Appendix

1 Basic convexity inequalities

The following inequalities are classical. See Nesterov 1998 for proofs. They hold for all $x$ & $y$, when $f \in S^{1,1}_{x,L}$.

(B1) $f(y) \leq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2} \|x - y\|^2$

(B2) $f(y) \geq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2L} \|f'(x) - f'(y)\|^2$

(B3) $f(y) \geq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2} \|x - y\|^2$

(B4) $\langle f'(x) - f'(y), x - y \rangle \geq \frac{1}{2} \|f'(x) - f'(y)\|^2$

(B5) $\langle f'(x) - f'(y), x - y \rangle \geq s \|x - y\|^2$

We also use variants of B2 and B3 that are summed over each $f_i$, with $x = \phi_i$ and $y = w$:

$f(w) \geq \frac{1}{n} \sum_i f_i(\phi_i) + \frac{1}{n} \sum_i \langle f'_i(\phi_i), w - \phi_i \rangle + \frac{1}{2Ln} \sum_i \|f'(x) - f'(y)\|^2$

These are used in the following negated and rearranged form:

$-f(w) - T_2 = -f(w) + \frac{1}{n} \sum_i f_i(\phi_i) + \frac{1}{n} \sum_i \langle f'_i(\phi_i), w - \phi_i \rangle$

(B6) $-f(w) - T_2 \leq -\frac{s}{2n} \sum_i \|w - \phi_i\|^2$

(B7) $-f(w) - T_2 \leq -\frac{1}{2Ln} \sum_i \|f'(w) - f'(\phi_i)\|^2$.

2 Lyapunov term bounds

Simplifying each Lyapunov term is fairly straightforward. We use extensively that $\phi_j^{(k+1)} = w$, and that $\phi_i^{(k+1)} = \phi_i$ for $i \neq j$. Note also that

(B8) $w^{(k+1)} - w = \frac{1}{n} \langle f'_{j}(\phi_j), w - \phi_j \rangle + \frac{1}{\alpha s} [f'_{j}(\phi_j) - f'_{j}(w)]$.

Lemma 6. Between steps $k$ and $k+1$, the $T_1 = f(\tilde{\phi})$ term changes as follows:

$E[T_1^{(k+1)}] - T_1 \leq \frac{1}{n} \langle f'(\tilde{\phi}), w - \tilde{\phi} \rangle + \frac{L}{2n^2} \sum_i \|w - \phi_i\|^2$.

Proof. First we use the standard Lipschitz upper bound (B1):

$f(y) \leq f(x) + \langle f'(x), y - x \rangle + \frac{L}{2} \|x - y\|^2$.

We can apply this using $y = \tilde{\phi}^{(k+1)} = \tilde{\phi} + \frac{1}{n} \langle f'_{j}(\phi_j), w - \phi_j \rangle$ and $x = \tilde{\phi}$:

$f(\tilde{\phi}^{(k+1)}) \leq f(\tilde{\phi}) + \frac{1}{n} \langle f'(\tilde{\phi}), w - \phi_j \rangle + \frac{L}{2n^2} \|w - \phi_j\|^2$.

We now take expectations over $j$, giving:

$E[f(\tilde{\phi}^{(k+1)})] - f(\tilde{\phi}) \leq \frac{1}{n} \langle f'(\tilde{\phi}), w - \tilde{\phi} \rangle + \frac{L}{2n^2} \sum_i \|w - \phi_i\|^2$.
Lemma 7. Between steps $k$ and $k+1$, the $T_2 = -\frac{1}{n} \sum_i f_i(\phi_i) - \frac{1}{n} \sum_i \langle f_i'(\phi_i), w - \phi_i \rangle$ term changes as follows:

\[
E[T_2^{(k+1)}] - T_2 \leq -\frac{1}{n} T_2 - \frac{1}{n} f(w) + \frac{1}{\alpha} \sum \|f_i'(w) - f_i'(\phi_i)\|^2 + \frac{1}{n} \sum \langle f_i'(w) - f_i'(\phi_i), w - \phi_i \rangle.
\]

Proof. We introduce the notation $T_{21} = -\frac{1}{n} \sum_i f_i(\phi_i)$ and $T_{22} = -\frac{1}{n} \sum_i \langle f_i'(\phi_i), w - \phi_i \rangle$. We simplify the change in $T_{21}$ first using $\phi_{j}^{(k+1)} = w$:

\[
T_{21}^{(k+1)} - T_{21} = -\frac{1}{n} \sum_i f_i(\phi_{i}^{(k+1)}) + \frac{1}{n} \sum_i f_i(\phi_i) = -\frac{1}{n} \sum_i f_i(\phi_i) + \frac{1}{n} f_j(\phi_j) - \frac{1}{n} f_j(w) + \frac{1}{n} \sum_i f_i(\phi_i) = \frac{1}{n} f_j(\phi_j) - \frac{1}{n} f_j(w)
\]

Now we simplify the change in $T_{22}$:

\[
T_{22}^{(k+1)} - T_{22} = -\frac{1}{n} \sum_i \langle f_i'(\phi_i^{(k+1)}), w^{(k+1)} - w + w - \phi_{i}^{(k+1)} \rangle - T_{22}
\]

\[
\therefore T_{22}^{(k+1)} - T_{22} = -\frac{1}{n} \sum_i \langle f_i'(\phi_i^{(k+1)}), w - \phi_{i}^{(k+1)} \rangle - T_{22} - \frac{1}{n} \sum_i \langle f_i'(\phi_i^{(k+1)}), w^{(k+1)} - w \rangle.
\] (1)

We now simplifying the first two terms using $\phi_{j}^{(k+1)} = w$:

\[-\frac{1}{n} \sum_i \langle f_i'(\phi_i^{(k+1)}), w - \phi_{i}^{(k+1)} \rangle - T_{22} = T_{22} - \frac{1}{n} \langle f_j'(\phi_j), w - \phi_j \rangle + \frac{1}{n} \langle f_j'(w), w - w \rangle - T_{22} = \frac{1}{n} \langle f_j'(\phi_j), w - \phi_j \rangle.
\]

The last term of Equation 1 expands further:

\[-\frac{1}{n} \sum_i \langle f_i'(\phi_i^{(k+1)}), w^{(k+1)} - w \rangle = -\frac{1}{n} \langle \sum f_i'(\phi_i) - f_j'(\phi_j) + f_j'(w), w^{(k+1)} - w \rangle
\]

\[-\frac{1}{n} \langle \sum f_i'(\phi_i), w^{(k+1)} - w \rangle - \frac{1}{n} \langle f_j'(w) - f_j'(\phi_j), w^{(k+1)} - w \rangle.\] (2)

The second inner product term in 2 simplifies further using B8:

\[-\frac{1}{n} \langle f_j'(w) - f_j'(\phi_j), w^{(k+1)} - w \rangle = -\frac{1}{n} \langle f_j'(w) - f_j'(\phi_j), \frac{1}{n}(w - \phi_j) + \frac{1}{\alpha sn} [f_j'(\phi_j) - f_j'(w)] \rangle
\]

\[-\frac{1}{n^2} \langle f_j'(w) - f_j'(\phi_j), w - \phi_j \rangle - \frac{1}{\alpha sn^2} \langle f_j'(w) - f_j'(\phi_j), f_j'(\phi_j) - f_j'(w) \rangle.
\]

We simplify the second term:

\[-\frac{1}{\alpha sn^2} \langle f_j'(w) - f_j'(\phi_j), f_j'(\phi_j) - f_j'(w) \rangle = \frac{1}{\alpha sn^2} \|f_j'(w) - f_j'(\phi_j)\|^2.
\]

Grouping all remaining terms gives:

\[
T_{2}^{(k+1)} - T_2 \leq \frac{1}{n} f_j(\phi_j) + \frac{1}{n} \langle f_j'(\phi_j), w - \phi_j \rangle - \frac{1}{n} f_j(w) + \frac{1}{\alpha sn^2} \|f_j'(w) - f_j'(\phi_j)\|^2 - \frac{1}{n^2} \langle f_j'(w) - f_j'(\phi_j), w - \phi_j \rangle - \frac{1}{n} \langle \sum f_i'(\phi_i), w^{(k+1)} - w \rangle.
\]
We now take expectations of each remaining term. For the bottom inner product we use Lemma 1:

\[-\frac{1}{n} \left\langle \sum_i f'_i(\phi_i), w^{(k+1)} - w \right \rangle = \frac{1}{\alpha sn^2} \left\langle \sum_i f'_i(\phi_i), f'(w) \right \rangle = \frac{1}{n} \langle \bar{\phi} - w, f'(w) \rangle.\]

Taking expectations of the remaining terms is straightforward. We get:

\[
E[T_2^{(k+1)}] - T_2 \leq \frac{1}{n^2} \sum_i f_i(\phi_i) - \frac{1}{n} f(w) + \sum_i \frac{1}{\alpha sn^2} \left\| f'_i(w) - f'_i(\phi_i) \right\|^2 - \frac{1}{n^2} \sum_i \left\langle f'_i(w), f'_i(\phi_i), w - \phi_i \right \rangle + \frac{1}{n} \langle \bar{\phi} - w, f'(w) \rangle.
\]

**Lemma 8.** Between steps \( k \) and \( k + 1 \), the \( T_3 = -\frac{1}{2n} \sum_i \|w - \phi_i\|^2 \) term changes as follows:

\[
E[T_3^{(k+1)}] - T_3 = -(1 + \frac{1}{n}) T_3 + \frac{1}{\alpha n} \langle f'(w), w - \bar{\phi} \rangle - \frac{1}{2\alpha^2 sn^2} \sum_i \left\| f'_i(\phi_i) - f'_i(w) \right\|^2.
\]

**Proof.** We expand as:

\[
T_3^{(k+1)} = -\frac{s}{2n} \sum_i \left\| w^{(k+1)} - \phi_i^{(k+1)} \right\|^2
= -\frac{s}{2n} \sum_i \left\| w^{(k+1)} - w + w - \phi_i^{(k+1)} \right\|^2
= -\frac{s}{2} \left\| w^{(k+1)} - w \right\|^2 - \frac{s}{2n} \sum_i \left\| w - \phi_i^{(k+1)} \right\|^2 - \frac{s}{n} \sum_i \left\langle w^{(k+1)} - w, w - \phi_i^{(k+1)} \right \rangle.
\]

We expand the three terms on the right separately. For the first term:

\[
-\frac{s}{2} \left\| w^{(k+1)} - w \right\|^2 = -\frac{s}{2} \left\| \frac{1}{n} (w - \phi_j) + \frac{1}{\alpha sn} (f_j(\phi_j) - f_j(w)) \right\|^2
= -\frac{s}{2n^2} \left\| w - \phi_j \right\|^2 - \frac{1}{2\alpha^2 sn^2} \left\| f_j(\phi_j) - f_j(w) \right\|^2
- \frac{1}{\alpha n} \langle f_j(\phi_j) - f_j(w), w - \phi_j \rangle.
\]

For the second term of Equation 4, using \( \phi_j^{(k+1)} = w \):

\[
-\frac{s}{2n} \sum_i \left\| w - \phi_i^{(k+1)} \right\|^2 = -\frac{s}{2n} \sum_i \left\| w - \phi_i \right\|^2 + \frac{s}{2n} \left\| w - \phi_j \right\|^2
= T_3 + \frac{s}{2n} \left\| w - \phi_j \right\|^2.
\]

For the third term of Equation 4:

\[
-\frac{s}{n} \sum_i \left\langle w^{(k+1)} - w, w - \phi_i^{(k+1)} \right \rangle = -\frac{s}{n} \sum_i \left\langle w^{(k+1)} - w, w - \phi_j \right \rangle + \frac{s}{n} \left\langle w^{(k+1)} - w, w - \phi_j \right \rangle
= -s \left\langle w^{(k+1)} - w, w - \frac{1}{n} \sum_i \phi_i \right \rangle + \frac{s}{n} \left\langle w^{(k+1)} - w, w - \phi_j \right \rangle.
\]

The second inner product term in Equation 6 becomes (using B8):

\[
\frac{s}{n} \left\langle w^{(k+1)} - w, w - \phi_j \right \rangle = \frac{s}{n} \left\langle \frac{1}{n} (w - \phi_j) + \frac{1}{\alpha sn} [f'_j(\phi_j) - f'_j(w)], w - \phi_j \right \rangle
= \frac{s}{n^2} \left\| w - \phi_j \right\|^2 + \frac{1}{\alpha n^2} \left\langle f'_j(\phi_j) - f'_j(w), w - \phi_j \right \rangle.
\]
Now using \(1\) note that the inner product term here cancels with the one in 5.

Now we can take expectations of each remaining term. Recall that \(E[w^{(k+1)}] - w = -\frac{1}{\alpha n} f'(w)\), so the first inner product term in 6 becomes:

\[-sE \left[ \langle w^{(k+1)} - w, w - \frac{1}{n} \sum_i \phi_i \rangle \right] = \frac{1}{\alpha n} \langle f'(w), w - \bar{\phi} \rangle.\]

All other terms don’t simplify under expectations. So the result is:

\[E[T^{(k+1)}_4] - T_3 = \left( \frac{1}{2} - \frac{1}{n} \right) \frac{s}{n^2} \sum_i \|w - \phi_i\|^2 + \frac{1}{\alpha n} \langle f'(w), w - \bar{\phi} \rangle - \frac{1}{2\alpha^2 sn^3} \sum_i \|f_i(\phi_i) - f_i(w)\|^2.\]

**Lemma 9.** Between steps \(k\) and \(k + 1\), the \(T_4 = \frac{s}{2n} \sum_i \|\bar{\phi} - \phi_i\|^2\) term changes as follows:

\[E[T^{(k+1)}_4] - T_4 = -\frac{s}{2n} \sum_i \|\bar{\phi} - \phi_i\|^2 + \frac{s}{2n} \|\bar{\phi} - w\|^2 - \frac{s}{2n} \sum_i \|w - \phi_i\|^2.\]

**Proof.** Note that \(\bar{\phi}^{(k+1)} - \bar{\phi} = \frac{1}{n} (w - \phi_j)\), so:

\[T^{(k+1)}_4 = \frac{s}{2n} \sum_i \|\bar{\phi}^{(k+1)} - \bar{\phi} - \phi_i^{(k+1)}\|^2 = \frac{s}{2n} \sum_i \left( \|\bar{\phi}^{(k+1)} - \bar{\phi}\|^2 + \|\bar{\phi} - \phi_i^{(k+1)}\|^2 \right) + 2 \langle \bar{\phi}^{(k+1)} - \bar{\phi}, \bar{\phi} - \phi_i^{(k+1)} \rangle \]

\[= \frac{s}{2n} \sum_i \left( \|\frac{1}{n} (w - \phi_j)\|^2 + \|\bar{\phi} - \phi_i^{(k+1)}\|^2 \right) + \frac{2}{n} \langle w - \phi_j, \bar{\phi} - \phi_i^{(k+1)} \rangle.\]

Now using \(\frac{1}{n} \sum_i (\bar{\phi} - \phi_i^{(k+1)}) = \bar{\phi} - \bar{\phi}^{(k+1)} = -\frac{1}{n} (w - \phi_j)\) to simplify the inner product term:

\[= \frac{s}{2n} \|w - \phi_j\|^2 + \frac{s}{2n} \sum_i \|\bar{\phi} - \phi_i^{(k+1)}\|^2 + \frac{s}{n^2} \langle w - \phi_j, \phi_j - w \rangle \]

\[= \frac{s}{2n^2} \|w - \phi_j\|^2 + \frac{s}{2n} \sum_i \|\bar{\phi} - \phi_i^{(k+1)}\|^2 - \frac{s}{2n} \|w - \phi_j\|^2 \]

\[= \frac{s}{2n} \sum_i \|\phi - \phi_i^{(k+1)}\|^2 - \frac{s}{2n} \|w - \phi_j\|^2 \]

\[= \frac{s}{2n} \sum_i \|\phi - \phi_i\|^2 - \frac{s}{2n} \|\phi - \phi_j\|^2 + \frac{s}{2n} \|\bar{\phi} - w\|^2.\] (7)

Taking expectations gives the result.

**Lemma 10.** Let \(f \in S_{s,L}\). Then we have:

\[f(x) \geq f(y) + \langle f'(y), x - y \rangle + \frac{1}{2(L - s)} \|f'(x) - f'(y)\|^2 + \frac{sL}{2(L - s)} \|y - x\|^2 + \frac{s}{(L - s)} \langle f'(x) - f'(y), y - x \rangle.\]

**Proof.** Define the function \(g\) as \(g(x) = f(x) - \frac{s}{2} \|x\|^2\). Then the gradient is \(g'(x) = f'(x) - sx\). \(g\) has a lipschitz gradient with with constant \(L - s\). By convexity we have:

\[g(x) \geq g(y) + \langle g'(y), x - y \rangle + \frac{s}{2(L - s)} \|g'(x) - g'(y)\|^2.\]

Now replacing \(g\) with \(f\):

\[f(x) - \frac{s}{2} \|x\|^2 \geq f(y) - \frac{s}{2} \|y\|^2 + \langle f'(y) - sy, x - y \rangle + \frac{1}{2(L - s)} \|f'(x) - sx - f'(y) + sy\|^2.\]

Note that
\[
\frac{1}{2(L-s)} \|f'(x) - sx - f'(y) + sy\|^2 = \frac{1}{2(L-s)} \|f'(x) - f'(y)\|^2 + \frac{s^2}{2(L-s)} \|y-x\|^2
\]
so:

\[
f(x) \geq f(y) + \langle f'(y), x - y \rangle + \frac{1}{2(L-s)} \|f'(x) - f'(y)\|^2 + \frac{s^2}{2(L-s)} \|y-x\|^2
\]
\[
+ \frac{s}{2} \|x\|^2 - \frac{s}{2} \|y\|^2 + \frac{s}{L-s} \langle f'(x) - f'(y), y-x \rangle - s \langle y, y-x \rangle.
\]
Now using:

\[
\frac{s}{2} \|x\|^2 - s \langle y, x \rangle = -\frac{s}{2} \|y\|^2 + \frac{s}{2} \|x-y\|^2,
\]
we get:

\[
f(x) \geq f(y) + \langle f'(y), x - y \rangle + \frac{1}{2(L-s)} \|f'(x) - f'(y)\|^2 + \frac{s^2}{2(L-s)} \|x-y\|^2
\]
\[
- s \|y\|^2 + \frac{s}{2} \|x-y\|^2 + \frac{s}{L-s} \langle f'(x) - f'(y), y-x \rangle + s \langle y, y \rangle
\]
Note the norm \(y\) terms cancel, and:

\[
\frac{s}{2} \|x-y\|^2 + \frac{s^2}{2(L-s)} \|x-y\|^2 = \frac{(L-s)s + s^2}{2(L-s)} \|x-y\|^2
\]
\[
= \frac{sL}{2(L-s)} \|x-y\|^2.
\]
So:

\[
f(x) \geq f(y) + \langle f'(y), x - y \rangle + \frac{1}{2(L-s)} \|f'(x) - f'(y)\|^2 + \frac{sL}{2(L-s)} \|y-x\|^2
\]
\[
+ \frac{s}{L-s} \langle f'(x) - f'(y), y-x \rangle
\]

\[\square\]

**Corollary 11.** Take \(f(x) = \frac{1}{n} \sum_i f_i(x)\), with the big data condition holding with constant \(\beta\). Then for any \(x\) and \(\phi_i\) vectors:

\[
f(x) \geq \frac{1}{n} \sum_i f_i(\phi_i) + \frac{1}{n} \sum_i \langle f'_i(\phi_i), x - \phi_i \rangle + \frac{\beta}{2sn^2} \sum_i \|f'_i(x) - f'_i(\phi_i)\|^2
\]
\[
+ \frac{\beta L}{2n^2} \sum_i \|x - \phi_i\|^2 + \frac{\beta}{n} \sum_i \langle f'_i(x) - f'_i(\phi_i), \phi_i - x \rangle.
\]

**Proof.** We apply Lemma 10 to each \(f_i\), but instead of using the actual constant \(L\), we use \(\frac{\beta L}{n} + s\), which under the big data assumption is larger than \(L\):

\[
f_i(x) \geq f_i(\phi_i) + \langle f'_i(\phi_i), x - \phi_i \rangle + \frac{\beta}{2sn} \|f'_i(x) - f'_i(\phi_i)\|^2 + \frac{\beta L}{2n} \|x - \phi_i\|^2 + \frac{\beta}{n} \langle f'_i(x) - f'_i(\phi_i), \phi_i - x \rangle.
\]

Averaging over \(i\) gives the result.

\[\square\]

### 3 Lower complexity bounds

In this section we use the following technical assumption, as used in Nesterov (1998):

**Assumption 1:** An optimization method at step \(k\) may only invoke the oracle with a point \(x^{(k)}\) that is of the form:

\[
x^{(k)} = x^{(0)} + \sum_i a_i y^{(i)},
\]

where \(y^{(i)}\) is the derivative returned by the oracle at step \(i\), and \(a_i \in R\).

This assumption prevents an optimization method from just guessing the correct solution without doing any work. Virtually all optimization methods fall into under this assumption.
Simple $(1 - \frac{1}{n})^k$ bound

Any procedure that minimizes a sum of the form $f(w) = \frac{1}{n} \sum f_i(w)$ by uniform random access of $f_i$ is restricted by the requirement that it has to actually see each term at least once in order to find the minimum. This leads to a $(1 - \frac{1}{n})^k$ rate in expectation. We now formalize such an argument. We will work in $\mathbb{R}^n$, matching the dimensionality of the problem to the number of terms in the summation.

Theorem 12. For any $f \in FS_{1,n,n}(\mathbb{R}^n)$, we have that a $k$ step optimization procedure gives:

$$E[f(w)] - f(w^*) \geq \left(1 - \frac{1}{n}\right)^k \left(f(w^0) - f(w^*)\right)$$

Proof. We will exhibit a simple worst-case problem. Without loss of generality we assume that the first oracle access by the optimization procedure is at $w = 0$. In any other case, we shift our space in the following argument appropriately.

Let $f(w) = \frac{1}{n} \sum_i \left[ \frac{1}{2} (w_i - 1)^2 + \frac{1}{2} \|w\|^2 \right]$. Then clearly the solution is $w_i = \frac{1}{2}$ for each $i$, with minimum of $f(w^*) = \frac{n}{4}$. For $w = 0$ we have $f(0) = \frac{n}{2}$. Since the derivative of each $f_j$ is 0 on the $i$th component if we have not yet seen $f_i$, the value of each $w_i$ remains 0 unless term $i$ has been seen.

Let $v(k)$ be the number of unique terms we have not seen up to step $k$. Between steps $k$ and $k+1$, $v$ decreases by 1 with probably $\frac{1}{n}$ and stays the same otherwise. So

$$E[v^{(k+1)}|v^{(k)}] = v^{(k)} - \frac{v^{(k)}}{n} = \left(1 - \frac{1}{n}\right) v^{(k)}.$$ 

So we may define the sequence $X^{(k)} = (1 - \frac{1}{n})^{-k} v^{(k)}$, which is then martingale with respect to $v$, as

$$E[X^{(k+1)}|v^{(k)}] = \left(1 - \frac{1}{n}\right)^{-k-1} E[v^{(k+1)}|v^{(k)}] = \left(1 - \frac{1}{n}\right)^{-k} v^{(k)} = X^{(k)}.$$ 

Now since $k$ is chosen in advance, stopping time theory gives that $E[X^{(k)}] = E[X^{(0)}]$. So

$$E\left[\left(1 - \frac{1}{n}\right)^{-k} v^{(k)}\right] = n,$$

\Rightarrow: $E[v^{(k)}] = \left(1 - \frac{1}{n}\right)^k n$.

By Assumption 1, the function can be at most minimized over the dimensions seen up to step $k$. The seen dimensions contribute a value of $\frac{1}{4}$ and the unseen terms $\frac{1}{2}$ to the function. So we have that:

$$E[f(w^{(k)})] - f(w^*) \geq \frac{1}{4} \left(n - E[v^{(k)}]\right) + \frac{1}{2} E[v^{(k)}] - \frac{n}{4} = \frac{1}{4} E[v^{(k)}] = \left(1 - \frac{1}{n}\right)^k \frac{n}{4} = \left(1 - \frac{1}{n}\right)^k \left[f(w^{(0)}) - f(w^*)\right].$$

Minimization of non-strongly convex finite sums

It is known that the class of convex, continuous & differentiable problems, with $L$-Lipschitz continuous derivatives $F_{L}^{1,1}(\mathbb{R}^m)$, has the following lower complexity bound when $k < m$:

$$f(x^{(k)}) - f^*(x^*) \geq \frac{L \|x^{(0)} - x^*\|^2}{8(k + 1)^2}.$$
which is proved via explicit construction of a worst-case function where it holds with equality. Let this worst case function be denoted $h^{(k)}$ at step $k$.

We will show that the same bound applies for the finite-sum case, on a per pass equivalent basis, by a simple construction.

**Theorem 13.** The following lower bound holds for $k$ a multiple of $n$:

$$f(x^{(k)}) - f^{(k)}(x^*) \geq \frac{L \|x^{(0)} - x^*\|^2}{8\left(\frac{k}{n} + 1\right)^2},$$

when $f$ is a finite sum of $n$ terms $f(x) = \frac{1}{n} \sum_i f_i(x)$, with each $f_i \in F_{L_1}^1(\mathbb{R}^m)$, and with $m > kn$, under the oracle model where the optimization method may choose the index $i$ to access at each step.

**Proof.** Let $h_i$ be a copy of $h^{(k)}$ redefined to be on the subset of dimensions $i + jn$, for $j = 1 \ldots k$, or in other words, $h_i^{(k)}(x) = h^{(k)}([x_i, x_{i+n}, \ldots, x_{i+jn}, \ldots])$. Then we will use:

$$f^{(k)}(x) = \frac{1}{n} \sum_i h_i^{(k)}(x)$$

as a worst case function for step $k$.

Since the derivatives are orthogonal between $h_i$ and $h_j$ for $i \neq j$, by Assumption 1, the bound on $h_i^{(k)}(x^{(k)}) - h_i^{(k)}(x^*)$ depends only on the number of times the oracle has been invoked with index $i$, for each $i$. Let this be denoted $c_i$. Then we have that:

$$f(x^{(k)}) - f^{(k)}(x^*) \geq \frac{L \|x^{(0)} - x^*\|^2}{8n} \sum_i \frac{1}{(c_i + 1)^2}.$$

Where $\|\cdot\|_{(i)}^2$ is the norm on the dimensions $i + jn$ for $j = 1 \ldots k$. We can combine these norms into a regular Euclidean norm:

$$f(x^{(k)}) - f^{(k)}(x^*) \geq \frac{L \|x^{(0)} - x^*\|^2}{8n} \sum_i \frac{1}{(c_i + 1)^2}.$$

Now notice that $\sum_i \frac{1}{(c_i + 1)^2}$ under the constraint $\sum c_i = k$ is minimized when each $c_i = \frac{k}{n}$. So we have:

$$f(x^{(k)}) - f^{(k)}(x^*) \geq \frac{L \|x^{(0)} - x^*\|^2}{8n} \sum_i \frac{1}{\left(\frac{k}{n} + 1\right)^2} = \frac{L \|x^{(0)} - x^*\|^2}{8\left(\frac{k}{n} + 1\right)^2},$$

which is the same lower bound as for $k/n$ iterations of an optimization method on $f$ directly. \qed