mathbf{x}^t-\mathbf{x}^{t+1}\|^2 + \frac{\rho}{2}(1+1/\delta_1)\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2,$$

which implies that

$$-\frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 + \frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^*\|^2 \le \frac{\rho}{2}(1+\delta_1)\|\mathbf{x}^t-\mathbf{x}^{t+1}\|^2 + \frac{\rho}{2}\frac{1}{\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2. \quad (75)$$

Putting (73), (74), and (75) into (72), we can bound the optimality gap of the function $\mathbf{F}$ by

$$\begin{aligned}&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*,\mathbf{x}^*)\\ &\le -\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2 + \frac{1}{2}\delta_1\|\mathbf{x}^{t+1}-\mathbf{x}^t\|_\Sigma^2 + \frac{1}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2 - \frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 + \frac{\rho}{2}\|\mathbf{x}^t-\mathbf{x}^*\|^2 - (P5)\\ &= -\frac{\sigma^2}{2}\|\mathbf{x}^*-\mathbf{x}^{t+1}\|^2 + \frac{1}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|_\Sigma^2 + \frac{\rho}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 + \frac{1}{2}\delta_1\|\mathbf{x}^{t+1}-\mathbf{x}^t\|_\Sigma^2 + \frac{\rho}{2}(1+\delta_1)\|\mathbf{x}^t-\mathbf{x}^{t+1}\|^2 - (P5).\end{aligned} \quad (76)$$

Now we bound (P5) on the RHS of (76), which is defined in (73). For some $\delta_1 > 0$ one has

$$
\begin{aligned}
&(P5)\\
&\geq -\frac{\delta_1}{2m}\sum_{j\in[m]}\|\nabla \mathbf{L}_j(\mathbf{x}^{t_j}) - \nabla \mathbf{L}_j(\mathbf{x}^t)\|^2 - \frac{1}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2\\
&\geq -\frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j}-\mathbf{x}^t\|^2 - \frac{1}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2.
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*,\mathbf{x}^*)\\
&\leq -\frac{\sigma^2}{2}\|\mathbf{x}^* - \mathbf{x}^{t+1}\|^2 + \frac{L}{\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2\\
&+ \frac{\rho}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2 + \delta_1 L\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2\\
&+ \frac{\rho}{2}(1+\delta_1)\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2\\
&+ \frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j}-\mathbf{x}^t\|^2 + \frac{1}{2\delta_1}\|\mathbf{x}^{t+1}-\mathbf{x}^*\|^2. \qquad (77)
\end{aligned}
$$

Let $\frac{\sigma^2}{2} > \frac{L}{\delta_1} + \frac{\rho}{2\delta_1} + \frac{1}{2\delta_1}$, i.e., $\sigma^2 > \frac{2L}{\delta_1} + \frac{\rho}{\delta_1} + \frac{1}{\delta_1}$, then it follows that

$$
\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*,\mathbf{x}^*) \leq \delta_1 L\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2\\
&+ \big(-\frac{\sigma^2}{2} + \frac{L}{\delta_1} + \frac{\rho}{2\delta_1} + \frac{1}{2\delta_1}\big)\|\mathbf{x}^* - \mathbf{x}^{t+1}\|^2\\
&+ \frac{\rho}{2}(1+\delta_1)\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 + \frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j}-\mathbf{x}^t\|^2\\
&\leq \delta_1 L\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2 + \frac{\rho}{2}(1+\delta_1)\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2\\
&+ \frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j}-\mathbf{x}^t\|^2, \qquad (78)
\end{aligned}
$$

from which we have

$$
\begin{aligned}
&\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\big(\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*,\mathbf{x}^*)\big)\\
&\leq \frac{\delta_1 L}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2\\
&+ \frac{\frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2\\
&+ \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j}-\mathbf{x}^t\|^2,
\end{aligned}
$$

therefore proving Lemma III.4.

E. Proof of Corollary III.1 and Corollary III.2

We first prove Corollary III.1. Following the proof steps in Lemma III.1, we have

$$
\begin{aligned}
&\frac{\rho}{2}\|\mathbf{x}^{t+1}-\mathbf{x}^t\|^2 + h(\mathbf{x}^{t+1}) - \frac{\rho}{2}\|\mathbf{x}^t - \mathbf{x}^{t-1}\|^2 - h(\mathbf{x}^t)\\
&\leq -\Big\langle \nabla \mathbf{L}_1(\mathbf{x}^{t+1}) + \frac{1}{m}\sum_{j=1}^m \nabla \mathbf{L}_j(\mathbf{x}^{t_j}) - \nabla \mathbf{L}_1(\mathbf{x}^{t_1}) - \epsilon^t,\\
&\Delta^{(t)}\Big\rangle - \frac{\gamma(\rho)}{2}\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2.
\end{aligned}
$$

In the second step, for the descent of function $\mathbf{F}$, similar to Lemma III.2 for any $\delta > 0$ it holds that

$$
\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1},\mathbf{x}^t) - \mathbf{F}(\mathbf{x}^t,\mathbf{x}^{t-1})\\
&\leq \left(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau\right)\big\|\Delta^{(t)}\big\|^2 - \frac{\rho}{2}\big\|\Delta^{(t-1)}\big\|^2\\
&\qquad + \frac{L}{\delta}\sum_{k=1}^{\tau}\big\|\Delta^{(t-k)}\big\|^2 + \langle \epsilon^t, \Delta^{(t)}\rangle. \qquad (79)
\end{aligned}
$$

Now following the similar path, the first conclusion of Theorem III.1 can be proved. Summing up (79) over t yields

$$
\begin{aligned}
&\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T) - \mathbf{F}(\mathbf{x}^1,\mathbf{x}^0)\\
&= \mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T) - \mathbf{F}(\mathbf{x}^T,\mathbf{x}^{T-1}) + \mathbf{F}(\mathbf{x}^T,\mathbf{x}^{T-1})\\
&- \mathbf{F}(\mathbf{x}^{T-1},\mathbf{x}^{T-2}) + \cdots\cdots + \mathbf{F}(\mathbf{x}^2,\mathbf{x}^1) - \mathbf{F}(\mathbf{x}^1,\mathbf{x}^0)\\
&\leq \sum_{t=1}^T\Big(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau\Big)\big\|\Delta^{(t)}\big\|^2\\
&+ \sum_{t=1}^T\Big(\frac{L\tau}{\delta} - \frac{\rho}{2}\Big)\big\|\Delta^{(t-1)}\big\|^2 + \sum_{t=1}^T\langle \epsilon^t, \Delta^{(t)}\rangle,
\end{aligned}
$$

which further implies that

$$
\begin{aligned}
&\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T) - \mathbf{F}(\mathbf{x}^1,\mathbf{x}^0)\\
&\leq \sum_{t=1}^T\left(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau\right)\big\|\Delta^{(t)}\big\|^2\\
&+ \sum_{t=1}^T\left(\frac{L\tau}{\delta} - \frac{\rho}{2}\right)\big\|\Delta^{(t-1)}\big\|^2 + \frac{1}{2}\sum_{t=1}^T(\|\epsilon^t\|^2 + \big\|\Delta^{(t)}\big\|^2)\\
&\leq \sum_{t=1}^T\left(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau + \frac{1}{2}\right)\big\|\Delta^{(t)}\big\|^2\\
&+ \sum_{t=1}^T\left(\frac{L\tau}{\delta} - \frac{\rho}{2} + \frac{1}{2}c_1\right)\big\|\Delta^{(t-1)}\big\|^2,
\end{aligned}
$$

where in the last inequality we use Assumption III.4. This inequality bounds the cumulative progress of function $\mathbf{F}$ by the weighted summation of the norm of $\Delta^{(t)}$ for $t = 1, \ldots, T$.

To simplify the above inequality, we define $\widetilde{c} := \min\big\{\frac{\gamma(\rho)}{2} - \frac{3L}{2} - L\delta\tau - \frac{1}{2}, \frac{\rho}{2} - \frac{L\tau}{\delta} - \frac{1}{2}c_1\big\}$. Assume that

$$
\gamma(\rho) > 3L + 2L\delta\tau + 1 \quad \text{and} \quad \rho > \frac{2L\tau}{\delta} + c_1, \qquad (80)
$$

then we have $\widetilde{c} > 0$. Therefore

$$
\mathbf{F}(\mathbf{x}^{T+1},\mathbf{x}^T) - \mathbf{F}(\mathbf{x}^1,\mathbf{x}^0) \leq -\widetilde{c}\sum_{t=0}^T\big\|\Delta^{(t)}\big\|^2. \qquad (81)
$$

Note that by Lemma III.3 the LHS of (81) is bounded from below. It follows that

$$
\big\|\Delta^{(t)}\big\| \to 0, \qquad t \to \infty.
$$

We now establish the second conclusion of Theorem III.1. From (15) we know that

$$
\begin{aligned}
\mathbf{x}^{t+1} = \mathbf{Prox}_h\Big[&\mathbf{x}^{t+1} - \Big(\nabla \mathbf{L}_1(\mathbf{x}^{t+1}) + \frac{1}{m}\sum_{j\in[m]}\nabla \mathbf{L}_j(\mathbf{x}^{t_j})\\
&-\nabla \mathbf{L}_1(\mathbf{x}^{t_1}) + \rho\left(\mathbf{x}^{t+1} - \mathbf{x}^t\right) + \epsilon^t\Big)\Big].
\end{aligned}
$$

This equation implies that

$$\begin{aligned}
&\left\|\mathbf{x}^t - \mathbf{Prox}_h\left(\mathbf{x}^t - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right)\right\| \\
&\leq \Bigg\|\mathbf{x}^t - \mathbf{x}^{t+1} + \mathbf{x}^{t+1} \\
&- \mathbf{Prox}_h\left(\mathbf{x}^t - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right)\Bigg\| \\
&\leq \left\|\mathbf{x}^t - \mathbf{x}^{t+1}\right\| \\
&+ \Bigg\|\mathbf{Prox}_h\Bigg[\mathbf{x}^{t+1} - \Bigg(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) + \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) \\
&- \nabla\mathbf{L}_1(\mathbf{x}^{t_1}) + \rho\big(\mathbf{x}^{t+1} - \mathbf{x}^t\big) - \epsilon^t\Bigg)\Bigg] \\
&- \mathbf{Prox}_h\left(\mathbf{x}^t - \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t)\right)\Bigg\| \\
&\overset{(\widetilde{a})}{\leq} \Bigg\|(1+\rho)(\mathbf{x}^{t+1} - \mathbf{x}^t) - \epsilon^t + \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^t) \\
&- \frac{1}{m}\sum_{j\in[m]}\nabla\mathbf{L}_j(\mathbf{x}^{t_j}) - \big(\nabla\mathbf{L}_1(\mathbf{x}^{t+1}) - \nabla\mathbf{L}_1(\mathbf{x}^{t_1})\big)\Bigg\| \\
&+ \left\|\Delta^{(t)}\right\| \\
&\leq (2+\rho)\left\|\Delta^{(t)}\right\| + 2L\sum_{k=0}^{\tau}\left\|\Delta^{(t-k)}\right\| + c_1^{\frac{1}{2}}\left\|\Delta^{(t-1)}\right\| \\
&\longrightarrow 0, \qquad t\to\infty.
\end{aligned} \tag{82}$$

Note that here inequality $(\widetilde{a})$ holds because of the nonexpansiveness of the operator $\mathbf{Prox}_h$. The last inequality follows from Assumption III.1. Therefore as in the proof of Theorem III.1, the second result holds.

The rest of the analysis is the same as that of Theorem III.1. Specifically, we can bound the norm of the proximal gradient by the cumulative progress of function $\mathbf{F}$:

$$\sum_{t=0}^{T}\left\|\widetilde{\nabla}\mathbf{L}\left(\mathbf{x}^t\right)\right\|^2 \leq \frac{\widetilde{\mu}}{\widetilde{c}}\left(\mathbf{F}(\mathbf{x}^1, \mathbf{x}^0) - \mathbf{F}(\mathbf{x}^{T+1}, \mathbf{x}^T)\right),$$

where $\widetilde{\mu} := 3\left((2+\rho)^2 + 4L^2\tau + c_1\right)$.

Recall that $T(\epsilon) := \min\left\{t \mid \left\|\widetilde{\nabla}\mathbf{L}(\mathbf{x}^t)\right\| \leq \epsilon, t\geq 0\right\}$. Thus it follows that

$$\epsilon \leq \frac{C\cdot\left(\mathbf{F}(\mathbf{x}^1, \mathbf{x}^0) - \underline{\mathbf{F}}\right)}{T(\epsilon)},$$

where $C := \frac{\widetilde{\mu}}{\widetilde{c}} > 0$. The conclusions in Corollary III.1 have been proved.

Now we prove Corollary III.2. Similar to Lemma III.4, the optimality gap of function $\mathbf{F}$ can be bounded as the following:

$$\begin{aligned}
&\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\left(\mathbf{F}(\mathbf{x}^{t+1}, \mathbf{x}^t) - \mathbf{F}(\mathbf{x}^*, \mathbf{x}^*)\right) \\
&\leq \frac{\delta_1 L}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 \\
&+ \frac{\frac{\rho}{2}(1+\delta_1)}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 \\
&+ \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j} - \mathbf{x}^t\|^2 \\
&+ \big(\frac{1}{2}\|\epsilon^t\|^2 + \frac{1}{2}\left\|\Delta^{(t)}\right\|^2\big)\frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}.
\end{aligned} \tag{83}$$

Following the steps of Lemma III.2, we have

$$\begin{aligned}
&\mathbf{F}(\mathbf{x}^{t+1}, \mathbf{x}^t) - \mathbf{F}(\mathbf{x}^t, \mathbf{x}^{t-1}) \\
&\leq \left(\frac{3L}{2} - \frac{\gamma(\rho)}{2} + L\delta\tau\right)\left\|\Delta^{(t)}\right\|^2 - \frac{\rho}{2}\left\|\Delta^{(t-1)}\right\|^2 \\
&\qquad + \frac{L}{\delta}\sum_{k=1}^{\tau}\left\|\Delta^{(t-k)}\right\|^2 + \langle\epsilon^t, \Delta^{(t)}\rangle.
\end{aligned} \tag{84}$$

Summing (83) and (84) and then applying the Assumption III.4 leads to

$$\begin{aligned}
&\left(1 + \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\right)\widetilde{\Delta}^{(t+1)} \\
&\leq \widetilde{\Delta}^{(t)} - \frac{\rho - c_1}{2}\left\|\Delta^{(t-1)}\right\|^2 + \frac{L}{\delta}\sum_{k=1}^{\tau}\left\|\Delta^{(t-k)}\right\|^2 \\
&+ \Big[\frac{\delta_1 L + \frac{\rho}{2}(1+\delta_1) + \frac{1}{2}}{\frac{\rho}{2}(1+\delta_1)+\delta_1} + \frac{3L}{2} - \frac{\rho-1}{2} + L\delta\tau\Big]\left\|\Delta^{(t)}\right\|^2 \\
&+ \frac{1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\frac{\delta_1}{2m}L^2\sum_{j\in[m]}\|\mathbf{x}^{t_j} - \mathbf{x}^t\|^2 \\
&+ \frac{c_1}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\left\|\Delta^{(t-1)}\right\|^2.
\end{aligned}$$

By a recursive argument similar to the proof of Theorem III.2, one can prove that if $\rho > 0$ satisfies that

$$\frac{\delta_1 L + \frac{\rho}{2}(1+\delta_1) + \frac{1}{2}}{\frac{\rho}{2}(1+\delta_1)+\delta_1} + \frac{3L}{2} - \frac{\rho-1}{2} + L\delta\tau < 0 \tag{85}$$

and

$$\begin{aligned}
&\frac{\delta_1 L + \frac{\rho}{2}(1+\delta_1) + \frac{1}{2}}{\frac{\rho}{2}(1+\delta_1)+\delta_1} + \frac{3L}{2} - \frac{\rho-1}{2} + L\delta\tau - \frac{(\rho-c_1)\eta}{2} \\
&+ \left(\frac{L}{\delta} + \frac{\frac{\delta_1}{2}L^2\tau}{\frac{\rho}{2}(1+\delta_1)+\delta_1}\right)\frac{\eta^\tau - 1}{\eta - 1} + \frac{c_1\eta}{\frac{\rho}{2}(1+\delta_1)+\delta_1} < 0,
\end{aligned} \tag{86}$$

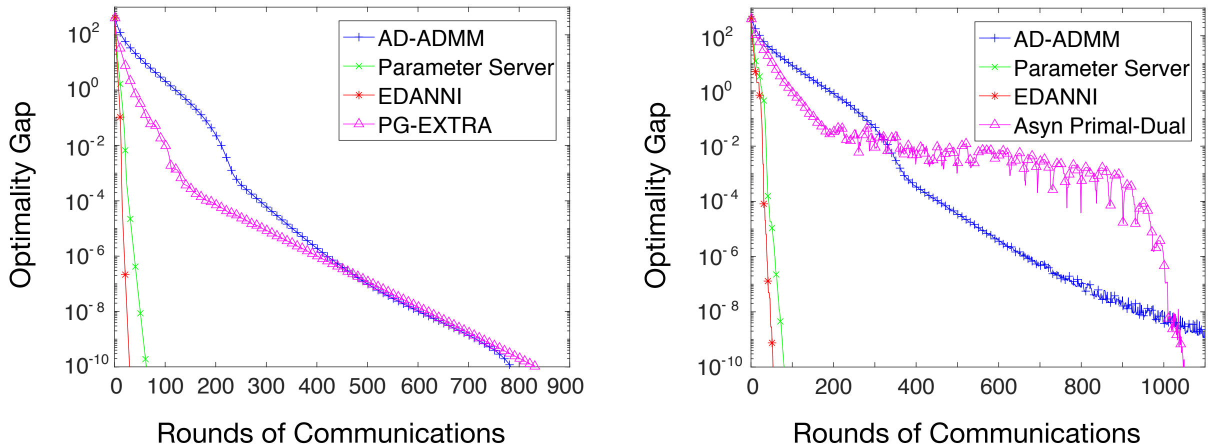
then we have

$$0 \leq \widetilde{\Delta}^{(t+1)} \leq \frac{1}{\eta^t}\widetilde{\Delta}^{(1)},$$

the corollary is proved.

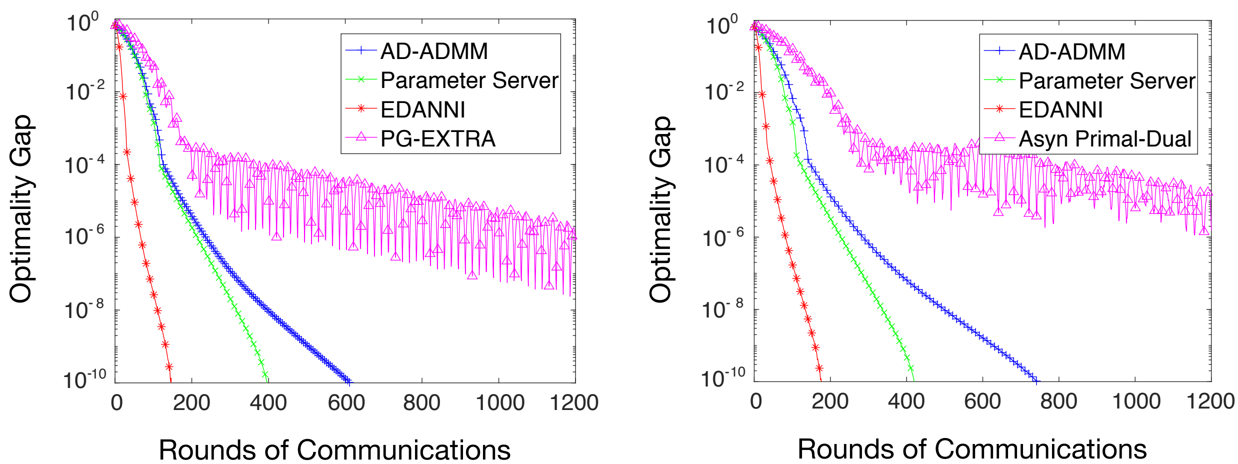
F. Additional Experiment Results

This section shows additional results for comparing the competing algorithms (including the fully decentralized ones) on the LASSO problems (Figure 12) and the sparse PCA problems (Figure 13). Moreover, the convergence and time table of the proposed algorithm for different choices of the delay bound τ when solving sparse PCA problems are shown in Figure 14.



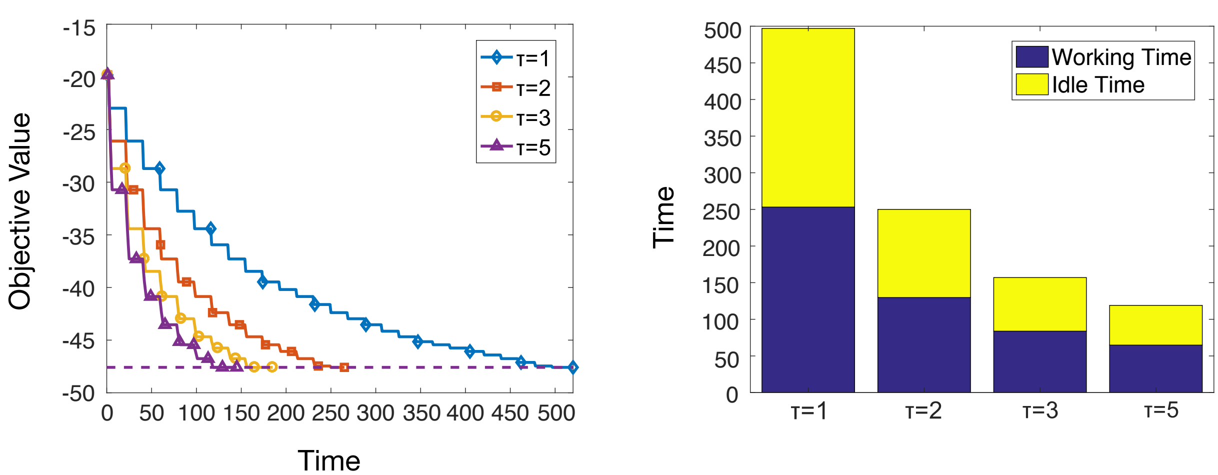
(a) Synchronous, (26, 2.3, **1**, 27.7) (b) Asynchronous ($\tau = 8$), (18.6, 1.7, **1**, 23.2)

Fig. 12. Comparison of candidate algorithms in LASSO when $m = 20$, $n = 10000$, $p = 800$, $s = 20$, $\rho = 12$, and $\theta = 0.01$.



(a) Synchronous, (4.2, 2.7, **1**, >20) (b) Asynchronous ($\tau = 8$), (4.3, 2.4, **1**, >20)

Fig. 13. Comparison of candidate algorithms in sparse PCA when $m = 16$, $n = 8000$, $p = 50$, $q = 100$, $s = 120$, and $\theta = 0.1$.



(a) Convergence for different τ's (b) Time table for different τ's

Fig. 14. Comparison of EDANNI in Sparse PCA with different delay bounds when $m = 8$, $n = 1000$, $p = 10000$, $q = 50$, $s = 2000$, $\theta = 0.01$.