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 $\epsilon$-suboptimal solution when $F$ is convex.

**Lemma A.1.** For any linear-span randomized first-order algorithm $A$ 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 $\{f_i\}_{i=1}^n \in \mathcal{Y}^{(L)}$, $F \in \mathcal{S}^{(0, L)}$ and $F(x_i^0) - \inf_{x \in \mathbb{R}^n} F(x) \leq \Delta$. In order to find $\hat{x} \in \mathbb{R}^n$ such that $\mathbb{E}F(\hat{x}) - \inf_{x \in \mathbb{R}^n} F(x) \leq \epsilon$, $A$ needs at least $\Omega(n)$ IFO calls.

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

$$\bar{f}_i(x) := \sqrt{n}f_{Nc}(U^{(i)}x; \alpha, T),$$

$$\bar{F}(x) := \frac{1}{n} \sum_{i=1}^{n} \bar{f}_i(x).$$

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

$$\text{dist}^2(0, \bar{x}^*) = \inf_{x^* \in \bar{X}^*} \|0 - x^*\|^2 = \inf_{x^* \in \bar{X}^*} \sum_{i=1}^{n} \|U^{(j)}x^*\|^2 \leq \frac{2nT}{3} \leq nT.$$

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

$$\bar{F}(x) - \inf_{z} \bar{F}(z) \geq \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}} [f_{Nc}(y^{(i)}, \alpha, T) - \inf_{z} f_{Nc}(z, \alpha, T)] \geq \sqrt{n}/(16T).$$

(A.1)

With above properties, we set the final functions as $f_i(x) = \lambda \bar{f}_i(x/\beta)$. We first consider any fixed index sequence $\{i_0\}$. For the case $\epsilon \leq LB^2/(16\sqrt{n})$, we set $\lambda, \beta, T$ 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 $f_i \in \mathcal{Y}^{(L)}$, $F \in \mathcal{S}^{(0, L)}$, $F(0) - \inf_{z} F(z) \leq \Delta$. By Proposition 3.5, we know that for any algorithm output $x^{(t)}$ where $t$ is less than

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

(A.2)

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

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

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 $LB^2/4 \geq \epsilon \geq LB^2/(16\sqrt{n})$, 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).
A.2. Proof of Theorem 4.7

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

\[
\tilde{f}_{i}(x) = Q(U^{(i)}x; \sqrt{\alpha}, T + 1, 0) + \frac{\alpha}{n} \Gamma(Ux),
\]

\[
\tilde{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_{i}(x).
\]

We have the following properties. First, we claim that each \( \tilde{f}_{i} \in S^{(-\alpha c_{s}/n, 4 + \alpha c_{s}/n)} \) because \( Q \in S^{(0.4)} \) and \( \Gamma \in S^{(-c_{s}, c_{s})} \). Next, note that

\[
\tilde{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_{i}(x)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} [Q(U^{(i)}x; \sqrt{\alpha}, T + 1, 0) + \alpha \Gamma(U^{(i)}x)]
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} f_{c}(U^{(i)}x; \sqrt{\alpha}, T + 1).
\]

Then we have

\[
\tilde{F}(0) - \inf_{x} \tilde{F}(x)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} f_{c}(0; \sqrt{\alpha}, T + 1) - \inf_{x} \frac{1}{n} \sum_{i=1}^{n} f_{c}(U^{(i)}x; \sqrt{\alpha}, T + 1)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} [f_{c}(0; \sqrt{\alpha}, T + 1) - \inf_{x} f_{c}(x; \sqrt{\alpha}, T + 1)]
\]

\[
\leq \sqrt{\alpha} + 10\alpha T,
\]

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

\[
\| \nabla \tilde{F}(x) \|_{2}^{2} \geq \frac{1}{n^{2}} \sum_{i \in I} \| U^{(i)} \nabla f_{c}(U^{(i)}x; \alpha, T) \|_{2}^{2}
\]

\[
\geq \frac{1}{n^{2}} \frac{n}{2} \left( \frac{\alpha^{3/4}}{4} \right)^{2}
\]

\[
= \alpha^{3/2}/(32n).
\]

(A.3)

With above properties, we set the final functions \( f_{i}(x) = \lambda \tilde{f}_{i}(x/\beta) \). We first consider any fixed index sequence \( \{ i_{t} \} \). We set \( \alpha, \lambda, \beta, T \) as

\[
\alpha = \min \left\{ 1, \frac{5n\sigma}{c_{s}L} \right\}
\]

\[
\lambda = \frac{160nc^{2}}{L\alpha^{3/2}}
\]

\[
\beta = \sqrt{5\lambda / L}
\]

\[
T = \frac{\Delta L}{1760nc^{2}} \sqrt{\min \left\{ 1, \frac{5n\sigma}{c_{s}L} \right\}}.
\]

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

\[
\frac{nT}{2} = \frac{\Delta L}{3520\epsilon^{2}} \sqrt{\min \left\{ 1, \frac{5n\sigma}{c_{s}L} \right\}},
\]

(A.4)
there exists $\mathcal{I} \subset [n], |\mathcal{I}| > n - nT/(2T) = n/2$ and for each $i$, $y^{(i)}_T = y^{(i)}_{T+1} = 0$ where $y^{(i)} = U^{(i)}x^{(i)}$. Thus, by (A.3), $x^{(i)}$ satisfies that
\[
\|\nabla F(x^{(i)})\|_2 \geq \lambda/\beta \cdot \sqrt{n^{3/2}/(32n)} \geq \epsilon.
\]
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 $x, y \in \mathbb{R}^{mn}$, we have that
\[
\mathbb{E}_r \|\nabla \hat{g}_i(x) - \nabla \hat{g}_i(y)\|_2^2 = \frac{1}{n} \sum_{i=1}^n \|\nabla[\sqrt{n}g(U^{(i)}x)] - \nabla[\sqrt{n}g(U^{(i)}y)]\|_2^2
\]
\[
= \sum_{i=1}^n \|U^{(i)}\top \nabla g(U^{(i)}x) - [U^{(i)}] \nabla g(U^{(i)}y)\|_2^2
\]
\[
= \sum_{i=1}^n \|\nabla g(U^{(i)}x) - \nabla g(U^{(i)}y)\|_2^2
\]
\[
\leq \beta^2 \sum_{i=1}^n \|U^{(i)}x - U^{(i)}y\|_2^2
\]
\[
= \beta^2 \|x - y\|_2^2,
\]
where the third and last equality holds due to the fact that $U^{(i)}[U^{(i)}] \top = I$ and $U^{(i)}[U^{(j)}] \top = 0$ for each $i \neq j$, and the inequality holds due to the fact that $g \in S^{(\xi, \zeta)}$. Thus, we have $\hat{g}_i \in V(\xi)$. To prove $\hat{G} \in S^{(\xi/\sqrt{n}, \zeta)}$, we have
\[
\nabla^2 \hat{G}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n U^{(i)}(U^{(i)}) \top \nabla^2 g(U^{(i)}x) \geq \frac{\xi}{\sqrt{n}} I,
\]
where the inequality holds due to the assumption that $g \in S^{(\xi, \beta)}$. With this fact and
\[
\|\nabla \hat{G}(x) - \nabla \hat{G}(y)\|_2^2 \leq \mathbb{E}_r \|\nabla \hat{g}_i(x) - \nabla \hat{g}_i(y)\|_2^2 \leq \beta^2 \|x - y\|_2^2
\]
which implies that $\nabla^2 \hat{G}(x) \preceq \beta I$, we conclude that $\hat{G} \in S^{(\xi/\sqrt{n}, \zeta)}$.

□

B.2. Proof of Lemma 5.2

Proof of Lemma 5.2. First we have $\{g_i\}_{i=1}^n \in \mathcal{V}^{(\lambda/\beta^2, L')}$ because for any $x, y \in \mathbb{R}^d$,
\[
\mathbb{E}_r \|\nabla g_i(x) - \nabla g_i(y)\|_2^2 = \lambda^2 \mathbb{E}_r \|\nabla g_i(x/\beta) - \nabla g_i(y/\beta/\beta)\|_2^2 \leq \lambda^2/\beta^2 \lambda^2 \|x/\beta - y/\beta\|_2^2
\]
\[
= (\lambda/\beta^2 \cdot L')\|x - y\|_2^2.
\]
Next we have $g_i \in S^{(\lambda/\beta^2 \xi', \lambda/\beta^2 \zeta')}$ because $\nabla^2 g_i(x) = \lambda/\beta^2 \nabla^2 \hat{g}_i(x/\beta)$ and for any $x \in \mathbb{R}^d$,
\[
\lambda/\beta^2 \cdot \xi' I \leq \lambda/\beta^2 \nabla^2 \hat{g}_i(x/\beta) \leq \lambda/\beta^2 \cdot \zeta' I.
\]
Next we have $G(0) - \inf_x G(x) \leq \lambda \Delta'$ because
\[
G(0) - \inf_x G(x) = \lambda \hat{G}(0) - \lambda \inf_x G(x) \leq \lambda \Delta'.
\]
Finally we have $\text{dist}(0, (Z')^*) \leq \beta B'$ because $(Z')^* = \beta \cdot Z^*$. □
B.3. Proof of Lemma 5.3

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

$$f_i(x) := -\sqrt{n}(x, e^{(i)}) + \frac{\|x\|^2}{2},$$

$$\bar{F}(x) := \frac{1}{n} \sum_{i=1}^n f_i(x),$$

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

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

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

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

which implies

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

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

B.4. Proof of Lemma A.1

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

$$f_i(x) := -\sqrt{n}(x, e^{(i)}) + \frac{\|x\|^2}{2},$$

$$\bar{F}(x) := \frac{1}{n} \sum_{i=1}^n f_i(x),$$

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

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

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