# Supplementary Material

## A Proof of Lemma 1

We first note that $F_t(\mathbf{y})$ is 2-strongly convex for any $t = 0, \ldots, T$, and Hazan and Kale [2012] have proved that for any -strongly convex function $f(\mathbf{x})$ over $\mathcal{K}$ and any $\mathbf{x} \in \mathcal{K}$, it holds that

$$\frac{}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2 \leq f(\mathbf{x}) - f(\mathbf{x}^*) \tag{21}$$

where $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathcal{K}} f(\mathbf{x})$.

Then, we consider the term $A = \sum_{t=1}^{T} G\|\mathbf{y}_t - \mathbf{y}_{t^0}\|_2$. If $T \leq 2d$, we have

$$A = \sum_{t=1}^{T} G\|\mathbf{y}_t - \mathbf{y}_{t^0}\|_2 \leq TGD \leq 2dGD \tag{22}$$

where the first inequality is due to Assumption 2. If $T > 2d$, we have

$$\begin{aligned} A &= \sum_{t=1}^{2d} G\|\mathbf{y}_t - \mathbf{y}_{t^0}\|_2 + \sum_{t=2d+1}^{T} G\|\mathbf{y}_t - \mathbf{y}_{t^0}\|_2 \\ &\leq 2dGD + \sum_{t=2d+1}^{T} G(\|\mathbf{y}_t - \mathbf{y}^*_t\|_2 + \|\mathbf{y}^*_t - \mathbf{y}^*_{t^0}\|_2 + \|\mathbf{y}^*_{t^0} - \mathbf{y}_{t^0}\|_2). \end{aligned} \tag{23}$$

Because of (21), for any $t \in [T + 1]$, we have

$$\|\mathbf{y}_t - \mathbf{y}^*_t\|_2 \leq \sqrt{F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}^*_t)} \leq \sqrt{}(t+2)^{-\ =2} \tag{24}$$

where the last inequality is due to $F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}^*_t) \leq \ (t+2)^{-}$ .

Moreover, for any $i \geq \ _t$, we have

$$\begin{aligned} \|\mathbf{y}^*_{\ t} - \mathbf{y}^*_i\|_2^2 &\leq F_{i-1}(\mathbf{y}^*_{\ t}) - F_{i-1}(\mathbf{y}^*_i) \\ &= F_{\ t-1}(\mathbf{y}^*_{\ t}) - F_{\ t-1}(\mathbf{y}^*_i) + \left\langle \sum_{k=\ _t}^{i-1} \mathbf{g}_{c_k}, \mathbf{y}^*_{\ t} - \mathbf{y}^*_i \right\rangle \\ &\leq \left\| \sum_{k=\ _t}^{i-1} \mathbf{g}_{c_k} \right\|_2 \|\mathbf{y}^*_{\ t} - \mathbf{y}^*_i\|_2 \\ &\leq G(i - \ _t)\|\mathbf{y}^*_{\ t} - \mathbf{y}^*_i\|_2 \end{aligned} \tag{25}$$

where the first inequality is still due to (21) and the last inequality is due to Assumption 1.

Because of $t' = t + d_t - 1 \geq t$, we have $\ _{t^0} \geq \ _t$. Then, from (25), we have

$$\|\mathbf{y}^*_{\ t} - \mathbf{y}^*_{t^0}\|_2 \leq \ G(\ _{t^0} - \ _t) = \ G \sum_{k=t}^{t^0-1} |\mathcal{F}_k|. \tag{26}$$

Then, by substituting (24) and (26) into (23), if $T > 2d$, we have

$$\begin{aligned} A &\leq 2dGD + \sum_{t=2d+1}^{T} G\left(\sqrt{}(\ _t + 2)^{-\ =2} + \ G \sum_{k=t}^{t^0-1} |\mathcal{F}_k| + \sqrt{}(\ _{t^0} + 2)^{-\ =2}\right) \\ &\leq 2dGD + \sum_{t=2d+1}^{T} 2G\sqrt{}(\ _t + 2)^{-\ =2} + \ G^2 \sum_{t=2d+1}^{T} \sum_{k=t}^{t^0-1} |\mathcal{F}_k| \\ &\leq 2dGD + \sum_{t=2d+1}^{T} 2G\sqrt{}(\ _t - 1)^{-\ =2} + \ G^2 \sum_{t=2d+1}^{T} \sum_{k=t}^{t^0-1} |\mathcal{F}_k| \end{aligned} \tag{27}$$

where the second inequality is due to $(\ _t + 2)^{-\ =2} \geq (\ _{t^0} + 2)^{-\ =2}$ for $\ _t \leq \ _{t^0}$ and $> 0$.

To bound the second term in the right side of (27), we introduce the following lemma.

**Lemma 7** *Let* $_t = 1 + \sum_{i=1}^{t-1} |\mathcal{F}_i|$ *for any* $t \in [T + d]$. *If* $T > 2d$, *for* $0 < \quad \leq 1$, *we have*

$$\sum_{t=2d+1}^{T} (\ _t - 1)^{-\ =2} \leq d + \frac{2}{2-} T^{1-\ =2}. \tag{28}$$

For the third term in the right side of (27), if $T > 2d$, we have

$$\begin{aligned}\sum_{t=2d+1}^{T} \sum_{k=t}^{t'-1} |\mathcal{F}_k| &\leq \sum_{t=1}^{T} \sum_{k=t}^{t'-1} |\mathcal{F}_k| \leq \sum_{t=1}^{T} \sum_{k=t}^{t+d_t-1} |\mathcal{F}_k| = \sum_{k=0}^{d-1} \sum_{t=1+k}^{T-k} |\mathcal{F}_t| \\ &\leq \sum_{k=0}^{d-1} \sum_{t=1}^{T+d-1} |\mathcal{F}_t| = dT\end{aligned} \tag{29}$$

where the second inequality is due to

$$t' - 1 < t' = t + d_t - 1 \leq t + d - 1.$$

By substituting (28) and (29) into (27) and combining with (22), we have

$$A \leq 2dGD + 2Gd\sqrt{\ } + \frac{4G\sqrt{\ }}{2-} T^{1-\ =2} + \quad G^2 dT. \tag{30}$$

Then, for the term $C = \sum_{t=s}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} G\|\mathbf{y}_{\ _t} - \mathbf{y}_i\|_2$, we have

$$\begin{aligned}C &= \sum_{i=\ _s}^{\ _{s+1}-1} G\|\mathbf{y}_{\ _t} - \mathbf{y}_i\|_2 + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} G\|\mathbf{y}_{\ _t} - \mathbf{y}_i\|_2 \\ &\leq |\mathcal{F}_s| GD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} G(\|\mathbf{y}_{\ _t} - \mathbf{y}^*_{\ _t}\|_2 + \|\mathbf{y}^*_{\ _t} - \mathbf{y}^*_i\|_2 + \|\mathbf{y}^*_i - \mathbf{y}_i\|_2) \\ &\leq |\mathcal{F}_s| GD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} G\ \ \sqrt{\ }(\ _t + 2)^{-\ =2} + \quad G(i - \ _t) + \sqrt{\ }(i + 2)^{-\ =2} \\ &\leq |\mathcal{F}_s| GD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} 2G\sqrt{\ }(\ _t + 2)^{-\ =2} + \quad G^2 \sum_{t=s+1}^{T+d-1} \sum_{k=0}^{\ _{t+1}-\ _t-1} k \\ &\leq |\mathcal{F}_s| GD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} 2G\sqrt{\ }(\ _t - 1)^{-\ =2} + \quad G^2 \sum_{t=s}^{T+d-1} \sum_{k=0}^{\ _{t+1}-\ _t-1} k\end{aligned} \tag{31}$$

where the first inequality is due to Assumption 2, the second inequality is due to (24) and (25), and the third inequality is due to $(\ _t + 2)^{-\ =2} \geq (i + 2)^{-\ =2}$ for $\ _t \leq i$ and $\quad > 0$.

Moreover, for any $t \in [T + d - 1]$ and $k \in \mathcal{F}_t$, since $1 \leq d_k \leq d$, we have

$$t - d + 1 \leq k = t - d_k + 1 \leq t$$

which implies that

$$|\mathcal{F}_t| \leq t - (t - d + 1) + 1 = d. \tag{32}$$

Then, it is easy to verify that

$$\ _{t+1} - \ _t - 1 < \ _{t+1} - \ _t = |\mathcal{F}_t| \leq d.$$

Therefore, by combining with (31), we have

$$\begin{aligned}C &\leq dGD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} 2G\sqrt{\ }(\ _t - 1)^{-\ =2} + \quad G^2 \sum_{t=s}^{T+d-1} \frac{|\mathcal{F}_t|^2}{2} \\ &\leq dGD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} 2G\sqrt{\ }(\ _t - 1)^{-\ =2} + \quad G^2 \sum_{t=s}^{T+d-1} \frac{d|\mathcal{F}_t|}{2} \\ &= dGD + \sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{\ _{t+1}-1} 2G\sqrt{\ }(\ _t - 1)^{-\ =2} + \frac{G^2 dT}{2}.\end{aligned} \tag{33}$$

Furthermore, we introduce the following lemma.

**Lemma 8** *Let* $_t = 1 + \sum_{i=1}^{t-1} |\mathcal{F}_i|$ *for any* $t \in [T + d]$ *and* $s = \min\{t | t \in [T + d - 1]; |\mathcal{F}_t| > 0\}$. *For* $0 < \quad \leq 1$, *we have*

$$\sum_{t=s+1}^{T+d-1} \sum_{i=\ _t}^{t-1} (\ _t - 1)^{-\ =2} \leq d + \frac{2}{2-} T^{1-\ =2}. \tag{34}$$

By substituting (34) into (33), we have

$$C \leq dGD + 2G\sqrt{d} + \frac{4G\sqrt{\ }}{2-} T^{1-\ =2} + \frac{G^2 dT}{2} \tag{35}$$

We complete the proof by combing (30) and (35).

## B Proof of Lemma 2

At the beginning of this proof, we recall the standard definition for smooth functions [Boyd and Vandenberghe, 2004].

**Definition 2** *A function* $f(\mathbf{x}) : \mathcal{K} \to \mathbb{R}$ *is called* $\ $*-smooth over* $\mathcal{K}$ *if for all* $\mathbf{x}; \mathbf{y} \in \mathcal{K}$, *it holds that* $f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}); \mathbf{y} - \mathbf{x} \rangle + \frac{\ }{2} \|\mathbf{y} - \mathbf{x}\|_2^2$.

It is not hard to verify that $F_t(\mathbf{y})$ is 2-smooth over $\mathcal{K}$ for any $t \in [T]$. This property will be utilized in the following.

For brevity, we define $h_t = F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}_t^*)$ for $t = 1; \cdots; T+1$ and $h_t(\mathbf{y}_{t-1}) = F_{t-1}(\mathbf{y}_{t-1}) - F_{t-1}(\mathbf{y}_t^*)$ for $t = 2; \cdots; T + 1$.

For $t = 1$, since $\mathbf{y}_1 = \operatorname{argmin}_{\mathbf{y} \in \mathcal{K}} \|\mathbf{y} - \mathbf{y}_1\|_2^2$, we have

$$h_1 = F_0(\mathbf{y}_1) - F_0(\mathbf{y}_1^*) = 0 \leq \frac{8D^2}{\sqrt{3}} = \frac{8D^2}{\sqrt{t+2}}. \tag{36}$$

Then, for any $T + 1 \geq t \geq 2$, we have

$$\begin{aligned} h_t(\mathbf{y}_{t-1}) =& F_{t-1}(\mathbf{y}_{t-1}) - F_{t-1}(\mathbf{y}_t^*) \\ =& F_{t-2}(\mathbf{y}_{t-1}) - F_{t-2}(\mathbf{y}_t^*) + \langle\ \mathbf{g}_{c_t\ \ 1}; \mathbf{y}_{t-1} - \mathbf{y}_t^* \rangle \\ \leq& F_{t-2}(\mathbf{y}_{t-1}) - F_{t-2}(\mathbf{y}_{t-1}^*) + \langle\ \mathbf{g}_{c_t\ \ 1}; \mathbf{y}_{t-1} - \mathbf{y}_t^* \rangle \\ \leq& h_{t-1} + \ \|\mathbf{g}_{c_t\ \ 1}\|_2 \|\mathbf{y}_{t-1} - \mathbf{y}_t^*\|_2 \\ \leq& h_{t-1} + \ \|\mathbf{g}_{c_t\ \ 1}\|_2 \|\mathbf{y}_{t-1} - \mathbf{y}_{t-1}^*\|_2 + \ \|\mathbf{g}_{c_t\ \ 1}\|_2 \|\mathbf{y}_{t-1}^* - \mathbf{y}_t^*\|_2 \\ \leq& h_{t-1} + \ G\|\mathbf{y}_{t-1} - \mathbf{y}_{t-1}^*\|_2 + \ G\|\mathbf{y}_{t-1}^* - \mathbf{y}_t^*\|_2 \end{aligned} \tag{37}$$

where the first inequality is due to $\mathbf{y}_{t-1}^* = \operatorname{argmin}_{\mathbf{y} \in \mathcal{K}} F_{t-2}(\mathbf{y})$ and the last inequality is due to Assumption 1.

Moreover, for any $T + 1 \geq t \geq 2$, we note that $F_{t-2}(\mathbf{x})$ is also 2-strongly convex, which implies that

$$\|\mathbf{y}_{t-1} - \mathbf{y}_{t-1}^*\|_2 \leq \sqrt{F_{t-2}(\mathbf{y}_{t-1}) - F_{t-2}(\mathbf{y}_{t-1}^*)} \leq \sqrt{h_{t-1}} \tag{38}$$

where the first inequality is due to (21).

Similarly, for any $T + 1 \geq t \geq 2$

$$\begin{aligned} \|\mathbf{y}_{t-1}^* - \mathbf{y}_t^*\|_2^2 \leq& F_{t-1}(\mathbf{y}_{t-1}^*) - F_{t-1}(\mathbf{y}_t^*) \\ =& F_{t-2}(\mathbf{y}_{t-1}^*) - F_{t-2}(\mathbf{y}_t^*) + \langle\ \mathbf{g}_{c_t\ \ 1}; \mathbf{y}_{t-1}^* - \mathbf{y}_t^* \rangle \\ \leq& \ \|\mathbf{g}_{c_t\ \ 1}\|_2 \|\mathbf{y}_{t-1}^* - \mathbf{y}_t^*\|_2 \end{aligned}$$

which implies that

$$\|\mathbf{y}_{t-1}^* - \mathbf{y}_t^*\|_2 \leq \ \|\mathbf{g}_{c_t\ \ 1}\|_2 \leq \ G. \tag{39}$$

By combining (37), (38), and (39), for any $T + 1 \geq t \geq 2$, we have

$$h_t(\mathbf{y}_{t-1}) \leq h_{t-1} + \ G\sqrt{h_{t-1}} + \ ^2 G^2. \tag{40}$$

Then, for any $T+1 \geq t \geq 2$, since $F_{t-1}(\mathbf{y})$ is 2-smooth, we have

$$\begin{aligned} h_t =& F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}_t^*) \\ =& F_{t-1}(\mathbf{y}_{t-1} + \ _{t-1}(\mathbf{v}_{t-1} - \mathbf{y}_{t-1})) - F_{t-1}(\mathbf{y}_t^*) \\ \leq& h_t(\mathbf{y}_{t-1}) + \langle \nabla F_{t-1}(\mathbf{y}_{t-1}); \ _{t-1}(\mathbf{v}_{t-1} - \mathbf{y}_{t-1})\rangle + \ ^2_{t-1}\|\mathbf{v}_{t-1} - \mathbf{y}_{t-1}\|_2^2. \end{aligned} \tag{41}$$

Moreover, for any $t \in [T]$, according to Algorithm 1, we have

$$_t = \operatorname*{argmin}_{\in[0;1]} \langle \ (\mathbf{v}_t - \mathbf{y}_t); \nabla F_t(\mathbf{y}_t)\rangle + \ ^2\|\mathbf{v}_t - \mathbf{y}_t\|_2^2. \tag{42}$$

Therefore, for $t = 2$, by combining (40) and (41), we have

$$\begin{aligned} h_2 \leq& h_1 + \ G\sqrt{h_1} + \ ^2G^2 + \langle \nabla F_1(\mathbf{y}_1); \ _1(\mathbf{v}_1 - \mathbf{y}_1)\rangle + \ ^2_1\|\mathbf{v}_1 - \mathbf{y}_1\|_2^2 \\ \leq& h_1 + \ G\sqrt{h_1} + \ ^2G^2 = \frac{D^2}{2(T+2)^{3=2}} \leq 4D^2 = \frac{8D^2}{\sqrt{t+2}} \end{aligned} \tag{43}$$

where the second inequality is due to (42), and the first equality is due to (36) and $= \frac{D}{\sqrt{2}G(T+2)^{3/4}}$.

Then, for any $t = 3; : : : ; T+1$, by defining $\ ^{\prime}_{t-1} = 2=\sqrt{t+1}$ and assuming $h_{t-1} \leq \frac{8D^2}{\sqrt{t+1}}$, we have

$$\begin{aligned} h_t \leq& h_t(\mathbf{y}_{t-1}) + \langle \nabla F_{t-1}(\mathbf{y}_{t-1}); \ ^{\prime}_{t-1}(\mathbf{v}_{t-1} - \mathbf{y}_{t-1})\rangle + (\ ^{\prime}_{t-1})^2\|\mathbf{v}_{t-1} - \mathbf{y}_{t-1}\|_2^2 \\ \leq& h_t(\mathbf{y}_{t-1}) + \langle \nabla F_{t-1}(\mathbf{y}_{t-1}); \ ^{\prime}_{t-1}(\mathbf{y}_t^* - \mathbf{y}_{t-1})\rangle + (\ ^{\prime}_{t-1})^2\|\mathbf{v}_{t-1} - \mathbf{y}_{t-1}\|_2^2 \\ \leq& (1 - \ ^{\prime}_{t-1})h_t(\mathbf{y}_{t-1}) + (\ ^{\prime}_{t-1})^2\|\mathbf{v}_{t-1} - \mathbf{y}_{t-1}\|_2^2 \\ \leq& (1 - \ ^{\prime}_{t-1})(h_{t-1} + \ G\sqrt{h_{t-1}} + \ ^2G^2) + (\ ^{\prime}_{t-1})^2D^2 \\ \leq& (1 - \ ^{\prime}_{t-1})h_{t-1} + \ G\sqrt{h_{t-1}} + \ ^2G^2 + (\ ^{\prime}_{t-1})^2D^2 \\ \leq& \ 1 - \frac{2}{\sqrt{t+1}} \ \ \frac{8D^2}{\sqrt{t+1}} + \frac{2D^2}{(T+2)^{3=4}(t+1)^{1=4}} + \frac{D^2}{2(T+2)^{3=2}} + \frac{4D^2}{t+1} \\ \leq& \ 1 - \frac{2}{\sqrt{t+1}} \ \ \frac{8D^2}{\sqrt{t+1}} + \frac{2D^2}{t+1} + \frac{D^2}{2(t+1)} + \frac{4D^2}{t+1} \\ \leq& \ 1 - \frac{2}{\sqrt{t+1}} \ \ \frac{8D^2}{\sqrt{t+1}} + \frac{8D^2}{t+1} \\ =& \ 1 - \frac{1}{\sqrt{t+1}} \ \ \frac{8D^2}{\sqrt{t+1}} \leq \frac{8D^2}{\sqrt{t+2}} \end{aligned} \tag{44}$$

where the first inequality is due to (41) and (42), the second inequality is due to $\mathbf{v}_{t-1} \in \operatorname{argmin}_{\mathbf{y}\in\mathcal{K}}\langle \nabla F_{t-1}(\mathbf{y}_{t-1}); \mathbf{y}\rangle$, the third inequality is due to the convexity of $F_{t-1}(\mathbf{y})$, the fourth inequality is due to (40), and the last inequality is due to

$$1 - \frac{1}{\sqrt{t+1}} \ \ \frac{1}{\sqrt{t+1}} \leq \frac{1}{\sqrt{t+2}} \tag{45}$$

for any $t \geq 0$.

Note that (45) can be derived by dividing $(t+1)\sqrt{t+2}$ into both sides of the following inequality

$$\sqrt{t+2}\sqrt{t+1} - \sqrt{t+2} \leq (\sqrt{t+1}+1)\sqrt{t+1} - \sqrt{t+2} \leq t+1+\sqrt{t+1} - \sqrt{t+2} \leq t+1.$$

By combining (36), (43), and (44), we complete this proof.

## C Proof of Lemma 3

In the beginning, we define $\mathbf{y}_t^* = \operatorname{argmin}_{\mathbf{y}\in\mathcal{K}} F_{t-1}(\mathbf{y})$ for any $t \in [T+1]$, where $F_t(\mathbf{y}) = \sum_{i=1}^{t}\langle \mathbf{g}_{c_i}; \mathbf{y}\rangle + \|\mathbf{y} - \mathbf{y}_1\|_2^2$.

Then, it is easy to verify that

$$\sum_{t=1}^{T}\langle \mathbf{g}_{c_t}; \mathbf{y}_t - \mathbf{x}^*\rangle = \sum_{t=1}^{T}\langle \mathbf{g}_{c_t}; \mathbf{y}_t - \mathbf{y}_t^*\rangle + \sum_{t=1}^{T}\langle \mathbf{g}_{c_t}; \mathbf{y}_t^* - \mathbf{x}^*\rangle. \tag{46}$$

Therefore, we will continue to upper bound the right side of (46). By applying Lemma 2, we have

$$\begin{aligned}\sum_{t=1}^{T}\langle \mathbf{g}_{c_t}, \mathbf{y}_t - \mathbf{y}_t^*\rangle &\leq \sum_{t=1}^{T}\|\mathbf{g}_{c_t}\|_2\|\mathbf{y}_t - \mathbf{y}_t^*\|_2 \leq \sum_{t=1}^{T} G\sqrt{F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}_t^*)} \\ &\leq \sum_{t=1}^{T}\frac{2\sqrt{2}GD}{(t+2)^{1/4}} \leq \frac{8\sqrt{2}GD(T+2)^{3/4}}{3}\end{aligned} \tag{47}$$

where the second inequality is due to (21) and Assumption 1, and the last inequality is due to $\sum_{t=1}^{T}(t+2)^{-1/4} \leq 4(T+2)^{3/4}/3$.

Then, to bound $\sum_{t=1}^{T}\langle \mathbf{g}_{c_t}, \mathbf{y}_t^* - \mathbf{x}^*\rangle$, we introduce the following lemma.

**Lemma 9** *(Lemma 6.6 of Garber and Hazan [2016]) Let $\{f_t(\mathbf{y})\}_{t=1}^{T}$ be a sequence of loss functions and let $\mathbf{y}_t^* \in \operatorname{argmin}_{\mathbf{y}}$*

where the first inequality is due to Assumption 2.

Then, if $T > 2d$, we have

$$
\begin{aligned}
E &= \frac{3\ D}{2} \sum_{t=1}^{2d} \|\mathbf{y}_t - \mathbf{y}^*_t\|_2 + \frac{3\ D}{2} \sum_{t=2d+1}^{T} \|\mathbf{y}_t - \mathbf{y}^*_t\|_2 \\
&\leq 3\ dD^2 + \frac{3\ D}{2} \sum_{t=2d+1}^{T} \|\mathbf{y}_t - \mathbf{y}^*_t\|_2 + \|\mathbf{y}^*_t - \mathbf{y}^*_t\|_2 + \|\mathbf{y}^*_t - \mathbf{y}^*_t\|_2 \ .
\end{aligned} \tag{50}
$$

Because $F_{t-1}(\mathbf{y})$ is $(t-1)$ -strongly convex for any $t = 2, \ldots, T + 1$, we have

$$
\|\mathbf{y}_t - \mathbf{y}^*_t\|_2 \leq \sqrt{\frac{2(F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}^*_t))}{(t-1)}} \leq \sqrt{\frac{2}{(t-1)^{1-}}} \tag{51}
$$

where the first inequality is due to (21) and the second inequality is due to $F_{t-1}(\mathbf{y}_t) - F_{t-1}(\mathbf{y}^*_t) \leq (t-1)$ .

Before considering $\|\mathbf{y}^*_t - \mathbf{y}^*_t\|_2$, we define $f_t(\mathbf{y}) = \langle \mathbf{g}_{c_t}, \mathbf{y} \rangle + \frac{}{2}\|\mathbf{y} - \mathbf{y}_t\|_2^2$ for any $t = 1, \ldots, T$. Note that $F_t(\mathbf{y}) = \sum_{i=1}^{t} f_i(\mathbf{y})$. Moreover, for any $\mathbf{x}, \mathbf{y} \in \mathcal{K}$ and $t = 1, \ldots, T$, we have

$$
\begin{aligned}
|f_t(\mathbf{x}) - f_t(\mathbf{y})| &= \langle \mathbf{g}_{c_t}, \mathbf{x} - \mathbf{y} \rangle + \frac{}{2}\|\mathbf{x} - \mathbf{y}_t\|_2^2 - \frac{}{2}\|\mathbf{y} - \mathbf{y}_t\|_2^2 \\
&= \mathrm{k}_2^2
\end{aligned}
$$

where the second inequality is due to $(\ _t - 1)^{1-\ } \leq (t-1)^{1-\ }$ for $t \geq\ _t > 1$ and $\ < 1$, and the last inequality is due to Lemma 7 and $0 < 1 - \ \leq 1$.

By combining (49) with the above inequality, we have

$$E \leq 3\ \ dD^2 + 3dD\sqrt{2} + \frac{6D\sqrt{2}}{1+}T^{(1+\ )=2} + 3D(G+\ \ D)d\ln T:$$

Then, we proceed to bound the term $C = \sum_{t=s}^{T+d-1}\sum_{i=\ _t}^{\ _{t+1}-1} G\|\mathbf{y}_{\ _t} - \mathbf{y}_i\|_2$. Similar to (31), we first have

$$C \leq |\mathcal{F}_s|GD + \sum_{t=s+1}^{T+d-1}\sum_{i=\ _t}^{\ _{t+1}-1} G(\|\mathbf{y}_{\ _t} - \mathbf{y}^*_{\ _t}\|_2 + \|\mathbf{y}^*_{\ _t} - \mathbf{y}^*_i\|_2 + \|\mathbf{y}^*_i - \mathbf{y}_i\|_2): \tag{55}$$

By combining (55) with $|\mathcal{F}_s| \leq d$, (51), and (53), we have

$$\begin{aligned} C \leq & dGD + \sum_{t=s+1}^{T+d-1}\sum_{i=\ _t}^{\ _{t+1}-1} G\left(\sqrt{\frac{2}{(\ _t-1)^{1-\ }}} + \frac{2(i-\ _t)(G+\ \ D)}{(i-1)} + \sqrt{\frac{2}{(i-1)^{1-\ }}}\right) \\ \leq & dGD + \sum_{t=s+1}^{T+d-1}\sum_{i=\ _t}^{\ _{t+1}-1} G\ \ 2\left(\sqrt{\frac{2}{(\ _t-1)^{1-\ }}} + \frac{2(i-\ _t)(G+\ \ D)}{(i-1)}\right) \\ \leq & dGD + 2dG\sqrt{\frac{2}{\ }} + \sqrt{\frac{2}{\ }}\frac{4G}{1+}T^{(1+\ )=2} + \sum_{t=s+1}^{T+d-1}\sum_{i=\ _t}^{\ _{t+1}-1}\frac{2dG(G+\ \ D)}{(i-1)} \end{aligned} \tag{56}$$

where the first inequality is due to $(\ _t - 1)^{1-\ } \leq (i-1)^{1-\ }$ for $0 <\ _t - 1 \leq i-1$ and the last inequality is due to Lemma 8, $0 < 1 - \ \leq 1$, and $i -\ _t \leq\ _{t+1} - 1 -\ _t \leq |\mathcal{F}_t| \leq$

Recall that we have defined

$$\mathcal{I}_t = \begin{cases} \emptyset; \text{ if } |\mathcal{F}_t| = 0; \\ \{\ _t;\ _t + 1; \cdots;\ _{t+1} - 1\}; \text{ otherwise:} \end{cases}$$

It is not hard to verify that

$$\cup_{t=s+1}^{T+d-1}\mathcal{I}_t = \{|F_s| + 1; \cdots; T\}; \mathcal{I}_i \cap \mathcal{I}_j = \emptyset; \forall i \neq j: \tag{57}$$

By combining (57) with (56), we have

$$\begin{aligned} C \leq & dGD + 2dG\sqrt{\frac{2}{\ }} + \sqrt{\frac{2}{\ }}\frac{4G}{1+}T^{(1+\ )=2} + \sum_{t=|F_s|+1}^{T}\frac{2dG(G+\ \ D)}{(t-1)} \\ \leq & dGD + 2dG\sqrt{\frac{2}{\ }} + \sqrt{\frac{2}{\ }}\frac{4G}{1+}T^{(1+\ )=2} + \sum_{t=2}^{T}\frac{2dG(G+\ \ D)}{(t-1)} \\ \leq & dGD + 2dG\sqrt{\frac{2}{\ }} + \sqrt{\frac{2}{\ }}\frac{4G}{1+}T^{(1+\ )=2} + \frac{2dG(G+\ \ D)(1+\ln T)}{\ }: \end{aligned} \tag{58}$$

Next, we proceed to bound the term $A = \sum_{t=1}^{T} G\|\mathbf{y}_{\ _t} - \mathbf{y}_{\ _{t^0}}\|_2$. Similar to (23), if $T > 2d$, we have

$$\begin{aligned} A \leq & 2dGD + \sum_{t=2d+1}^{T} G(\|\mathbf{y}_{\ _t} - \mathbf{y}^*_{\ _t}\|_2 + \|\mathbf{y}^*_{\ _t} - \mathbf{y}^*_{\ _{t^0}}\|_2 + \|\mathbf{y}^*_{\ _{t^0}} - \mathbf{y}_{\ _{t^0}}\|_2) \\ \leq & 2dGD + \sum_{t=2d+1}^{T} G\left(\sqrt{\frac{2}{(\ _t-1)^{1-\ }}} + \frac{2(\ _{t^0} -\ _t)(G+\ \ D)}{(\ _{t^0}-1)} + \sqrt{\frac{2}{\ }}\right. \end{aligned}$$

**Lemma 10** *Let* $h_k = \sum_{i=1}^{k} |\mathcal{F}_i|$. *If* $T > 2d$, *we have*

$$\sum_{t=2d+1}^{T} \sum_{k=t}^{T-1} \frac{|\mathcal{F}_k|}{h_k} \leq d + d \ln T.$$

By applying Lemmas 7 and 10 to (59) and combining with (22), we have

$$A \leq 2dGD + 2dG \sqrt{\frac{2}{\ }} + \sqrt{\frac{2}{\ }} \frac{4G}{1+\ } T^{(1+\ )/2} + \frac{2G(G+\ D)d(1+\ln T)}{\ }. \tag{60}$$

Finally, by combining (58) and (60), we complete this proof.

## E Proof of Lemmas 5 and 6

Recall that $F$

k

where the first inequality is due to $(\ _t - 1)^{-\ \ /2} \leq 1$ for $\ > 0$ and $\ _t \geq 2$, and the second inequality is due to (61) and $\ > 0$.

## G Proof of Lemma 8

Because of $\ _t = 1 + \sum_{i=1}^{t-1} |\mathcal{F}_i|$, we have

$$\sum_{t=s+1}^{T+\cancel{X}-1} \ _t$$