A. Proofs of Theorems 4.2 and 4.7

A.1. Proof of Theorem 4.2

We need the following lemma to guarantee an $\Omega(n)$ lower bound for finding an ϵ-suboptimal solution when F is convex.

Lemma A.1. For any linear-span randomized first-order algorithm $\mathcal{A}$ and any $L, \sigma, n, \Delta, \epsilon$ with $\epsilon < \Delta/4$, there exist functions $\{f_i\}_{i=1}^n : \mathbb{R}^n \to \mathbb{R}$ and $F = \sum_{i=1}^n f_i/n$ which satisfy that $\{f_i\}_{i=1}^n \in \mathcal{V}^{(L)}$, $F \in \mathcal{S}^{(0,L)}$ and $F(\mathbf{x}^{(0)}) - \inf_{\mathbf{x}\in\mathbb{R}^n} F(\mathbf{x}) \le \Delta$. In order to find $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that $\mathbb{E}F(\widehat{\mathbf{x}}) - \inf_{\mathbf{x}\in\mathbb{R}^n} F(\mathbf{x}) \le \epsilon$, $\mathcal{A}$ needs at least $\Omega(n)$ IFO calls.

Proof of Theorem 4.2. Let $\{\mathbf{U}^{(i)}\}_{i=1}^n \in \mathcal{O}(2T-1, (2T-1)n, n)$. We choose $\bar{f}_i(\mathbf{x}) : \mathbb{R}^{Tn} \to \mathbb{R}$ as follows:

$$\bar{f}_i(\mathbf{x}) := \sqrt{n} f_{\mathcal{N}\mathrm{c}}(\mathbf{U}^{(i)}\mathbf{x}; \alpha, T),$$
$$\bar{F}(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^n \bar{f}_i(\mathbf{x}).$$

We have the following properties. First, we claim that $\{\bar{f}_i(\mathbf{x})\} \in \mathcal{V}^{(1)}$ because of Lemma 5.1 where $f_{\mathcal{N}\mathrm{c}} \in \mathcal{S}^{(0,1)} \subset \mathcal{S}^{(-1,1)}$. Next, suppose that $\bar{\mathcal{X}}^* = \mathrm{argmin}_{\mathbf{z}} \bar{F}(\mathbf{z})$, then by definition, we have that for any $\bar{\mathbf{x}}^* \in \bar{\mathcal{X}}^*$, $\mathbf{U}^{(i)}\bar{\mathbf{x}}^* \in (\mathcal{X}^*)^{(i)}$, where $(\mathcal{X}^*)^{(i)} = \mathrm{argmin}_{\mathbf{z}} f_{\mathcal{N}\mathrm{c}}(\mathbf{z}; \alpha, T)$. Thus, we have

$$\mathrm{dist}^2(0, \bar{\mathcal{X}}^*) = \inf_{\bar{\mathbf{x}}^*\in\bar{\mathcal{X}}^*} \|0 - \bar{\mathbf{x}}^*\|_2^2 = \inf_{\bar{\mathbf{x}}^*\in\bar{\mathcal{X}}^*} \sum_{i=1}^n \|\mathbf{U}^{(i)}\bar{\mathbf{x}}^*\|_2^2 \le \frac{2nT}{3} \le nT.$$

Finally, let $\mathbf{y}^{(i)} = \mathbf{U}^{(i)}\mathbf{x} \in \mathbb{R}^T$. If there exists $\mathcal{I} \subset [n], |\mathcal{I}| > n/2$ and for each $i \in \mathcal{I}$, $\mathbf{y}_T^{(i)} = ... = \mathbf{y}_{2T-1}^{(i)} = 0$, then by Proposition 3.9, we have

$$\begin{aligned}
&\bar{F}(\mathbf{x}) - \inf_{\mathbf{z}} \bar{F}(\mathbf{z}) \\
&\ge \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} [f_{\mathcal{N}\mathrm{sc}}(\mathbf{y}^{(i)}, \alpha, T) - \inf_{\mathbf{z}} f_{\mathcal{N}\mathrm{sc}}(\mathbf{z}, \alpha, T)] \\
&\ge \sqrt{n}/(16T).
\end{aligned} \tag{A.1}$$

With above properties, we set the final functions as $f_i(\mathbf{x}) = \lambda \bar{f}_i(\mathbf{x}/\beta)$. We first consider any fixed index sequence $\{i_t\}$. For the case $\epsilon \le LB^2/(16\sqrt{n})$, we set λ, β, T as

$$\lambda = \frac{B\sqrt{16\epsilon L}}{n^{3/4}}, \beta = \sqrt{\lambda/L}, T = \frac{B\sqrt{L}}{4n^{1/4}\epsilon^{1/2}},$$

Since Then by Lemma 5.2, we have that $f_i \in \mathcal{V}^{(L)}$, $F \in \mathcal{S}^{(0,L)}$, $F(0) - \inf_{\mathbf{z}} F(\mathbf{z}) \le \Delta$. By Proposition 3.5, we know that for any algorithm output $\mathbf{x}^{(t)}$ where t is less than

$$\frac{nT}{2} = 8n^{3/4}B\sqrt{\frac{L}{\epsilon}}, \tag{A.2}$$

there exists $\mathcal{I} \subset [n], |\mathcal{I}| > n - nT/(2T) = n/2$ and for each $i \in \mathcal{I}$, $\mathbf{y}_T^{(i)} = ... = \mathbf{y}_{2T-1}^{(i)} = 0$, where $\mathbf{y}^{(i)} = \mathbf{U}^{(i)}\mathbf{x}^{(t)}$. Thus, $\mathbf{x}^{(t)}$ satisfies that

$$\bar{F}(\mathbf{x}^{(t)}) - \inf_{\mathbf{z}} \bar{F}(\mathbf{z}) \ge \lambda\sqrt{n}/(16T) \ge \epsilon,$$

where the first inequality holds due to (A.1). Then, applying Yao's minimax theorem, we have that for any randomized index sequence $\{i_t\}$, we have the lower bound (A.2). For the case $LB^2/4 \ge \epsilon \ge LB^2/(16\sqrt{n})$, by Lemma A.1 we know that there exists an $\Omega(n)$ lower bound. Thus, with all above statements, we have the lower bound (4.2). $\square$

A.2. Proof of Theorem 4.7

Proof of Theorem 4.7. Let $\{\mathbf{U}^{(i)}\}_{i=1}^n \in \mathcal{O}(T+1,(T+1)n,n)$. We choose $\bar{f}_i(\mathbf{x}): \mathbb{R}^{(T+1)n} \to \mathbb{R}$ as follows:

$$\bar{f}_i(\mathbf{x}) := Q(\mathbf{U}^{(i)}\mathbf{x}; \sqrt{\alpha}, T+1, 0) + \frac{\alpha}{n}\Gamma(\mathbf{U}\mathbf{x}),$$
$$\bar{F}(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^n \bar{f}_i(\mathbf{x}).$$

We have the following properties. First, we claim that each $\bar{f}_i \in \mathcal{S}^{(-\alpha c_\gamma/n, 4+\alpha c_\gamma/n)}$ because $Q \in \mathcal{S}^{(0,4)}$ and $\Gamma \in \mathcal{S}^{(-c_\gamma, c_\gamma)}$. Next, note that

$$\begin{aligned}\bar{F}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \bar{f}_i(\mathbf{x})\\ &= \frac{1}{n}\sum_{i=1}^n [Q(\mathbf{U}^{(i)}\mathbf{x}; \sqrt{\alpha}, T+1, 0) + \alpha\Gamma(\mathbf{U}^{(i)}\mathbf{x})]\\ &= \frac{1}{n}\sum_{i=1}^n f_{\mathcal{C}}(\mathbf{U}^{(i)}\mathbf{x}; \sqrt{\alpha}, T+1).\end{aligned}$$

Then we have

$$\begin{aligned}&\bar{F}(0) - \inf_{\mathbf{x}} \bar{F}(\mathbf{x})\\ &= \frac{1}{n}\sum_{i=1}^n f_{\mathcal{C}}(0; \sqrt{\alpha}, T+1) - \inf_{\mathbf{x}} \frac{1}{n}\sum_{i=1}^n f_{\mathcal{C}}(\mathbf{U}^{(i)}\mathbf{x}; \sqrt{\alpha}, T+1)\\ &= \frac{1}{n}\sum_{i=1}^n [f_{\mathcal{C}}(0; \sqrt{\alpha}, T+1) - \inf_{\mathbf{x}} f_{\mathcal{C}}(\mathbf{x}; \sqrt{\alpha}, T+1)]\\ &\le \sqrt{\alpha} + 10\alpha T,\end{aligned}$$

where the second equality holds due to the fact that $\inf_{\mathbf{x}} \sum_{i=1}^n f_{\mathcal{C}}(\mathbf{U}^{(i)}\mathbf{x}; \alpha, T) = \sum_{i=1}^n \inf_{\mathbf{x}} f_{\mathcal{C}}(\mathbf{x}; \alpha, T)$. Finally, let $\mathbf{y}^{(i)} = \mathbf{U}^{(i)}\mathbf{x}$. If there exists $\mathcal{I}, |\mathcal{I}| > n/2$ and for each $i \in \mathcal{I}$, $\mathbf{y}_T^{(i)} = \mathbf{y}_{T+1}^{(i)} = 0$, then by Proposition 3.11, we have

$$\begin{aligned}\|\nabla \bar{F}(\mathbf{x})\|_2^2 &\ge \frac{1}{n^2}\sum_{i\in\mathcal{I}} \|\mathbf{U}^{(i)}\nabla[f_{\mathcal{C}}(\mathbf{U}^{(i)}\mathbf{x}; \alpha, T)]\|_2^2\\ &\ge \frac{1}{n^2}\frac{n}{2}(\alpha^{3/4}/4)^2\\ &= \alpha^{3/2}/(32n).\end{aligned} \tag{A.3}$$

With above properties, we set the final functions $f_i(\mathbf{x}) = \lambda \bar{f}_i(\mathbf{x}/\beta)$. We first consider any fixed index sequence $\{i_t\}$. We set $\alpha, \lambda, \beta, T$ as

$$\begin{aligned}\alpha &= \min\left\{1, \frac{5n\sigma}{c_\gamma L}\right\}\\ \lambda &= \frac{160n\epsilon^2}{L\alpha^{3/2}}\\ \beta &= \sqrt{5\lambda/L}\\ T &= \frac{\Delta L}{1760n\epsilon^2}\sqrt{\min\left\{1, \frac{5n\sigma}{c_\gamma L}\right\}},\end{aligned}$$

Then by Lemma 5.2, we have that $f_i \in \mathcal{S}^{(-\sigma, L)}$, $F(0) - \inf_{\mathbf{z}} F(\mathbf{z}) \le \Delta$ with the assumption that $\epsilon^2 \le \Delta L\alpha/(1760n)$. By Proposition 3.5, we know that for any algorithm output $\mathbf{x}^t$ where t is less than

$$\frac{nT}{2} = \frac{\Delta L}{3520\epsilon^2}\sqrt{\min\left\{1, \frac{5n\sigma}{c_\gamma L}\right\}}, \tag{A.4}$$

there exists $\mathcal{I} \subset [n], |\mathcal{I}| > n - nT/(2T) = n/2$ and for each i, $\mathbf{y}_T^{(i)} = \mathbf{y}_{T+1}^{(i)} = 0$ where $\mathbf{y}^{(i)} = \mathbf{U}^{(i)}\mathbf{x}^{(t)}$. Thus, by (A.3), $\mathbf{x}^{(t)}$ satisfies that

$$\|\nabla F(\mathbf{x}^{(t)})\|_2 \geq \lambda/\beta \cdot \sqrt{\alpha^{3/2}/(32n)} \geq \epsilon.$$

Applying Yao's minimax theorem, we have that for any randomized index sequence $\{i_t\}$, we have the lower bound (A.4), which implies (4.4).

□

B. Proofs of Technical Lemmas

B.1. Proof of Lemma 5.1

Proof of Lemma 5.1. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{mn}$, we have that

$$\begin{aligned}
\mathbb{E}_i\|\nabla\bar{g}_i(\mathbf{x}) - \nabla\bar{g}_i(\mathbf{y})\|_2^2 &= \frac{1}{n}\sum_{i=1}^n \|\nabla[\sqrt{n}g(\mathbf{U}^{(i)}\mathbf{x})] - \nabla[\sqrt{n}g(\mathbf{U}^{(i)}\mathbf{y})\|_2^2] \\
&= \sum_{i=1}^n \|[\mathbf{U}^{(i)}]^\top \nabla g(\mathbf{U}^{(i)}\mathbf{x}) - [\mathbf{U}^{(i)}]^\top \nabla g(\mathbf{U}^{(i)}\mathbf{y})\|_2^2 \\
&= \sum_{i=1}^n \|\nabla g(\mathbf{U}^{(i)}\mathbf{x}) - \nabla g(\mathbf{U}^{(i)}\mathbf{y})\|_2^2 \\
&\leq \beta^2 \sum_{i=1}^n \|\mathbf{U}^{(i)}\mathbf{x} - \mathbf{U}^{(i)}\mathbf{y}\|_2^2 \\
&= \beta^2\|\mathbf{x} - \mathbf{y}\|_2^2,
\end{aligned}$$

where the third and last equality holds due to the fact that $\mathbf{U}^{(i)}[\mathbf{U}^{(i)}]^\top = \mathbf{I}$ and $\mathbf{U}^{(i)}[\mathbf{U}^{(j)}]^\top = \mathbf{0}$ for each $i \neq j$, and the inequality holds due to the fact that $g \in \mathcal{S}^{(-\zeta,\zeta)}$. Thus, we have $\{\bar{g}_i\}_{i=1}^n \in \mathcal{V}^{(\zeta)}$. To prove $\bar{G} \in \mathcal{S}^{(\xi/\sqrt{n},\zeta)}$, we have

$$\nabla^2\bar{G}(\mathbf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{U}^{(i)}(\mathbf{U}^{(i)})^\top \nabla^2 g(\mathbf{U}^{(i)}\mathbf{x}) \succeq \frac{\xi}{\sqrt{n}}\mathbf{I},$$

where the inequality holds due to the assumption that $g \in \mathcal{S}^{(\xi,\beta)}$. With this fact and $\|\nabla\bar{G}(\mathbf{x}) - \nabla\bar{G}(\mathbf{y})\|_2^2 \leq \mathbb{E}_i\|\nabla\bar{g}_i(\mathbf{x}) - \nabla\bar{g}_i(\mathbf{y})\|_2^2 \leq \beta^2\|\mathbf{x} - \mathbf{y}\|_2^2$ which implies that $\nabla^2\bar{G}(\mathbf{x}) \preceq \beta\mathbf{I}$, we conclude that $\bar{G} \in \mathcal{S}^{(\xi/\sqrt{n},\beta)}$. □

B.2. Proof of Lemma 5.2

Proof of Lemma 5.2. First we have $\{g_i\}_{i=1}^n \in \mathcal{V}^{(\lambda/\beta^2 \cdot L')}$ because for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\begin{aligned}
\mathbb{E}_i\|\nabla g_i(\mathbf{x}) - \nabla g_i(\mathbf{y})\|_2^2 &= \lambda^2\mathbb{E}_i\|\nabla\bar{g}_i(\mathbf{x}/\beta)/\beta - \nabla\bar{g}_i(\mathbf{y}/\beta)/\beta\|_2^2 \\
&\leq \lambda^2/\beta^2(L')^2\|\mathbf{x}/\beta - \mathbf{y}/\beta\|_2^2 \\
&= (\lambda/\beta^2 \cdot L')^2\|\mathbf{x} - \mathbf{y}\|_2^2.
\end{aligned}$$

Next we have $g_i \in \mathcal{S}^{(\lambda/\beta^2\cdot\xi',\lambda/\beta^2\cdot\zeta')}$ because $\nabla^2 g_i(\mathbf{x}) = \lambda/\beta^2\nabla^2\bar{g}_i(\mathbf{x}/\beta)$ and for any $\mathbf{x} \in \mathbb{R}^d$,

$$\lambda/\beta^2 \cdot \xi'\mathbf{I} \preceq \lambda/\beta^2\nabla^2\bar{g}_i(\mathbf{x}/\beta) \preceq \lambda/\beta^2 \cdot \zeta'\mathbf{I}.$$

Next we have $G(0) - \inf_{\mathbf{x}} G(\mathbf{x}) \leq \lambda\Delta'$ because

$$G(0) - \inf_{\mathbf{x}} G(\mathbf{x}) = \lambda\bar{G}(0) - \lambda\inf_{\mathbf{x}} G(\mathbf{x}) \leq \lambda\Delta'.$$

Finally we have $\mathrm{dist}(0, (\mathbf{Z}')^*) \leq \beta B'$ because $(\mathbf{Z}')^* = \beta \cdot \mathbf{Z}^*$. □

B.3. Proof of Lemma 5.3

Proof of Lemma 5.3. Suppose the initial point $\mathbf{x}^{(0)} = \mathbf{0}$. Consider the following function $\{\bar{f}_i\}_{i=1}^n$, $\bar{f}_i : \mathbb{R}^n \to \mathbb{R}$, where

$$\bar{f}_i(\mathbf{x}) := -\sqrt{n}\langle \mathbf{x}, \mathbf{e}^{(i)} \rangle + \frac{\|\mathbf{x}\|_2^2}{2},$$
$$\bar{F}(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^n \bar{f}_i(\mathbf{x}),$$

$\mathbf{e}^{(i)}$ is the i-th coordinate vector. We have that $\{\bar{f}_i\}_{i=1}^n \in \mathcal{V}^{(1)}$ and the global minimizer of $\bar{F}$ is

$$\mathbf{x}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{e}^{(i)}.$$

Thus we have $\text{dist}(\mathbf{0}, \mathbf{x}^*) = 1$ and $\bar{F}(0) - \inf_{\mathbf{x}} \bar{F}(\mathbf{x}) = 1/2$. Moreover, if point $\mathbf{x}$ satisfies that $|\text{supp}\{\mathbf{x}\}| \leq n/2$, then

$$\bar{F}(\mathbf{x}) = \frac{\|\mathbf{x}\|_2^2}{2} - \frac{1}{\sqrt{n}}\sum_{i=1}^n \langle \mathbf{x}, \mathbf{e}^{(i)} \rangle = \frac{\|\mathbf{x}\|_2^2}{2} - \frac{1}{\sqrt{n}}\sum_{i \in \text{supp}\{\mathbf{x}\}} \langle \mathbf{x}, \mathbf{e}^{(i)} \rangle \geq -1/4,$$

which implies

$$\bar{F}(\mathbf{x}) - \inf_{\mathbf{x}} F(\mathbf{x}) \geq 1/4. \tag{B.1}$$

Next we choose $f_i = \lambda \bar{f}_i(\mathbf{x}/\beta)$, where $\lambda = 2\Delta, \beta = \sqrt{2\Delta/L}$, then we can check that $\{f_i\}_{i=1}^n \in \mathcal{V}^{(L)}$, $F(0) - \inf_{\mathbf{x}} F(\mathbf{x}) \leq \Delta$, $F \in \mathcal{S}^{(L,L)} \subset \mathcal{S}^{(\sigma,L)}$. Moreover, since $\nabla f_i(\mathbf{x}) = -\lambda\sqrt{n}\mathbf{e}^{(i)}/\beta + \lambda\mathbf{x}/\beta^2$, then for some $\mathbf{x}$, i is in the support set of $\mathbf{x}$ only if f_i has been called. Thus, if less than $n/2$ IFO calls have been made, then current point $\mathbf{x}$ satisfies that $|\text{supp}\{\mathbf{x}\}| \leq n/2$. With (B.1), we have that $F(\mathbf{x}) - \inf_{\mathbf{z}} F(\mathbf{z}) \geq \Delta/4 \geq \epsilon$. □

B.4. Proof of Lemma A.1

Proof of Lemma A.1. Suppose the initial point $\mathbf{x}^{(0)} = \mathbf{0}$. Consider the following function $\{\bar{f}_i\}_{i=1}^n$, $\bar{f}_i : \mathbb{R}^n \to \mathbb{R}$, where

$$\bar{f}_i(\mathbf{x}) := -\sqrt{n}\langle \mathbf{x}, \mathbf{e}^{(i)} \rangle + \frac{\|\mathbf{x}\|_2^2}{2},$$
$$\bar{F}(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^n \bar{f}_i(\mathbf{x}),$$

$\mathbf{e}^{(i)}$ is the i-th coordinate vector. Then by the proof of Lemma 5.3, we know that $\{\bar{f}_i\}_{i=1}^n \in \mathcal{V}^{(1)}$, $\text{dist}(\mathbf{0}, \bar{\mathbf{x}}^*) = 1$ where $\bar{\mathbf{x}}^*$ is the global minimizer of $\bar{F}$ and for any $\mathbf{x}$ satisfying $|\text{supp}\{\mathbf{x}\}| \leq n/2$,

$$\bar{F}(\mathbf{x}) - \inf_{\mathbf{x}} F(\mathbf{x}) \geq 1/4. \tag{B.2}$$

Next we choose $f_i = \lambda \bar{f}_i(\mathbf{x}/\beta)$, where $\lambda = LB^2, \beta = B$, then we can check that $\{f_i\}_{i=1}^n \in \mathcal{V}^{(L)}$, $\text{dist}(\mathbf{0}, \mathbf{x}^*) = B$ where $\mathbf{x}^*$ is the global minimizer of F, $F \in \mathcal{S}^{(L,L)} \subset \mathcal{S}^{(0,L)}$. Moreover, since $\nabla f_i(\mathbf{x}) = -\lambda\sqrt{n}\mathbf{e}^{(i)}/\beta + \lambda\mathbf{x}/\beta^2$, then for some $\mathbf{x}$, i is in the support set of $\mathbf{x}$ only if f_i has been called. Thus, if less than $n/2$ IFO calls have been made, then current point $\mathbf{x}$ satisfies that $|\text{supp}\{\mathbf{x}\}| \leq n/2$. With (B.1), we have that $F(\mathbf{x}) - \inf_{\mathbf{z}} F(\mathbf{z}) \geq \lambda/4 \geq \epsilon$. □