Supplementary Materials

Proof of Theorem 1

To help theoretical analysis, we denote the objective functions in (6) and (7) as

$$g_m(x_m) = \lambda R_m(x_m) + \langle y^t, \mathcal{D}_m x_m \rangle + \frac{\rho}{2} \Big\| \sum_{\substack{k=1\\k\neq m}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m - z^t \Big\|^2,$$

$$h(z) = l(z) - \langle y^t, z \rangle + \frac{\rho}{2} \Big\| \sum_{m=1}^{M} \mathcal{D}_m x_m^{t+1} - z \Big\|^2, \tag{18}$$

correspondingly. We prove the following four lemmas to help prove the theorem.

Lemma 2 *Under Assumption 1, we have*

$$\nabla l(z^{t+1}) = y^{t+1},$$

and

$$\|y^{t+1} - y^t\|^2 \leq L^2 \|z^{t+1} - z^t\|^2.$$

Proof. By the optimality in (7), we have

$$\nabla l(z^{t+1}) - y^t + \rho\big(z^{t+1} - \sum_{m=1}^{M} \mathcal{D}_m x_m^{t+1}\big) = 0.$$

Combined with (8), we can get

$$\nabla l(z^{t+1}) = y^{t+1}. \tag{19}$$

Combined with Assumption 1.1, we have

$$\|y^{t+1} - y^t\|^2 = \|\nabla l(z^{t+1}) - \nabla l(z^t)\|^2 \leq L^2 \|z^{t+1} - z^t\|^2. \tag{20}$$

□

Lemma 3 *We have*

$$\begin{aligned}
&\big(\| \sum_{m=1}^{M} x_m^{t+1} - z\|^2 - \| \sum_{m=1}^{M} x_m^t - z\|^2\big) - \sum_{m=1}^{M} \big(\| \sum_{\substack{k=1\\k\neq m}}^{M} x_k^t + x_m^{t+1} - z\|^2 - \| \sum_{m=1}^{M} x_m^t - z\|^2\big) \\
&\leq \sum_{m=1}^{M} \|x_m^{t+1} - x_m^t\|^2.
\end{aligned} \tag{21}$$

Proof.

$$\begin{aligned}
\text{LHS} =& \big(\sum_{m=1}^{M} (x_m^{t+1} + x_m^t) - 2z\big)\top\big(\sum_{m=1}^{M} x_m^{t+1} - \sum_{m=1}^{M} x_m^t\big) - \sum_{m=1}^{M} \big(\sum_{\substack{k=1\\k\neq m}}^{M} 2x_k^t + x_m^t + x_m^{t+1} - 2z\big)\top(x_m^{t+1} - x_m^t) \\
=& - \sum_{m=1}^{M} \sum_{\substack{k=1\\k\neq m}}^{M} (x_k^{t+1} - x_k^t)\top(x_m^{t+1} - x_m^t) \\
=& - \| \sum_{m=1}^{M} (x_m^{t+1} - x_m^t)\|^2 + \sum_{m=1}^{M} \|x_m^{t+1} - x_m^t\|^2 \\
\leq& \sum_{m=1}^{M} \|x_m^{t+1} - x_m^t\|^2.
\end{aligned}$$

□

Lemma 4 *Suppose Assumption 1 holds. We have*

$$\begin{aligned}
&\mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^{t+1}) - \mathcal{L}(\{x_m^t\}, z^t; y^t) \\
&\leq \sum_{m=1}^{M} -\left(\frac{\gamma_m(\rho)}{2} - \sigma_{max}(\mathcal{D}_m^\top \mathcal{D}_m)\right) \|x_m^{t+1} - x_m^t\|^2 - \left(\frac{\gamma(\rho)}{2} - \frac{L^2}{\rho}\right) \|z^{t+1} - z^t\|^2.
\end{aligned}$$

Proof. The LFH can be decomposed into two parts as

$$\begin{aligned}
&\mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^{t+1}) - \mathcal{L}(\{x_m^t\}, z^t; y^t) \\
=&\big(\mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^{t+1}) - \mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^t)\big) \\
&+ \big(\mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^t) - \mathcal{L}(\{x_m^t\}, z^t; y^t)\big).
\end{aligned} \tag{22}$$

For the first term, we have

$$\begin{aligned}
&\mathcal{L}(\{x_j^{t+1}\}, z^{t+1}; y^{t+1}) - \mathcal{L}(\{x_j^{t+1}\}, z^{t+1}; y^t) \\
=&\langle y^{t+1} - y^t, \sum_j D_j x_j^{t+1} - z^{t+1}\rangle \\
=&\frac{1}{\rho}\|y^{t+1} - y^t\|^2 \quad \text{(by (8))} \\
=&\frac{L^2}{\rho}\|z^{t+1} - z^t\|^2 \quad \text{(by Lemma 2).}
\end{aligned} \tag{23}$$

For the second term, we have

$$\begin{aligned}
&\mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^t) - \mathcal{L}(\{x_m^t\}, z^t; y^t) \\
=&\mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^t) - \mathcal{L}(\{x_m^t\}, z^{t+1}; y^t) + \mathcal{L}(\{x_m^t\}, z^{t+1}; y^t) - \mathcal{L}(\{x_m^t\}, z^t; y^t) \\
\leq&\Bigg(\big(\lambda \sum_{m=1}^M R_m(x_m^{t+1}) + \langle y^t, \sum_{m=1}^M \mathcal{D}_m x_m^{t+1}\rangle + \frac{\rho}{2}\big\|\sum_{k=1}^M \mathcal{D}_k x_k^{t+1} - z^{t+1}\big\|^2\big) \\
&- \big(\lambda \sum_{m=1}^M R_m(x_m^t) + \langle y^t, \sum_{m=1}^M \mathcal{D}_m x_m^t\rangle + \frac{\rho}{2}\big\|\sum_{k=1}^M \mathcal{D}_k x_k^t - z^{t+1}\big\|^2\big)\Bigg) \\
&+ \Bigg(\big(l(z^{t+1}) - \langle y^t, z^{t+1}\rangle + \frac{\rho}{2}\big\|\sum_{m=1}^M \mathcal{D}_m x_m^{t+1} - z^{t+1}\big\|^2\big) - \big(l(z^t) - \langle y^t, z^t\rangle + \frac{\rho}{2}\big\|\sum_{m=1}^M \mathcal{D}_m x_m^{t+1} - z^t\big\|^2\big)\Bigg) \\
\leq&\sum_{m=1}^M \Bigg(\big(\lambda R_m(x_m^{t+1}) + \langle y^t, \mathcal{D}_m x_m^{t+1}\rangle + \frac{\rho}{2}\big\|\sum_{\substack{k=1\\k\neq m}}^M \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^{t+1}\big\|^2\big) \\
&- \big(\lambda R_m(x_m^t) + \langle y^t, \mathcal{D}_m x_m^t\rangle + \frac{\rho}{2}\big\|\sum_{k=1}^M \mathcal{D}_k x_k^t - z^{t+1}\big\|^2\big)\Bigg) + \sum_{m=1}^M \|\mathcal{D}_m(x_m^{t+1} - x_m^t)\|^2 \\
&+ \Bigg(\big(l(z^{t+1}) - \langle y^t, z^{t+1}\rangle + \frac{\rho}{2}\big\|\sum_{m=1}^M \mathcal{D}_m x_m^{t+1} - z^{t+1}\big\|^2\big) \\
&- \big(l(z^t) - \langle y^t, z^t\rangle + \frac{\rho}{2}\big\|\sum_{m=1}^M \mathcal{D}_m x_m^{t+1} - z^t\big\|^2\big)\Bigg) \quad \text{(by Lemma 3)} \\
=&\sum_{m=1}^M \big(g_m(x_m^{t+1}) - g_m(x_m^t)\big) + (h(z^{t+1}) - h(z^T))) + \sum_{m=1}^M \|\mathcal{D}_m(x_m^{t+1} - x_m^t)\|^2 \\
\leq&\sum_{m=1}^M \big(\langle \nabla g_m(x_m^{t+1}), x_m^{t+1} - x_m^t\rangle - \frac{\gamma_m(\rho)}{2}\|x_m^{t+1} - x_m^t\|^2\big) + \langle \nabla h(z^{t+1}), z^{t+1} - z^t\rangle - \frac{\gamma(\rho)}{2}\|z^{t+1} - z^t\|^2 \\
&+ \sum_{m=1}^M \|\mathcal{D}_m(x_m^{t+1} - x_m^t)\|^2 \quad \text{(by strongly convexity from Assumption 1.2)}
\end{aligned}$$

$$\leq -\sum_{m=1}^{M} \frac{\gamma_m(\rho)}{2}\|x_m^{t+1}-x_m^t\|^2 - \frac{\gamma(\rho)}{2}\|z^{t+1}-z^t\|^2 + \sum_{m=1}^{M}\|\mathcal{D}_m(x_m^{t+1}-x_m^t)\|^2$$

(by optimality condition for subproblem in (6) and (7))

$$\leq \sum_{m=1}^{M} -\left(\frac{\gamma_m(\rho)}{2} - \sigma_{\max}(\mathcal{D}_m^\top\mathcal{D}_m)\right)\|x_m^{t+1}-x_m^t\|^2 - \frac{\gamma(\rho)}{2}\|z^{t+1}-z^t\|^2. \tag{24}$$

Note that we have abused the notation $\nabla g_m(x_m)$ and denote it as the subgradient when g is non-smooth but convex. Combining (22), (23) and (24), the lemma is proved. □

Lemma 5 *Suppose Assumption 1 holds. Then the following limit exists and is bounded from below:*

$$\lim_{t\to\infty} \mathcal{L}(\{x^{t+1}\}, z^{t+1}; y^{t+1}). \tag{25}$$

Proof.

$$\begin{aligned}
&\mathcal{L}(\{x^{t+1}\}, z^{t+1}; y^{t+1})\\
=&l(z^{t+1}) + \lambda\sum_{m=1}^{M} R_m(x_m^{t+1}) + \langle y^{t+1}, \sum_{m=1}^{M}\mathcal{D}_m x_m^{t+1} - z^{t+1}\rangle + \frac{\rho}{2}\|\mathcal{D}_m x_m^{t+1} - z^{t+1}\|^2\\
=&\lambda\sum_{m=1}^{M} R_m(x_m^{t+1}) + l(z^{t+1}) + \langle \nabla l(z^{t+1}), \sum_{m=1}^{M}\mathcal{D}_m x_m^{t+1} - z^{t+1}\rangle + \frac{\rho}{2}\|\mathcal{D}_m x_m^{t+1} - z^{t+1}\|^2 \quad \text{(by Lemma 2)}\\
\geq&\lambda\sum_{m=1}^{M} R_m(x_m^{t+1}) + l(\sum_{m=1}^{M}\mathcal{D}_m x_m^{t+1}) + \frac{\rho-L}{2}\|\mathcal{D}_m x_m^{t+1} - z^{t+1}\|^2.
\end{aligned} \tag{26}$$

Combined with Assumption 1.3, $\mathcal{L}(\{x^{t+1}\}, z^{t+1}; y^{t+1})$ is lower bounded. Furthermore, by Assumption 1.2 and Lemma 4, $\mathcal{L}(\{x^{t+1}\}, z^{t+1}; y^{t+1})$ is decreasing. These complete the proof. □

Now we are ready for the proof.

Part 1. By Assumption 1.2, Lemma 4 and Lemma 5, we have

$$\begin{aligned}
&\|x_m^{t+1} - x_m^t\| \to 0 \quad \forall m = 1, 2, \ldots, M,\\
&\|z^{t+1} - z^t\| \to 0.
\end{aligned}$$

Combined with Lemma 2, we have

$$\|y^{t+1} - y^t\|^2 \to 0.$$

Combined with (8), we complete the proof for Part 1.

Part 2. Due to the fact that $\|y^{t+1} - y^t\|^2 \to 0$, by taking limit in (8), we can get (12).

At each iteration $t+1$, by the optimality of the subproblem in (7), we have

$$\nabla l(z^{t+1}) - y^t + \rho(z^{t+1} - \sum_{m=1}^{M}\mathcal{D}_m x_m^{t+1}) = 0. \tag{27}$$

Combined with (12) and taking the limit, we get (11).

Similarly, by the optimality of the subproblem in (6), for $\forall m \in \mathcal{M}$ there exists $\eta_m^{t+1} \in \partial R_m(x_m^{t+1})$, such that

$$\left\langle x_m - x_m^{t+1}, \lambda\eta_m^{t+1} + \mathcal{D}_m^T y^t + \rho\mathcal{D}_m^T\big(\sum_{\substack{k=1\\k\leq m}}^{M}\mathcal{D}_k x_k^{t+1} + \sum_{\substack{k=1\\k>j}}^{M}\mathcal{D}_k x_k^t - z^t\big)\right\rangle \geq 0 \quad \forall x_m \in X_m. \tag{28}$$

Since R_m is convex, we have

$$\lambda R_m(x_m) - \lambda R_m(x_m^{t+1}) + \left\langle x - x_m^{t+1}, \mathcal{D}_m^\top y^t + \rho\big(\sum_{\substack{k=1\\k\leq m}}^{M}\mathcal{D}_k x_k^{t+1} + \sum_{\substack{k=1\\k>j}}^{M}\mathcal{D}_k x_k^t - z^t\big)^T\mathcal{D}_j\right\rangle \geq 0 \quad \forall x_m \in X_m. \tag{29}$$

Combined with (12) and the fact $\|x_m^{t+1} - x_m^t\| \to 0$, by taking the limit, we get

$$\lambda R_m(x_m) - \lambda R_m(x_m^*) + \left\langle x - x_m^*, \mathcal{D}_m^\top y^*\right\rangle \geq 0 \quad \forall x_m \in X_m, \forall m, \tag{30}$$

which is equivalent to

$$\lambda R_m(x) + \langle y^*, \mathcal{D}_m x\rangle - \lambda R_m(x_m^*) - \langle y^*, \mathcal{D}_m x_m^*\rangle \geq 0 \quad \forall x \in X_m, \forall m. \tag{31}$$

And we can get the result in (9).
When $m \notin \mathcal{M}$, we have

$$\Big\langle x_m - x_m^{t+1}, \lambda \nabla R_m(x_m^{t+1}) + \mathcal{D}_m^\top y^t + \rho\big(\sum_{\substack{k=1\\k\leq m}}^{M} \mathcal{D}_k x_k^{t+1} + \sum_{\substack{k=1\\k>j}}^{M} \mathcal{D}_k x_k^t - z^t\big)^T \mathcal{D}_m\Big\rangle \geq 0 \quad \forall x_m \in X_m. \tag{32}$$

Taking the limit and we can get (10).

Part 3. We first show that there exists a limit point for each of the sequences $\{x_m^t\}$, $\{z^t\}$ and $\{y^t\}$. Since $X_m, \forall m$ is compact, $\{x_m^t\}$ must have a limit point. With Theorem 1.1, we can get that $\{z^t\}$ is also compact and has a limit point. Furthermore, with Lemma 2, we can get $\{y^t\}$ is also compact and has a limit point.

We prove Part 3 by contradiction. Since $\{x_m^t\}$, $\{z^t\}$ and $\{y^t\}$ lie in some compact set, there exists a subsequence $\{x_m^{t_k}\}$, $\{z^{t_k}\}$ and $\{y^{t_k}\}$, such that

$$(\{x_m^{t_k}\}, z^{t_k}; y^{t_k}) \to (\{\hat{x}_m\}, \hat{z}; \hat{y}), \tag{33}$$

where $(\{\hat{x}_m\}, \hat{z}; \hat{y})$ is some limit point and by part 2, we have $(\{\hat{x}_m\}, \hat{z}; \hat{y}) \in Z^*$. Suppose that $\{\{x_m^t\}, z^t; y^t\}$ does not converge to Z^*, since $(\{x_m^{t_k}\}, z^{t_k}; y^{t_k})$ is a subsequence of it, there exists some $\gamma > 0$, such that

$$\lim_{k\to\infty} \quad \mathrm{dist}\big((\{x_m^{t_k}\}, z^{t_k}; y^{t_k}); Z^*\big) = \gamma > 0. \tag{34}$$

From (33), there exists some $J(\gamma) > 0$, such that

$$\|(\{x_m^{t_k}\}, z^{t_k}; y^{t_k}) - (\{\hat{x}_m\}, \hat{z}; \hat{y})\| \leq \frac{\gamma}{2}, \quad \forall k \geq J(\gamma). \tag{35}$$

Since $(\{\hat{x}_m\}, \hat{z}; \hat{y}) \in Z^*$, we have

$$\mathrm{dist}\big((\{x_m^{t_k}\}, z^{t_k}; y^{t_k}); Z^*\big) \leq \mathrm{dist}\big((\{x_m^{t_k}\}, z^{t_k}; y^{t_k}); (\{\hat{x}_m\}, \hat{z}; \hat{y})\big). \tag{36}$$

From the above two inequalities, we must have

$$\mathrm{dist}\big((\{x_m^{t_k}\}, z^{t_k}; y^{t_k}); Z^*\big) \leq \frac{\gamma}{2}, \quad \forall k \geq J(\gamma), \tag{37}$$

which contradicts to (34), completing the proof. $\square$

Proof of Theorem 2

We first show an upper bound for V^t.

1. Bound for $\tilde{\nabla}_{x_m}\mathcal{L}(\{x_m^t\}, z^t; y^t)$. When $m \in \mathcal{M}$, from the optimality condition in (6), we have

$$0 \in \lambda \partial_{x_m} R_j(x_m^{t+1}) + \mathcal{D}^\top y^t + \rho \mathcal{D}_j^\top\big(\sum_{\substack{k=1\\k\neq j}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^t\big).$$

By some rearrangement, we have

$$\big(x_m^{t+1} - \mathcal{D}^\top y^t - \rho \mathcal{D}_m^\top\big(\sum_{\substack{k=1\\k\neq j}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^t\big)\big) - x_m^{t+1} \in \lambda \partial_{x_m} R_m(x_m^{t+1}),$$

which is equivalent to

$$x_m^{t+1} = \mathrm{prox}_{\lambda R_m}\big[x_m^{t+1} - \mathcal{D}^\top y^t - \rho \mathcal{D}_m^\top\big(\sum_{\substack{k=1\\k\neq j}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^t\big)\big]. \tag{38}$$

Therefore,

$$
\begin{aligned}
&\|x_m^t - \mathrm{prox}_{\lambda R_m}\big[x_m^t - \nabla_{x_m}\big(\mathcal{L}(\{x_m^t\}, z^t; y^t) - \lambda \sum_{m=1}^{M} R_m(x_m^t)\big)\big]\| \\
=&\big\|x_m^t - x_m^{t+1} + x_m^{t+1} - \mathrm{prox}_{\lambda R_m}\big[x_m^t - \mathcal{D}^\top y^t - \rho \mathcal{D}_m^\top\big(\sum_{k=1}^{M} \mathcal{D}_k x_k^t - z^t\big)\big]\big\| \\
\leq&\|x_m^t - x_m^{t+1}\| + \big\|\mathrm{prox}_{\lambda R_m}\big[x_m^{t+1} - \mathcal{D}^\top y^t - \rho \mathcal{D}_m^T\big(\sum_{\substack{k=1\\k\neq m}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^t\big)\big] \\
&- \mathrm{prox}_{\lambda R_m}\big[x_m^t - \mathcal{D}^\top y^t - \rho \mathcal{D}_m^T\big(\sum_{k=1}^{M} \mathcal{D}_k x_k^t - z^t\big)\big]\big\| \\
&\leq 2\|x_m^t - x_m^{t+1}\| + \rho\|\mathcal{D}_m^\top \mathcal{D}_m(x_m^{t+1} - x_m^t)\|.
\end{aligned} \tag{39}
$$

When $m \notin \mathcal{M}$, similarly, we have

$$
\lambda \nabla_{x_m} R_m(x_m^{t+1}) + \mathcal{D}^\top y^t + \rho \mathcal{D}_m^\top\big(\sum_{\substack{k=1\\k\neq j}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^t\big) = 0. \tag{40}
$$

Therefore,

$$
\begin{aligned}
&\|\nabla_{x_m}\mathcal{L}(\{x_m^t\}, z^t; y^t)\| \\
=&\|\lambda \nabla_{x_m} R_m(x_m^t) + \mathcal{D}^\top y^t + \rho \mathcal{D}_m^\top\big(\sum_{k=1}^{M} \mathcal{D}_k x_k^t - z^t\big)\| \\
=&\|\lambda \nabla_{x_m} R_m(x_m^t) + \mathcal{D}^\top y^t + \rho \mathcal{D}_m^T\big(\sum_{k=1}^{M} \mathcal{D}_k x_k^t - z^t\big) \\
&- \big(\lambda \nabla_{x_m} R_m(x_m^{t+1}) + \mathcal{D}^\top y^t + \rho \mathcal{D}_m^T\big(\sum_{\substack{k=1\\k\neq j}}^{M} \mathcal{D}_k x_k^t + \mathcal{D}_m x_m^{t+1} - z^t\big)\big)\| \\
\leq&\lambda\|\nabla_{x_m} R_m(x_m^t) - \nabla_{x_m} R_m(x_m^{t+1})\| + \rho\|\mathcal{D}_m^\top \mathcal{D}_m(x_m^{t+1} - x_m^t)\| \\
\leq& L_m\|x_m^{t+1} - x_m^t\| + \rho\|\mathcal{D}_m^\top \mathcal{D}_m(x_m^{t+1} - x_m^t)\|. \quad \text{(by Assumption 1.4)}
\end{aligned} \tag{41}
$$

2. Bound for $\|\nabla_z \mathcal{L}(\{x_m^t\}, z^t; y^t)\|$. By optimality condition in (7), we have

$$
\nabla l(z^{t+1}) - y^t + \rho\big(z^{t+1} - \sum_{m=1}^{M} \mathcal{D}_m x_m^{t+1}\big) = 0.
$$

Therefore

$$
\begin{aligned}
&\|\nabla_z \mathcal{L}(\{x_m^t\}, z^t; y^t)\| \\
=&\|l(z^t) - y^t + \rho\big(z^t - \sum_{m=1}^{M} \mathcal{D}_m x_m^t\big)\| \\
=&\|l(z^t) - y^t + \rho\big(z^t - \sum_{m=1}^{M} \mathcal{D}_m x_m^t\big) - \big(l(z^{t+1}) - y^t + \rho\big(z^{t+1} - \sum_{m=1}^{M} \mathcal{D}_m x_m^{t+1}\big)\big)\| \\
\leq&(L + \rho)\|z^{t+1} - z^t\| + \rho \sum_{m=1}^{M} \|\mathcal{D}_m(x_m^{t+1} - x_m^t)\|.
\end{aligned} \tag{42}
$$

3. Bound for $\|\sum_{m=1}^{M} \mathcal{D}_m x_m^t - z^t\|$. According to Lemma 2, we have

$$
\|\sum_{m=1}^{M} \mathcal{D}_m x_m^t - z^t\| = \frac{1}{\rho}\|y^{t+1} - y^t\| \leq \frac{L}{\rho}\|z^{t+1} - z^t\|. \tag{43}
$$

Combining (39), (41), (42) and (43), we can conclude that there exists some $C_1 > 0$, such that

$$V^t \leq C_1(\|z^{t+1} - z^t\|^2 + \sum_{m=1}^{M} \|x_m^{t+1} - x_m^t\|^2), \tag{44}$$

By Lemma 4, there exists some constant $C_2 = \min\{\sum_{m=1}^{M} \frac{\gamma_m(\rho)}{2}, \frac{\gamma(\rho)}{2} - \frac{L^2}{\rho}\}$, such that

$$\begin{aligned}&\mathcal{L}(\{x_m^t\}, z^t; y^t) - \mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^{t+1})\\ \geq& C_2(\|z^{t+1} - z^t\|^2 + \sum_{m=1}^{M} \|x_m^{t+1} - x_m^t\|^2).\end{aligned} \tag{45}$$

By (44) and (45), we have

$$V^t \leq \frac{C_1}{C_2}\mathcal{L}(\{x_m^t\}, z^t; y^t) - \mathcal{L}(\{x_m^{t+1}\}, z^{t+1}; y^{t+1}). \tag{46}$$

Taking the sum over $t = 1, \ldots, T$, we have

$$\begin{aligned}\sum_{t=1}^{T} V^t \leq& \frac{C_1}{C_2}\mathcal{L}(\{x^1\}, z^1; y^1) - \mathcal{L}(\{x^{t+1}\}, z^{t+1}; y^{t+1})\\ \leq& \frac{C_1}{C_2}(\mathcal{L}(\{x^1\}, z^1; y^1) - \underline{f}).\end{aligned} \tag{47}$$

By the definition of $T(\epsilon)$, we have

$$T(\epsilon)\epsilon \leq \frac{C_1}{C_2}(\mathcal{L}(\{x^1\}, z^1; y^1) - \underline{f}). \tag{48}$$

By taking $C = \frac{C_1}{C_2}$, we complete the proof. □

Proof of Lemma 1

From the optimality condition of the x update procedure in (16), we can get

$$\begin{aligned}\mathcal{D}_m x_{m,\mathcal{D}_m}^{t+1} &= -\mathcal{D}_m(\rho\mathcal{D}_m^\top\mathcal{D}_m)^{-1}\left[\lambda R'_m(x_{m,\mathcal{D}_m}^{t+1}) + \mathcal{D}_m^\top y^t + \rho\mathcal{D}_m^\top(\sum_{\substack{k=1\\k\neq m}}^{M}\mathcal{D}_k\tilde{x}_k - z)\right],\\ \mathcal{D}'_m x_{m,\mathcal{D}'_m}^{t+1} &= -\mathcal{D}'_m(\rho\mathcal{D}'^\top_m\mathcal{D}'_m)^{-1}\left[\lambda R'_m(x_{m,\mathcal{D}'_m}^{t+1}) + \mathcal{D}'^\top_m y^t + \rho\mathcal{D}'^\top_m(\sum_{\substack{k=1\\k\neq m}}^{M}\mathcal{D}_k\tilde{x}_k - z)\right].\end{aligned}$$

Therefore we have

$$
\begin{aligned}
&\mathcal{D}_m x^{t+1}_{m,\mathcal{D}_m} - \mathcal{D}'_m x^{t+1}_{m,\mathcal{D}'_m} \\
&= -\mathcal{D}_m(\rho \mathcal{D}_m^\top \mathcal{D}_m)^{-1} \left[\lambda R'_m(x^{t+1}_{m,\mathcal{D}_m}) + \mathcal{D}_m^\top y^t \mathcal{D}_m + \rho \mathcal{D}_m^\top (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right] \\
&\quad + \mathcal{D}'_m(\rho \mathcal{D}'^\top_m \mathcal{D}'_m)^{-1} \left[\lambda R'_m(x^{t+1}_{m,\mathcal{D}'_m}) + \mathcal{D}'^\top_m y^t + \rho \mathcal{D}'^\top_m (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right] \\
&= \mathcal{D}_m(\rho \mathcal{D}_m^\top \mathcal{D}_m)^{-1} \\
&\quad \times \left[\lambda (R'_m(x^{t+1}_{m,\mathcal{D}'_m}) - R'_m(x^{t+1}_{m,\mathcal{D}_m})) + (\mathcal{D}'_m - \mathcal{D}_m)^\top y^t + \rho (\mathcal{D}'_m - \mathcal{D}_m)^\top (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right] \\
&\quad + [\mathcal{D}'_m(\rho \mathcal{D}'^\top_m \mathcal{D}'_m)^{-1} - \mathcal{D}_m(\rho \mathcal{D}_m^\top \mathcal{D}_m)^{-1}] \\
&\quad \times \left(\lambda R'_m(x^{t+1}_{m,\mathcal{D}'_m}) + \mathcal{D}'^\top_m y^t + \rho \mathcal{D}'^\top_m (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right).
\end{aligned}
$$

Denote

$$
\begin{aligned}
\Phi_1 &= \mathcal{D}_m(\rho \mathcal{D}_m^\top \mathcal{D}_m)^{-1} \\
&\quad \times \left[\lambda (R'_m(x^{t+1}_{m,\mathcal{D}'_m}) - R'_m(x^{t+1}_{m,\mathcal{D}_m})) + (\mathcal{D}'_m - \mathcal{D}_m)^\top y^t + \rho (\mathcal{D}'_m - \mathcal{D}_m)^\top (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right], \\
\Phi_2 &= [\mathcal{D}'_m(\rho \mathcal{D}'^\top_m \mathcal{D}'_m)^{-1} - \mathcal{D}_m(\rho \mathcal{D}_m^\top \mathcal{D}_m)^{-1}] \\
&\quad \times \left(\lambda R'_m(x^{t+1}_{m,\mathcal{D}'_m}) + \mathcal{D}'^\top_m y^t + \rho \mathcal{D}'^\top_m (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right).
\end{aligned}
$$

As a result:

$$
\mathcal{D}_m x^{t+1}_{m,\mathcal{D}_m} - \mathcal{D}'_m x^{t+1}_{m,\mathcal{D}'_m} = \Phi_1 + \Phi_2. \tag{49}
$$

In the following, we will analyze the components in (49) term by term. The object is to prove $\max_{\substack{\mathcal{D}_m, D'_m \\ \|\mathcal{D}_m - D'_m\| \leq 1}} \|x^{t+1}_{m,\mathcal{D}_m} - x^{t+1}_{m,\mathcal{D}'_m}\|$ is bounded. To see this, notice that

$$
\begin{aligned}
&\max_{\substack{\mathcal{D}_m, D'_m \\ \|\mathcal{D}_m - D'_m\| \leq 1}} \|\mathcal{D}_m x^{t+1}_{m,\mathcal{D}_m} - \mathcal{D}'_m x^{t+1}_{m,\mathcal{D}'_m}\| \\
&\leq \max_{\substack{\mathcal{D}_m, D'_m \\ \|\mathcal{D}_m - D'_m\| \leq 1}} \|\Phi_1\| + \max_{\substack{\mathcal{D}_m, D'_m \\ \|\mathcal{D}_m - D'_m\| \leq 1}} \|\Phi_2\|.
\end{aligned}
$$

For $\max_{\substack{\mathcal{D}_m, D'_m \\ \|\mathcal{D}_m - D'_m\| \leq 1}} \|\Phi_2\|$, from assumption 2.3, we have

$$
\begin{aligned}
&\max_{\substack{\mathcal{D}_m, D'_m \\ \|\mathcal{D}_m - D'_m\| \leq 1}} \|\Phi_2\| \\
&\leq \left\| \frac{2}{d_m \rho} \left(\lambda R'_m(x^{t+1}_{m,\mathcal{D}'_m}) + \mathcal{D}'^\top_m y^t + \rho \mathcal{D}'^\top_m (\sum_{\substack{k=1 \\ k\neq m}}^M \mathcal{D}_k \tilde{x}_k - z) \right) \right\|.
\end{aligned}
$$

By mean value theorem, we have

$$
\begin{aligned}
&\left\| \frac{2}{d_m \rho} \left(\lambda \mathcal{D}_m'^{\top} R_m''(x_*) + \mathcal{D}_m'^{\top} y^t + \rho \mathcal{D}_m'^{\top} (\sum_{\substack{k=1 \\ k \neq m}}^{M} \mathcal{D}_k \tilde{x}_k - z) \right) \right\| \\
&\leq \frac{2}{d_m \rho} \left[\lambda \| R_m''(\cdot) \| + \| y^t \| + \rho \| (\sum_{\substack{k=1 \\ k \neq m}}^{M} \mathcal{D}_k \tilde{x}_k - z) \| \right].
\end{aligned}
$$

For $\max_{\substack{\mathcal{D}_m, D_m' \\ \|\mathcal{D}_m - D_m'\| \leq 1}} \|\Phi_1\|$, we have

$$
\begin{aligned}
&\max_{\substack{\mathcal{D}_m, D_m' \\ \|\mathcal{D}_m - D_m'\| \leq 1}} \|\Phi_1\| \leq \Big\| \mathcal{D}_m (\rho \mathcal{D}_m^{\top} \mathcal{D}_m)^{-1} \\
&\times \left[\lambda (R_m'(x_{m,\mathcal{D}_m'}^{t+1}) - R_m'(x_{m,\mathcal{D}_m}^{t+1})) + (\mathcal{D}_m' - \mathcal{D}_m)^{\top} y^t + \rho (\mathcal{D}_m' - \mathcal{D}_m)^{\top} (\sum_{\substack{k=1 \\ k \neq m}}^{M} \mathcal{D}_k \tilde{x}_k - z) \right] \Big\| \\
&\leq \rho^{-1} \| (\mathcal{D}_m^{\top} \mathcal{D}_m)^{-1} \| \left[\lambda \| R_m''(\cdot) \| + \| y^t \| + \rho \| (\sum_{\substack{k=1 \\ k \neq m}}^{M} \mathcal{D}_k \tilde{x}_k - z)^{\top} \| \right] \\
&= \frac{1}{d_m \rho} \left[\lambda \| R_m''(\cdot) \| + \| y^t \| + \rho \| (\sum_{\substack{k=1 \\ k \neq m}}^{M} \mathcal{D}_k \tilde{x}_k - z) \| \right].
\end{aligned}
$$

Thus by assumption 2.1-2.2

$$
\begin{aligned}
&\max_{\substack{\mathcal{D}_m, D_m' \\ \|\mathcal{D}_m - D_m'\| \leq 1}} \| \mathcal{D}_m x_{m,\mathcal{D}_m}^{t+1} - \mathcal{D}_m' x_{m,\mathcal{D}_m'}^{t+1} \| \\
&\leq \frac{3}{d_m \rho} \left[\lambda c_1 + \| y^t \| + \rho \| (\sum_{\substack{k=1 \\ k \neq m}}^{M} \mathcal{D}_k \tilde{x}_k - z)^{\top} \| \right] \\
&\leq \frac{3}{d_m \rho} \left[\lambda c_1 + \| y^t \| + \rho \| z \| + \rho \sum_{\substack{k=1 \\ k \neq m}}^{M} \| \tilde{x}_k \| \right] \\
&\leq \frac{3}{d_m \rho} \left[\lambda c_1 + (1 + M\rho) b_1 \right]
\end{aligned}
$$

is bounded. □

Proof of Theorem 3

Proof: The privacy loss from $D_m \tilde{x}_m^{t+1}$ is calculated by:

$$
\left| \ln \frac{P(\mathcal{D}_m \tilde{x}_m^{t+1} | \mathcal{D}_m)}{P(\mathcal{D}_m' \tilde{x}_m^{t+1} | \mathcal{D}_m')} \right| = \left| \ln \frac{P(\mathcal{D}_m \tilde{x}_{m,\mathcal{D}_m}^{t+1} + \mathcal{D}_m \xi_m^{t+1})}{P(\mathcal{D}_m' \tilde{x}_{m,\mathcal{D}_m'}^{t+1} + \mathcal{D}_m' \xi_m^{\prime,t+1})} \right| = \left| \ln \frac{P(\mathcal{D}_m \xi_m^{t+1})}{P(\mathcal{D}_m' \xi_m^{\prime,t+1})} \right|.
$$

Since $\mathcal{D}_m\xi_m^{t+1}$ and $\mathcal{D}'_m\xi_m^{\prime,t+1}$ are sampled from $\mathcal{N}(0,\sigma_{m,t+1}^2)$, combine with lemma 1, we have

$$\begin{aligned}
&\left|\ln\frac{P(\mathcal{D}_m\xi_m^{t+1})}{P(\mathcal{D}'_m\xi_m^{\prime,t+1})}\right| \\
&=\left|\frac{2\xi_m^{t+1}\|\mathcal{D}_m x_{m,\mathcal{D}_m}^{t+1}-\mathcal{D}'_m x_{m,\mathcal{D}'_m}^{t+1}\|+\|\mathcal{D}_m x_{m,\mathcal{D}_m}^{t+1}-\mathcal{D}'_m x_{m,\mathcal{D}'_m}^{t+1}\|^2}{2\sigma_{m,t+1}^2}\right| \\
&\le\left|\frac{2\mathcal{D}_m\xi_m^{t+1}\mathbb{C}+\mathbb{C}^2}{2\frac{\mathbb{C}^2\cdot 2\ln(1.25/\sigma)}{\varepsilon^2}}\right| \\
&=\left|\frac{(2\mathcal{D}_m\xi_m^{t+1}+\mathbb{C})\varepsilon^2}{4\mathbb{C}\ln(1.25/\sigma)}\right|.
\end{aligned}$$

In order to make $\left|\frac{(2\mathcal{D}_m\xi_m^{t+1}+\mathbb{C})\varepsilon^2}{4\mathbb{C}\ln(1.25/\sigma)}\right|\le\varepsilon$, we need to make sure

$$\left|\mathcal{D}_m\xi_m^{t+1}\right|\le\frac{2\mathbb{C}\ln(1.25/\sigma)}{\varepsilon}-\frac{\mathbb{C}}{2}.$$

In the following, we need to proof

$$P(\left|\mathcal{D}_m\xi_m^{t+1}\right|\ge\frac{2\mathbb{C}\ln(1.25/\sigma)}{\varepsilon}-\frac{\mathbb{C}}{2})\le\delta \tag{50}$$

holds. However, we will proof a stronger result that lead to (50). Which is

$$P(\mathcal{D}_m\xi_m^{t+1}\ge\frac{2\mathbb{C}\ln(1.25/\sigma)}{\varepsilon}-\frac{\mathbb{C}}{2})\le\frac{\delta}{2}.$$

Since the tail bound of normal distribution $\mathcal{N}(0,\sigma_{m,t+1}^2)$ is:

$$P(\mathcal{D}_m\xi_m^{t+1}>r)\le\frac{\sigma_{m,t+1}}{r\sqrt{2\pi}}e^{-\frac{r^2}{2\sigma_{m,t+1}^2}}.$$

Let $r=\frac{2\mathbb{C}\ln(1.25/\sigma)}{\varepsilon}-\frac{\mathbb{C}}{2}$, we then have

$$\begin{aligned}
&P(\mathcal{D}_m\xi_m^{t+1}\ge\frac{2\mathbb{C}\ln(1.25/\sigma)}{\varepsilon}-\frac{\mathbb{C}}{2}) \\
&\le\frac{\mathbb{C}\sqrt{2\ln(1.25/\sigma)}}{r\sqrt{2\pi}\varepsilon}\exp\left[-\frac{[4\ln(1.25/\sigma)-\varepsilon]^2}{8\ln(1.25/\sigma)}\right].
\end{aligned}$$

When δ is small and let $\varepsilon\le 1$, we then have

$$\frac{\sqrt{2\ln(1.25/\sigma)}2}{(4\ln(1.25/\sigma)-\varepsilon)\sqrt{2\pi}}\le\frac{\sqrt{2\ln(1.25/\sigma)}2}{(4\ln(1.25/\sigma)-1)\sqrt{2\pi}}<\frac{1}{\sqrt{2\pi}}. \tag{51}$$

As a result, we can proof that

$$-\frac{[4\ln(1.25/\sigma)-\varepsilon]^2}{8\ln(1.25/\sigma)}<\ln(\sqrt{2\pi}\frac{\delta}{2})$$

by equation (51). Thus we have

$$P(\mathcal{D}_m\xi_m^{t+1}\ge\frac{2\mathbb{C}\ln(1.25/\sigma)}{\varepsilon}-\frac{\mathbb{C}}{2})<\frac{1}{\sqrt{2\pi}}\exp(\ln(\sqrt{2\pi}\frac{\delta}{2})=\frac{\delta}{2}.$$

Thus we proved (50) holds. Define

$$\begin{aligned}
\mathbb{A}_1&=\{\mathcal{D}_m\xi_m^{t+1}:|\mathcal{D}_m\xi_m^{t+1}|\le\frac{1}{\sqrt{2\pi}}\exp(\ln(\sqrt{2\pi}\frac{\delta}{2}\},\\
\mathbb{A}_2&=\{\mathcal{D}_m\xi_m^{t+1}:|\mathcal{D}_m\xi_m^{t+1}|>\frac{1}{\sqrt{2\pi}}\exp(\ln(\sqrt{2\pi}\frac{\delta}{2}\}.
\end{aligned}$$

Thus we obtain the desired result:

$$\begin{aligned}
&P(\mathcal{D}'_m\tilde{x}_m^{t+1}|\mathcal{D}_m)\\
&=P(\mathcal{D}_m x_{m,\mathcal{D}_m}^{t+1}+\mathcal{D}_m\xi_m^{t+1}:\mathcal{D}_m\xi_m^{t+1}\in\mathbb{A}_1)\\
&+P(\mathcal{D}_m x_{m,\mathcal{D}_m}^{t+1}+\mathcal{D}_m\xi_m^{t+1}:\mathcal{D}_m\xi_m^{t+1}\in\mathbb{A}_2)\\
&<e^{\varepsilon}P(\mathcal{D}_m x_{m,\mathcal{D}'_m}^{t+1}+\mathcal{D}_m\xi_m^{\prime,t+1})+\delta=e^{\varepsilon}P(\mathcal{D}_m\tilde{x}_m^{t+1}|\mathcal{D}'_m)+\delta.
\end{aligned}$$

$\square$