Supplementary Material

A. Proof of Theorem 2

In the beginning, we define several auxiliary variables, which will be used in this proof.

Let $\bar{\mathbf{z}}(m)=\frac{1}{n}\sum_{i=1}^n\mathbf{z}_i(m)$ and $\bar{\mathbf{g}}(m)=\frac{1}{n}\sum_{i=1}^n\widehat{\mathbf{g}}_i(m)$. Then, we define

$$\bar{F}_{m+1}(\mathbf{x})=\eta\bar{\mathbf{z}}(m+1)^\top\mathbf{x}+\|\mathbf{x}\|_2^2$$

and $\bar{\mathbf{x}}(m+1)=\underset{\mathbf{x}\in\mathcal{K}_\delta}{\operatorname{argmin}}\,\bar{F}_{m+1}(\mathbf{x})$. Similarly, let $\widehat{\mathbf{x}}_i(m)=\underset{\mathbf{x}\in\mathcal{K}_\delta}{\operatorname{argmin}}\,\eta\mathbf{z}_i(m)^\top\mathbf{x}+\|\mathbf{x}\|_2^2$.

Moreover, we introduce the following two lemmas with respect to the theoretical guarantees of δ-smoothed function.

Lemma 8 *(Lemma 2.6 in Hazan (2016)) Let $f(\mathbf{x}):\mathbb{R}^d\to\mathbb{R}$ be convex and G-Lipschitz over a convex and compact set $\mathcal{K}\subset\mathbb{R}^d$. Then, $\widehat{f}_\delta(\mathbf{x})$ is convex and G-Lipschitz over $\mathcal{K}_\delta$, and it holds that $|\widehat{f}_\delta(\mathbf{x})-f(\mathbf{x})|\leq\delta G$ for any $\mathbf{x}\in\mathcal{K}_\delta$.*

Lemma 9 *(Lemma 4 in Garber & Kretzu (2019)) Let $f(\mathbf{x}):\mathbb{R}^d\to\mathbb{R}$ be convex and suppose that all subgradients of f are upper bounded by G in ℓ_2-norm over a convex and compact set $\mathcal{K}\subset\mathbb{R}^d$. For any $\mathbf{x}\in\mathcal{K}_\delta$, $\|\nabla\widehat{f}_\delta(\mathbf{x})\|_2\leq G$.*

We first assume that for all $i\in V$ and $m=1,\cdots,B$,

$$\|\widehat{\mathbf{g}}_i(m)\|_2\leq\beta.$$

Let $\mathbf{x}^*\in\underset{\mathbf{x}\in\mathcal{K}}{\operatorname{argmin}}\sum_{t=1}^T f_t(\mathbf{x})$ and $\widetilde{\mathbf{x}}^*=(1-\delta/r)\mathbf{x}^*$. For any $i,j\in V$, we have

$$\begin{aligned}\sum_{t=1}^T f_{t,j}(\mathbf{y}_i(t))-\sum_{t=1}^T f_{t,j}(\mathbf{x}^*)&=\sum_{t=1}^T f_{t,j}(\mathbf{x}_i(m_t)+\delta\mathbf{u}_i(t))-\sum_{t=1}^T f_{t,j}(\mathbf{x}^*)\\&\leq\sum_{t=1}^T(f_{t,j}(\mathbf{x}_i(m_t))+G\|\delta\mathbf{u}_i(t)\|_2)-\sum_{t=1}^T(f_{t,j}(\widetilde{\mathbf{x}}^*)-G\|\widetilde{\mathbf{x}}^*-\mathbf{x}^*\|_2)\\&\leq\sum_{t=1}^T f_{t,j}(\mathbf{x}_i(m_t))-\sum_{t=1}^T f_{t,j}(\widetilde{\mathbf{x}}^*)+\delta GT+\frac{\delta GRT}{r}\\&\leq\sum_{t=1}^T(\widehat{f}_{t,j,\delta}(\mathbf{x}_i(m_t))+\delta G)-\sum_{t=1}^T(\widehat{f}_{t,j,\delta}(\widetilde{\mathbf{x}}^*)-\delta G)+\delta GT+\frac{\delta GRT}{r}\\&\leq\sum_{t=1}^T(\widehat{f}_{t,j,\delta}(\mathbf{x}_i(m_t))-\widehat{f}_{t,j,\delta}(\widetilde{\mathbf{x}}^*))+3\delta GT+\frac{\delta GRT}{r}\end{aligned}\tag{16}$$

where the first inequality is due to Assumption 1 and the third inequality is due to Lemma 8.

Then, similar to the proof of Theorem 1, we derive an upper bound of $\|\mathbf{x}_i(m)-\bar{\mathbf{x}}(m)\|_2$ by further introducing the following lemma.

Lemma 10 *Let $\widehat{\mathbf{x}}_i(m)=\underset{\mathbf{x}\in\mathcal{K}_\delta}{\operatorname{argmin}}\,F_{m,i}(\mathbf{x})$, for $m\in[B]$. Assume $\|\widehat{\mathbf{g}}_i(m)\|_2\leq\beta$ for any $i\in V$ and $m\in[B]$, Algorithm 3 with $\epsilon\leq 8R^2$ and $L=\frac{16R^2}{\epsilon^2}(\eta\alpha\beta\sqrt{\epsilon}+\eta^2\alpha^2\beta^2)$ has*

$$F_{m,i}(\mathbf{x}_i(m))-F_{m,i}(\widehat{\mathbf{x}}_i(m))\leq\epsilon$$

for any $i\in V$ and $m\in[B]$, where $\alpha=\frac{1+\sigma_2(P)}{1-\sigma_2(P)}\sqrt{n}+1$.

Applying Lemma 2 with $\|\widehat{\mathbf{g}}_i(m)\|_2\leq\beta$, we have

$$\|\mathbf{z}_i(m)-\bar{\mathbf{z}}(m)\|_2\leq\alpha'\beta\tag{17}$$

where $\alpha' = \frac{\sqrt{n}}{1-\sigma_2(P)}$.

Furthermore, applying Lemma 3 with (17), we have

$$\|\widehat{\mathbf{x}}_i(m) - \bar{\mathbf{x}}(m)\|_2 \leq \eta\|\mathbf{z}_i(m) - \bar{\mathbf{z}}(m)\|_2 \leq \eta\alpha'\beta$$

which implies that

$$\begin{aligned}\|\mathbf{x}_i(m) - \bar{\mathbf{x}}(m)\|_2 \leq &\|\mathbf{x}_i(m) - \widehat{\mathbf{x}}_i(m)\|_2 + \|\widehat{\mathbf{x}}_i(m) - \bar{\mathbf{x}}(m)\|_2 \\ \leq &\sqrt{F_{m,i}(\mathbf{x}_i(m)) - F_{m,i}(\widehat{\mathbf{x}}_i(m))} + \eta\alpha'\beta \\ \leq &\sqrt{\epsilon} + \eta\alpha'\beta\end{aligned} \tag{18}$$

where the second inequality is due to the fact $F_{m,i}(\mathbf{x})$ is 2-strongly convex and (5), and the last inequality is due to Lemma 10.

For brevity, let $\epsilon' = \sqrt{\epsilon} + \eta\alpha'\beta$. Then, we can use (18) to bound the first term in the right side of (16) as

$$\begin{aligned}&\sum_{t=1}^{T}(\widehat{f}_{t,j,\delta}(\mathbf{x}_i(m_t)) - \widehat{f}_{t,j,\delta}(\widetilde{\mathbf{x}}^*)) \\ \leq &\sum_{t=1}^{T}(\widehat{f}_{t,j,\delta}(\bar{\mathbf{x}}(m_t)) - \widehat{f}_{t,j,\delta}(\widetilde{\mathbf{x}}^*)) + \sum_{t=1}^{T} G\|\bar{\mathbf{x}}(m_t) - \mathbf{x}_i(m_t)\|_2 \\ \leq &\sum_{t=1}^{T}(\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t)) - \widehat{f}_{t,j,\delta}(\widetilde{\mathbf{x}}^*)) + \sum_{t=1}^{T} G\|\bar{\mathbf{x}}(m_t) - \mathbf{x}_j(m_t)\|_2 + GT\epsilon' \\ \leq &\sum_{t=1}^{T} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\mathbf{x}_j(m_t) - \widetilde{\mathbf{x}}^*) + 2GT\epsilon' \\ = &\sum_{t=1}^{T} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\mathbf{x}_j(m_t) - \bar{\mathbf{x}}(m_t)) + \sum_{t=1}^{T} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) + 2GT\epsilon' \\ \leq &\sum_{t=1}^{T} \|\nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))\|_2\|\mathbf{x}_j(m_t) - \bar{\mathbf{x}}(m_t)\|_2 + \sum_{t=1}^{T} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) + 2GT\epsilon' \\ \leq &\sum_{t=1}^{T} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) + \sum_{t=1}^{T} G\|\bar{\mathbf{x}}(m_t) - \mathbf{x}_j(m_t)\|_2 + 2GT\epsilon' \\ \leq &\sum_{t=1}^{T} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) + 3GT\epsilon'\end{aligned} \tag{19}$$

where the third inequality is due to the convexity of $\widehat{f}_{t,j,\delta}(\mathbf{x})$ and the fifth inequality is due to Lemma 9.

Combining (16), (19) and $\epsilon' = \sqrt{\epsilon} + \eta\alpha'\beta$, for any $i \in V$, we have

$$\begin{aligned}&\sum_{t=1}^{T}\sum_{j=1}^{n} f_{t,j}(\mathbf{y}_i(t)) - \sum_{t=1}^{T}\sum_{j=1}^{n} f_{t,j}(\mathbf{x}^*) \\ \leq &\sum_{t=1}^{T}\sum_{j=1}^{n} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) + 3\delta nGT + \frac{\delta nGRT}{r} + 3nGT\left(\sqrt{\epsilon} + \eta\alpha'\beta\right).\end{aligned}$$

Moreover, to bound $\sum_{t=1}^{T}\sum_{j=1}^{n} \nabla\widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*)$, we introduce the following lemma.

Lemma 11 *Let* $\bar{\mathbf{z}}(m) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(m)$ *and* $\bar{\mathbf{g}}(m) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{g}}_i(m)$. *Moreover, we define*

$$\bar{F}_{m+1}(\mathbf{x}) = \eta\bar{\mathbf{z}}(m+1)^\top\mathbf{x} + \|\mathbf{x}\|_2^2$$

and $\bar{\mathbf{x}}(m+1) = \underset{\mathbf{x}\in\mathcal{K}_\delta}{\operatorname{argmin}} \bar{F}_{m+1}(\mathbf{x})$. *Assume* $\|\widehat{\mathbf{g}}_i(m)\|_2 \leq \beta$ *for any* $i \in V$ *and* $m \in [B]$, *with probability at least* $1-\gamma$, *Algorithm 3 has*

$$\sum_{t=1}^{T}\sum_{j=1}^{n} \nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^\top (\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) \leq 2nR(KG+\beta)\sqrt{2B\ln\frac{1}{\gamma}} + \frac{nR^2}{\eta} + n\eta B\beta^2$$

where $\widetilde{\mathbf{x}}^* = (1-\delta/r)\mathbf{x}^*$ *and* $\mathbf{x}^* \in \underset{\mathbf{x}\in\mathcal{K}}{\operatorname{argmin}} \sum_{t=1}^{T} f_t(\mathbf{x})$.

According to Lemma 11, assume that $\|\widehat{\mathbf{g}}_i(m)\|_2 \leq \beta$ for any $i \in V$ and $m \in [B]$, with probability at least $1-\gamma$, we have

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{j=1}^{n} f_{t,j}(\mathbf{y}_i(t)) - \sum_{t=1}^{T}\sum_{j=1}^{n} f_{t,j}(\mathbf{x}^*) \\
\leq& 2nR(KG+\beta)\sqrt{2B\ln\frac{1}{\gamma}} + \frac{nR^2}{\eta} + n\eta B\beta^2 + 3\delta nGT + \frac{\delta nGRT}{r} + 3nGT\left(\sqrt{\epsilon} + \eta\alpha'\beta\right).
\end{aligned}$$

Substituting $\eta = \frac{cR}{\alpha_T dM}T^{-3/4}$, $\delta = cT^{-1/4}$, $\epsilon = 4R^2T^{-1/2}$, $\beta = \alpha_T\frac{dM\sqrt{K}}{\delta} + KG$ and $K = T^{1/2}$ into the above inequality, we have

$$\begin{aligned}
R_{T,i} \leq& 2nR\left(2G + \frac{\alpha_T dM}{c}\right)\sqrt{2\ln\frac{1}{\gamma}}T^{3/4} + \frac{\alpha_T ndMR}{c}T^{3/4} \\
&+ n\left(R + \frac{cRG}{\alpha_T dM}\right)\left(\frac{\alpha_T dM}{c} + G\right)T^{3/4} \\
&+ 3cnGT^{3/4} + \frac{cnGR}{r}T^{3/4} + 6nGRT^{3/4} \\
&+ 3\alpha' nG\left(R + \frac{cRG}{\alpha_T dM}\right)T^{3/4} \\
\leq& O\left(\alpha_T T^{3/4}\right).
\end{aligned}$$

Let $\mathcal{A}$ denote the event of $\|\widehat{\mathbf{g}}_i(m)\|_2 \leq \beta, \forall i \in V, m \in [B]$. Because we have used the event $\mathcal{A}$ as a fact, the above result should be formulated as

$$\Pr\left(R_{T,i} \leq O\left(\alpha_T T^{3/4}\right) \middle| \mathcal{A}\right) \geq 1-\gamma. \tag{20}$$

Furthermore, we introduce the following lemma with respect to the probability of the event $\mathcal{A}$.

Lemma 12 *For all* $i \in V$ *and* $m \in [B]$, *Algorithm 3 has*

$$\|\widehat{\mathbf{g}}_i(m)\|_2 \leq \left(1 + \sqrt{8\ln\frac{nB}{\gamma}}\right)\frac{dM\sqrt{K}}{\delta} + KG$$

with probability at least $1-\gamma$.

Then, applying Lemma 12 with $B = T/K = \sqrt{T}$, we have

$$\Pr(\mathcal{A}) \geq 1-\gamma. \tag{21}$$

Combining (20) with (21), we complete the proof.

B. Proof of Lemma 10

For $m = 1$, because $\mathbf{x}_i(1) = \widehat{\mathbf{x}}_i(1) = \underset{\mathbf{x}\in\mathcal{K}_\delta}{\operatorname{argmin}} \|\mathbf{x}\|_2^2$, we have

$$F_{1,i}(\mathbf{x}_i(1)) - F_{1,i}(\widehat{\mathbf{x}}_i(1)) = 0 \leq \epsilon. \tag{22}$$

Then, for $m = 2$, we have

$$
\begin{aligned}
&F_{m,i}(\mathbf{x}_i(m-1)) - F_{m,i}(\widehat{\mathbf{x}}_i(m)) \\
=&F_{m-1,i}(\mathbf{x}_i(m-1)) + \eta(\mathbf{z}_i(m) - \mathbf{z}_i(m-1))^\top \mathbf{x}_i(m-1) \\
&- F_{m-1,i}(\widehat{\mathbf{x}}_i(m)) - \eta(\mathbf{z}_i(m) - \mathbf{z}_i(m-1))^\top \widehat{\mathbf{x}}_i(m) \\
\leq&F_{m-1,i}(\mathbf{x}_i(m-1)) - F_{m-1,i}(\widehat{\mathbf{x}}_i(m-1)) \\
&+ \eta(\mathbf{z}_i(m) - \mathbf{z}_i(m-1))^\top (\mathbf{x}_i(m-1) - \widehat{\mathbf{x}}_i(m)) \\
\leq&\epsilon + \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\|\mathbf{x}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2 \\
\leq&\epsilon + \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\|\mathbf{x}_i(m-1) - \widehat{\mathbf{x}}_i(m-1)\|_2 \\
&+ \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\|\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2 \\
\leq&\epsilon + \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\sqrt{F_{m-1,i}(\mathbf{x}_i(m-1)) - F_{m-1,i}(\widehat{\mathbf{x}}_i(m-1))} \\
&+ \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\|\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2 \\
\leq&\epsilon + \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\sqrt{\epsilon} + \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\|\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2
\end{aligned} \tag{23}
$$

where the first inequality is due to $\widehat{\mathbf{x}}_i(m-1) = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{K}_\delta} F_{m-1,i}(\mathbf{x})$ and the fourth inequality is due to that $F_{m-1}(\mathbf{x})$ is 2-strongly convex and (5).

Moreover, because for each $m = 1, \cdots, B$, $F_{m,i}(\mathbf{x})$ is 2-strongly convex, we also have

$$
\begin{aligned}
\|\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2^2 \leq&F_{m,i}(\widehat{\mathbf{x}}_i(m-1)) - F_{m,i}(\widehat{\mathbf{x}}_i(m)) \\
=&F_{m-1,i}(\widehat{\mathbf{x}}_i(m-1)) + \eta(\mathbf{z}_i(m) - \mathbf{z}_i(m-1))^\top \widehat{\mathbf{x}}_i(m-1) \\
&- F_{m-1,i}(\widehat{\mathbf{x}}_i(m)) - \eta(\mathbf{z}_i(m) - \mathbf{z}_i(m-1))^\top \widehat{\mathbf{x}}_i(m) \\
=&F_{m-1,i}(\widehat{\mathbf{x}}_i(m-1)) - F_{m-1,i}(\widehat{\mathbf{x}}_i(m)) \\
&+ \eta(\mathbf{z}_i(m) - \mathbf{z}_i(m-1))^\top (\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)) \\
\leq&\eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\|\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2
\end{aligned}
$$

which further implies that

$$
\|\widehat{\mathbf{x}}_i(m-1) - \widehat{\mathbf{x}}_i(m)\|_2 \leq \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2. \tag{24}
$$

For $m \in [B]$, applying Lemma 6 with $\|\widehat{\mathbf{g}}_i(m)\|_2 \leq \beta$, we have

$$
\|\mathbf{z}_i(m+1) - \mathbf{z}_i(m)\|_2 \leq \alpha\beta. \tag{25}
$$

Substituting (24) and (25) into (23), we have

$$
\begin{aligned}
F_{m,i}(\mathbf{x}_i(m-1)) - F_{m,i}(\widehat{\mathbf{x}}_i(m)) \leq&\epsilon + \eta\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2\sqrt{\epsilon} + \eta^2\|\mathbf{z}_i(m) - \mathbf{z}_i(m-1)\|_2^2 \\
\leq&\epsilon + \eta\alpha\beta\sqrt{\epsilon} + \eta^2\alpha^2\beta^2.
\end{aligned}
$$

According to Algorithm 3, we have $\mathbf{x}_i(m) = \text{CGSC}(\mathcal{K}_\delta, \epsilon, L, F_{m,i}(\mathbf{x}), \mathbf{x}_i(m-1))$. Because $F_{m,i}(\mathbf{x})$ is 2-smooth and 2-strongly convex, $\epsilon \leq 8R^2$ and $L = \frac{16R^2}{\epsilon^2}(\eta\alpha\beta\sqrt{\epsilon} + \eta^2\alpha^2\beta^2)$, applying Lemma 7 with $\mathcal{K}' = \mathcal{K}_\delta$, we have

$$
F_{m,i}(\mathbf{x}_i(m)) - F_{m,i}(\widehat{\mathbf{x}}_i(m)) \leq \epsilon
$$

for $m = 2$. By induction, we can complete the proof for $m = 1, \cdots, B$.

C. Proof of Lemma 11

We first introduce the classical Azuma's inequality (Azuma, 1967) for martingales in the following lemma.

Lemma 13 *Suppose $D_1, \cdots, D_r$ is a martingale difference sequence and*

$$|D_j| \leq c_j$$

almost surely. Then, we have

$$\Pr\left(\sum_{j=1}^{r} D_j \geq \Delta\right) \leq \exp\left(\frac{-\Delta^2}{2\sum_{j=1}^{r} c_j^2}\right).$$

To apply Lemma 13, with $\mathcal{T}_m = \{(m-1)K+1, \cdots, mK\}$, we define

$$\begin{aligned} D_m &= \sum_{t\in\mathcal{T}_m}\sum_{j=1}^{n}\left(\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m)) - \mathbf{g}_j(t)\right)^{\top}(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*) \\ &= \sum_{j=1}^{n}\left(\sum_{t\in\mathcal{T}_m}\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m)) - \widehat{\mathbf{g}}_j(m)\right)^{\top}(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*). \end{aligned} \tag{26}$$

According to Algorithm 3 and Lemma 1, we have

$$\mathbb{E}\left[D_m | \, \mathbf{x}_1(m), \cdots, \mathbf{x}_n(m), \bar{\mathbf{x}}(m)\right] = 0$$

which further implies that $D_1, \cdots, D_B$ is a martingale difference sequence with

$$\begin{aligned} |D_m| &= \left|\sum_{j=1}^{n}\left(\sum_{t\in\mathcal{T}_m}\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m)) - \widehat{\mathbf{g}}_j(m)\right)^{\top}(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*)\right| \\ &\leq \sum_{j=1}^{n}\left\|\sum_{t\in\mathcal{T}_m}\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m)) - \widehat{\mathbf{g}}_j(m)\right\|_2 \|(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*)\|_2 \\ &\leq 2R\sum_{j=1}^{n}\left(\left\|\sum_{t\in\mathcal{T}_m}\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m))\right\|_2 + \|\widehat{\mathbf{g}}_j(m)\|_2\right) \\ &\leq 2R\sum_{j=1}^{n}\sum_{t\in\mathcal{T}_m}\left\|\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m))\right\|_2 + 2nR\beta \\ &\leq 2nRKG + 2nR\beta \end{aligned}$$

where the last inequality is due to Lemma 9.

Then, applying Lemma 13 with $\Delta = 2nR(KG+\beta)\sqrt{2B\ln\frac{1}{\gamma}}$, with probability at least $1-\gamma$, we have

$$\sum_{m=1}^{B} D_m \leq \Delta = 2nR(KG+\beta)\sqrt{2B\ln\frac{1}{\gamma}}. \tag{27}$$

Additionally, combining (26) with $\bar{\mathbf{g}}(m) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{g}}_i(m)$, we further have

$$\sum_{t=1}^{T}\sum_{j=1}^{n}\nabla \widehat{f}_{t,j,\delta}(\mathbf{x}_j(m_t))^{\top}(\bar{\mathbf{x}}(m_t) - \widetilde{\mathbf{x}}^*) = \sum_{m=1}^{B} D_m + n\sum_{m=1}^{B}\bar{\mathbf{g}}(m)^{\top}(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*). \tag{28}$$

Therefore, we still need to bound $\sum_{m=1}^{B}\bar{\mathbf{g}}(m)^{\top}(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*)$. According to Assumption 4, it is easy to verify that

$$\bar{\mathbf{z}}(m+1) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(m+1) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j\in N_i} P_{ij}\mathbf{z}_j(m) + \widehat{\mathbf{g}}_i(m)\right) = \bar{\mathbf{z}}(m) + \bar{\mathbf{g}}(m).$$

Moreover, according to the definition, we have

$$\bar{\mathbf{x}}(m+1) = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{K}_\delta} \bar{F}_{m+1}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{K}_\delta} \eta\bar{\mathbf{z}}(m+1)^\top\mathbf{x} + \|\mathbf{x}\|_2^2.$$

So, applying Lemma 5 with the linear loss functions $\left\{\bar{\mathbf{g}}(m)^\top\mathbf{x}\right\}_{m=1}^B$, the decision set $\mathcal{K} = \mathcal{K}_\delta$ and the regularizer $\mathcal{R}(\mathbf{x}) = \frac{\|\mathbf{x}\|_2^2}{\eta}$, we have

$$\begin{aligned}
\sum_{m=1}^B \bar{\mathbf{g}}(m)^\top(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*) &\le \frac{\|\widetilde{\mathbf{x}}^*\|_2^2}{\eta} - 0 + \sum_{m=1}^B \bar{\mathbf{g}}(m)^\top(\bar{\mathbf{x}}(m) - \bar{\mathbf{x}}(m+1)) \\
&\le \frac{R^2}{\eta} + \sum_{m=1}^B \|\bar{\mathbf{g}}(m)\|_2\|\bar{\mathbf{x}}(m) - \bar{\mathbf{x}}(m+1)\|_2.
\end{aligned} \tag{29}$$

Then, it is easy to verify that $\bar{F}_{m+1}(\mathbf{x})$ is 2-strongly convex, which implies that

$$\begin{aligned}
\|\bar{\mathbf{x}}(m) - \bar{\mathbf{x}}(m+1)\|_2^2 &\le \bar{F}_{m+1}(\bar{\mathbf{x}}(m)) - \bar{F}_{m+1}(\bar{\mathbf{x}}(m+1)) \\
&= \bar{F}_m(\bar{\mathbf{x}}(m)) + \eta\bar{\mathbf{g}}(m)^\top\bar{\mathbf{x}}(m) - \bar{F}_m(\bar{\mathbf{x}}(m+1)) - \eta\bar{\mathbf{g}}(m)^\top\bar{\mathbf{x}}(m+1) \\
&= \bar{F}_m(\bar{\mathbf{x}}(m)) - \bar{F}_m(\bar{\mathbf{x}}(m+1)) + \eta\bar{\mathbf{g}}(m)^\top\left(\bar{\mathbf{x}}(m) - \bar{\mathbf{x}}(m+1)\right) \\
&\le \eta\|\bar{\mathbf{g}}(m)\|_2\|\bar{\mathbf{x}}(m) - \bar{\mathbf{x}}(m+1)\|_2.
\end{aligned}$$

The above inequality can be simplified as

$$\|\bar{\mathbf{x}}(m) - \bar{\mathbf{x}}(m+1)\|_2 \le \eta\|\bar{\mathbf{g}}(m)\|_2. \tag{30}$$

Substituting (30) into (29), we have

$$\begin{aligned}
\sum_{m=1}^B \bar{\mathbf{g}}(m)^\top(\bar{\mathbf{x}}(m) - \widetilde{\mathbf{x}}^*) &\le \frac{R^2}{\eta} + \eta\sum_{m=1}^B \|\bar{\mathbf{g}}(m)\|_2^2 \\
&= \frac{R^2}{\eta} + \eta\sum_{m=1}^B \left\|\frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{g}}_i(m)\right\|_2^2 \\
&\le \frac{R^2}{\eta} + \frac{\eta}{n}\sum_{m=1}^B\sum_{i=1}^n \|\widehat{\mathbf{g}}_i(m)\|_2^2 \\
&= \frac{R^2}{\eta} + \eta B\beta^2.
\end{aligned} \tag{31}$$

Finally, substituting (27) and (31) into (28), we complete the proof.

D. Proof of Lemma 12

According to Algorithm 3, for any $i \in V$ and $m = 1, \cdots, B$, conditioned on $\mathbf{x}_i(m)$,

$$\mathbf{g}_i((m-1)K+1), \cdots, \mathbf{g}_i(mK)$$

are K independent random vectors.

For brevity, for $j = 1, \cdots, K$, let

$$X_j = \mathbf{g}_i(t_j)$$

where $t_j = (m-1)K + j$, and let $N = \left\|\sum_{j=1}^K X_j\right\|_2$, $\widehat{S}_j = \sum_{k\neq j} X_k$ and $\widehat{\mathbf{X}}_j$ be the set

$$\{X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_K\}.$$

To bound N by using Lemma 13, we define $\mathbf{X}_0 = \emptyset$, $\mathbf{X}_j = \{X_1, \cdots, X_j\}$ for $j \geq 1$ and a sequence $D_1, \cdots, D_K$ as

$$D_j = \mathbb{E}[N|\mathbf{X}_j] - \mathbb{E}[N|\mathbf{X}_{j-1}].$$

It is not hard to verify that

$$\mathbb{E}[D_j|\mathbf{X}_{j-1}] = \mathbb{E}[(\mathbb{E}[N|\mathbf{X}_j] - \mathbb{E}[N|\mathbf{X}_{j-1}])|\mathbf{X}_{j-1}] = 0$$

which implies that $D_1, \cdots, D_K$ is a martingale difference sequence.

Moreover, we have

$$|D_j| = |\mathbb{E}[N|\mathbf{X}_j] - \mathbb{E}[N|\mathbf{X}_{j-1}]| \leq \sup_{\widehat{\mathbf{X}}_j} \left| N - \mathbb{E}[N|\widehat{\mathbf{X}}_j] \right|. \tag{32}$$

Using the triangle inequality, we have

$$N \leq \|\widehat{S}_j\|_2 + \|X_j\|_2 \text{ and } N \geq \|\widehat{S}_j\|_2 - \|X_j\|_2. \tag{33}$$

According to the Algorithm 3, we have

$$\|X_j\|_2 = \left\| \frac{d}{\delta} f_{t_j,i}(\mathbf{y}_i(t_j))\mathbf{u}_i(t_j) \right\|_2 \leq \frac{dM}{\delta}.$$

Therefore, combining (32) with (33) and the above inequality, we further have

$$|D_j| \leq \|X_j\|_2 + \mathbb{E}[\|X_j\|_2|\widehat{\mathbf{X}}_j] \leq \frac{2dM}{\delta}. \tag{34}$$

Let $\Delta = \frac{\sqrt{K}dM}{\delta}\sqrt{8\ln\frac{nB}{\gamma}}$. Then, applying Lemma 13, with probability at least $1 - \frac{\gamma}{nB}$, we have

$$N - \mathbb{E}[N] = \mathbb{E}[N|\mathbf{X}_K] - \mathbb{E}[N|\mathbf{X}_0] = \sum_{j=1}^{K} D_j \leq \frac{\sqrt{K}dM}{\delta}\sqrt{8\ln\frac{nB}{\gamma}}$$

which implies that

$$\|\widehat{\mathbf{g}}_i(m)\|_2 = N \leq \frac{\sqrt{K}dM}{\delta}\sqrt{8\ln\frac{nB}{\gamma}} + \mathbb{E}[N] \leq \frac{\sqrt{K}dM}{\delta}\sqrt{8\ln\frac{nB}{\gamma}} + \sqrt{\mathbb{E}[N^2]}. \tag{35}$$

It is easy to provide an upper bound of $\mathbb{E}[N^2]$ by following the proof of Lemma 5 in Garber & Kretzu (2019). We include the detailed proof for completeness.

According to the definition, we have

$$\begin{aligned}
\mathbb{E}[N^2] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K} X_j^\top X_j \middle| \mathbf{x}_i(m)\right]\right] + \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K}\sum_{k\in[K]\cap k\neq j} X_j^\top X_k \middle| \mathbf{x}_i(m)\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K} \|X_j\|_2^2 \middle| \mathbf{x}_i(m)\right]\right] + \mathbb{E}\left[\sum_{j=1}^{K}\sum_{k\in[K]\cap k\neq j} \mathbb{E}\left[X_j|\,\mathbf{x}_i(m)\right]^\top \mathbb{E}\left[X_k|\,\mathbf{x}_i(m)\right]\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K} \|X_j\|_2^2 \middle| \mathbf{x}_i(m)\right]\right] + \mathbb{E}\left[\sum_{j=1}^{K}\sum_{k\in[K]\cap k\neq j} \|\mathbb{E}\left[X_j|\,\mathbf{x}_i(m)\right]\|_2\|\mathbb{E}\left[X_k|\,\mathbf{x}_i(m)\right]\|_2\right] \\
&\leq K\left(\frac{dM}{\delta}\right)^2 + \mathbb{E}\left[\sum_{j=1}^{K}\sum_{k\in[K]\cap k\neq j} \|\mathbb{E}\left[X_j|\,\mathbf{x}_i(m)\right]\|_2\|\mathbb{E}\left[X_k|\,\mathbf{x}_i(m)\right]\|_2\right] \\
&\leq K\left(\frac{dM}{\delta}\right)^2 + (K^2 - K)G^2 \\
&\leq K\left(\frac{dM}{\delta}\right)^2 + K^2G^2
\end{aligned}$$

where the third inequality is due to Lemmas 1 and 9.

Combining the above inequality with (35), with probability at least $1-\frac{\gamma}{nB}$, we have

$$\|\widehat{\mathbf{g}}_i(m)\|_2 \leq \left(1+\sqrt{8\ln\frac{nB}{\gamma}}\right)\frac{dM\sqrt{K}}{\delta}+KG.$$

Finally, using the union bound, we complete the proof for all $i \in V$ and $m = 1, \cdots, B$.