DRAFT: FAST ACCESS TO ALL GRADIENTS IN N ITERATIONS

JIALIN YU

1. Abstract

In this draft, we first introduce some assumptions on the function class. Then we propose an algorithm that takes $O\left(Nd^2 \log^2 d/\sigma_{\min}^2(A)\right)$ iterations to initialize the stepping point x_s, from which we can access all N gradient vectors within exactly N iterations. By storing x_s, we only need to initialize for once. Finally, we will show that a randomly generated function from the function class given in Marsden et.al satisfies our assumptions with high probability, and $\sigma_{\min}(A) = \Omega\left(\sqrt{d}\right)$ indicates that the query complexity of this algorithm will not exceed $O(Nd \log^2 d)$, which approximately equals the optimal rate of optimizing the function class established in Marsden et.al. Therefore, when designing an algorithm for the function class given in Marsden et.al, using our algorithm for access to all gradients as a subroutine will not affect the query complexity.

2. Notations

For vector v, let $\|v\|_\infty$ and $\|v\|$ denote ℓ^∞ norm and Euclidean norm, respectively. Let $\mathbb{B}$ denote the unit ball in $\mathbb{R}^d$ under Euclidean norm. Let $\widehat{v}$ denote $v/\|v\|$. For matrix M, let $\sigma_{\min}(M)$ and $\sigma_{\max}(M)$ denote its smallest and largest singular value, respectively. For a finite set S, we use $|S|$ to denote its cardinality, $\overline{S}$ to denote its complement. For functionals $f(n)$ and $g(n)$, we denote $f(n) \gtrsim g(n)$ if $f(n) \geq cg(n)$, $f(n) \lesssim g(n)$ if $f(n) \leq Cg(n)$, and $f(n) \asymp g(n)$ if $cg(n) \leq f(n) \leq Cg(n)$ for some positive constants c and C. We denote $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$.

3. Theories and proofs

Let $f_v(x) = \langle v, x\rangle - \gamma_v$ with $\{\gamma_{v_i}\}_{i=1}^N$ to be different from each other, and

$$F(x) = \max\left\{\max_{i=1\ldots N} f_{v_i}(x), d^5\|Ax\|_\infty - 1\right\}, x \in \mathbb{B} \tag{3.1}$$

where $A \in \mathbb{R}^{p\times d}$. Let l_F and u_F denote the lower and upper bound respectively, of $\max_{i=1\ldots N} f_{v_i}(x), x \in \mathbb{B}$. Note that in the special case of Marsden et al. $f_{v_i}(x) = \langle v_i, x\rangle - i\gamma$, and $l_F = -O(1/\sqrt{N})$. Further, consider

$$U = \left\{x \middle| \|Ax\|_\infty \leq d^{-5}(1 + u_F)\right\}$$
$$L = \left\{x \middle| \|Ax\|_\infty \leq d^{-5}(1 + l_F)\right\}$$

Write $R(V)$ for the short hand of $R(v_1, \ldots, v_N)$, then it's easy to see that

$$L \subseteq R(V) \subseteq U.$$

Construct the refined projection algorithm (RP) as follows:

$$RP(x) = \begin{cases} P\left(P(x)/\left\|P(x)\right\|\right) \cdot \left\|P(x)\right\| & x \notin R(V) \\ x & x \in R(V) \end{cases}, \tag{3.2}$$

where $P(x)$ is the result of running the memory-constrainted projection algorithm starting from x. To put it in words, if $x \notin R(V)$, RP projects for two times with the second time acting on the normalized result of the first projection, and if $x \in R(V)$, RP does not do anything. It is easy to see that for any vector $x \in \mathbb{B}$, $RP(x) \in R(V) \subseteq U$, and provided that $x \notin R(V)$,

$$\left\|A\left(RP(x)\right)\right\|_\infty \leq d^{-5}(1+u_F)\left\|P(x)\right\| < d^{-5}(1+u_F)\left\|x\right\| \quad , \tag{3.3}$$

where the last inequality results from the contraction property of the Projection Algorithm. ((3.3) might be further improved, but it seems enough for now.)

Let $V_{\min}, V_{\max}$ be two constants such that $V_{\min} \leq \min_{i=1\ldots N}\left\|v_i\right\| \leq \max_{i=1\ldots N}\left\|v_i\right\| \leq V_{\max}$. Let $\gamma_{\max} := \max_{i=1\ldots N}\gamma_{v_i}$.

Assumption 1. *The following assumptions are adopted:*

(a) *For all $i \in [N]$, $\left|\left\langle RP\left(v_i\right), v_i\right\rangle\right| \geq \left\|v_i\right\|^2/2$; (sorry, but this has to be changed into $\left\|v_i\right\|^2 \cdot (p/d)$, which does not affect the results)*

(b) *For all $i, j \in [N]$ with $i \neq j$, $\left|\left\langle RP\left(v_i\right), v_j\right\rangle\right| \leq \beta$ for some apriori upper-bound β with $N\beta = o(1)$;*

(c) *$\sigma_{\min}(A) = \Omega(1)$;*

(d) *$\left(\sum_{i=1}^N\left(\gamma_{v_i} - \gamma_{\max}\right)^2\right)^{1/2} \leq r$ for some positive constant r with*

$$2r/V_{\min} < (1-\alpha)\frac{1+l_F}{1+u_F};$$

(e) *$N < d$; (this might not be sharp, but enough for the following analysis)*

We will show that the function class given in Marsden et. al satisfies Assumption 1 w.h.p in next section. Moreover,

$$\begin{aligned} \left|\langle RP(v_i), v_j - RP(v_j)\rangle\right| &= \left|\langle RP(v_i) - \mathrm{Proj}_{\mathrm{Ker}(A)}(v_i), v_j - RP(v_j)\rangle\right| \\ &\leq \left\|RP(v_i) - \mathrm{Proj}_{\mathrm{Ker}(A)}(v_i)\right\| \cdot \left\|v_j - RP(v_j)\right\| \\ &\leq \frac{\sqrt{p}}{\sigma_{\min}(A)}\left\|A\left(RP(x)\right)\right\|_\infty \cdot \left\|v_j - RP(v_j)\right\| \\ &\leq \frac{\sqrt{p}}{\sigma_{\min}(A)} d^{-5}(1+u_F)\cdot 2\left\|v_j\right\|, \end{aligned}$$

the last inequality uses (3.3). Therefore, we conclude that

$$\begin{aligned} \left|\langle RP(v_i), RP(v_j)\rangle\right| &\leq \left|\langle RP(v_i), v_j - RP(v_j)\rangle\right| + \left|\left\langle RP\left(v_i\right), v_j\right\rangle\right| \\ &\leq \frac{\sqrt{p}}{\sigma_{\min}(A)} d^{-5}(1+u_F)\cdot 2\left\|v_j\right\| + \beta = (1+o(1))\beta. \end{aligned} \tag{3.4}$$

Lemma 2. *If $x = \sum_{i=1}^N \lambda_i \cdot RP(v_i)$ satisfies $\left|\langle \lambda_i \cdot RP(v_i), v_i\rangle\right| \leq \Delta_i$ for some $\Delta_i > 0, i = 1\ldots N$, then*

$$\left\|x\right\| \leq (2+o(1)) \cdot \left(\sum_{i\in\mathcal{G}} \Delta_i^2/\left\|v_i\right\|^2\right)^{1/2}$$

where $\mathcal{G} = \{i \in [N] : \lambda_i \neq 0\}$.

Proof. Without loss of generality, assume $\mathcal{G} = [N]$. With (3.4), compute

$$\begin{aligned}
\|x\|^2 &= \sum_{i=1}^{N} \lambda_i^2 \|RP(v_i)\|^2 + \sum_{i\neq j} \lambda_i\lambda_j \langle RP(v_i), RP(v_j)\rangle \\
&\leq \sum_{i=1}^{N} 4\Delta_i^2 / \|v_i\|^4 \cdot \|RP(v_i)\|^2 + \sum_{i\neq j} 4\left(\Delta_i / \|v_i\|^2\right)\left(\Delta_j / \|v_j\|^2\right)(1+o(1))\beta \\
&\leq \sum_{i=1}^{N} 4\Delta_i^2 / \|v_i\|^2 + \sum_{i\neq j} 2(\Delta_i^2 / \|v_i\|^4 + \Delta_j^2 / \|v_j\|^4) \cdot (1+o(1))\beta \\
&= \sum_{i=1}^{N} 4\Delta_i^2 / \|v_i\|^2 + \sum_{i=1}^{N} 2(\Delta_i^2 / \|v_i\|^4) \cdot (1+o(1))\beta \cdot (2N-2) \\
&= \sum_{i=1}^{N} 4\Delta_i^2 / \|v_i\|^2 (1+o(1)).
\end{aligned}$$

The second last equality follows from the lower boundedness of $\|v_i\|$ and the last equality follows from $N\beta = o(1)$. □

3.1. **Initializing the stepping point x_s.** We would like to explain some intuition of algorithm 1. Let v_{i_k} be the gradient at x_k. Let $T < \min_{i=1\ldots N} f_{v_i}(x_0)$ be the target of algorithm in this round. Consider turning the value of $f_{v_{i_k}}(x)$ into the target T by moving from the current point x_k to $\widetilde{x_{k+1}} = x_k - \left(f_{v_{i_k}}(x_k) - T\right) \cdot RP\left(v_{i_k}\right) / \langle RP\left(v_{i_k}\right), v_{i_k}\rangle$. We move every step towards the target value by only accounting $f_{v_{i_k}}(x)$ for some $i_k \in [N]$, this would introduce a perturbation to the rest of the components of (3.1), i.e. all $f_{v_j}(x)$ with $j \neq i_k$. To be more specific, let $\Delta_k \coloneqq f_{v_{i_k}}(x_k) - T$ for all $0 \leq k \leq N_m - 1$. And the perturbation is

$$\begin{aligned}
\left|f_{v_j}(\widetilde{x_{k+1}}) - f_{v_j}(x_k)\right| &= |\Delta_k \cdot \langle RP\left(v_{i_k}\right), v_j\rangle / \langle RP\left(v_{i_k}\right), v_{i_k}\rangle| \\
&\leq |\Delta_k| \cdot 2\beta / \|v_{i_k}\|^2
\end{aligned} \tag{3.5}$$

The last inequality holds with Assumption 1. There is also another type of perturbation given $\widetilde{x_{k+1}} \notin R(V)$, which results from the projection of $\widetilde{x_{k+1}}$ to $x_{k+1} = RP(\widetilde{x_{k+1}})$. To analyze this type of perturbation, we first compute

$$\begin{aligned}
&\left\|A(\widetilde{x_{k+1}} - x_k)\right\|_\infty \\
&= \left\|\Delta_k \cdot A\left(RP\left(v_{i_k}\right)\right) / \langle RP\left(v_{i_k}\right), v_{i_k}\rangle\right\|_\infty \\
&\leq 2\left|\Delta_k\right| \cdot \left\|A\left(RP\left(v_{i_k}\right)\right)\right\|_\infty / \|v_{i_k}\|^2 \\
&\leq 2\left|\Delta_k\right| \cdot d^{-5}(1+u_F) / \|v_{i_k}\|^2
\end{aligned} \tag{3.6}$$

The last inequality holds by (3.3). Now we can analyze the process of the projection from $\widetilde{x_{k+1}}$ to $x_{k+1} = RP(\widetilde{x_{k+1}})$ given $\widetilde{x_{k+1}} \notin R(V)$ and $x_k \in R(V)$.

Algorithm 1: Initialize x_s

Input: $N, \beta, V_{\min}, \gamma_{\max}$;
Initialize $x = \mathbf{0}, T = -\gamma_{\max} - \left(2\sqrt{N}\beta/V_{\min}^2\right), m_1 = \lceil -\log N/\log(7N\beta/V_{\min}^2)\rceil, m_2 = O(\log d)$;
for $m = 1, 2, \ldots, m_1$ **do**
  $History = \emptyset$;
  $(v, f) \leftarrow Query(x)$;
  while $\langle v, x\rangle - f \notin History$ **do**
    add $\langle v, x\rangle - f$ to $History$;
    $x = x - (f - T) \cdot RP\left(v\right) / \langle RP\left(v\right), v\rangle$;
    $x = RP(x)$;
    $(v, f) \leftarrow Query(x)$;
  end
  $T = T - (2\sqrt{N}\beta/V_{\min}^2)(7N\beta/V_{\min}^2)^{m-1}$;
end
$x = P\left(P(x)/\left\|P(x)\right\|\right) \cdot \|P(x)\|$;
for $m = m_1 + 1, \ldots, m_2$ **do**
  $History = \emptyset$;
  $(v, f) \leftarrow Query(x)$;
  while $\langle v, x\rangle - f \notin History$ **do**
    add $\langle v, x\rangle - f$ to $History$;
    $x = x - (f - T) \cdot RP\left(v\right) / \langle RP\left(v\right), v\rangle$;
    $x = RP(x)$;
    $(v, f) \leftarrow Query(x)$;
  end
  $T = T - (2\sqrt{N}\beta/V_{\min}^2)(7N\beta/V_{\min}^2)^{m-1}$;
end
Output x as x_s.

Recall that $A \in \mathbb{R}^{p\times d}$,

$$
\begin{aligned}
\left\|\widetilde{x_{k+1}} - x_{k+1}\right\| &\leq \frac{\sqrt{p}}{\sigma_{\min}(A)} \left\|A\left(\widetilde{x_{k+1}} - RP(\widetilde{x_{k+1}})\right)\right\|_\infty \\
&\leq \frac{\sqrt{p}}{\sigma_{\min}(A)} \left(\frac{2d^{-5}\left|\Delta_k\right|}{\left\|v_{i_k}\right\|^2} \cdot (1 + u_F) + \left\|A\left(x_k - RP(\widetilde{x_{k+1}})\right)\right\|_\infty\right) \\
&\leq \frac{\sqrt{p}}{\sigma_{\min}(A)} \left(\frac{2d^{-5}\left|\Delta_k\right|}{\left\|v_{i_k}\right\|^2} \cdot (1 + u_F) + 2d^{-5}(1 + u_F)\right),
\end{aligned} \tag{3.7}
$$

The second inequality holds with (3.6), and the last inequality holds because $RP(\widetilde{x_{k+1}}) \in R(V) \subseteq U$ and $x_k \in R(V) \subseteq U$. Recall that v_{i_k} is the gradient at x_k. For $j \in [N]$, the perturbation is

$$
\begin{aligned}
\left|f_{v_j}(\widetilde{x_{k+1}}) - f_{v_j}(x_{k+1})\right| &= \left|\langle v_j, \widetilde{x_{k+1}} - x_{k+1}\rangle\right| \\
&\leq \|v_j\| \cdot \left\|\widetilde{x_{k+1}} - x_{k+1}\right\|
\end{aligned} \tag{3.8}
$$

Let N_m denote the number of iterations in round m, which indicates that the gradient at x_{N_m} coincides with the gradient at x_0. For some $1 \le l \le N_m - 1$, consider the perturbation on $f_{v_{i_l}}(x)$ before observing the gradient vector v_{i_l}:

$$\begin{aligned}\left|f_{v_{i_l}}(x_l) - f_{v_{i_l}}(x_0)\right| &\le \sum_{k=0}^{l-1} \left|f_{v_{i_l}}(x_{k+1}) - f_{v_{i_l}}(x_k)\right| \\ &\le \sum_{k=0}^{l-1} \left|f_{v_{i_l}}(x_{k+1}) - f_{v_{i_l}}(\widetilde{x_{k+1}})\right| + \sum_{k=0}^{l-1} |\Delta_k| \cdot 2\beta / \|v_{i_k}\|^2\end{aligned}$$

The last inequality follows from (3.5). By (3.2), $x_{k+1} \ne \widetilde{x_{k+1}}$ if and only if $\widetilde{x_{k+1}} \notin R(V)$. Let $S_m \coloneqq \left\{0 \le k \le N_m - 1 : \widetilde{x_{k+1}} \notin R(V)\right\}$. Combining (3.7) and (3.8) yields:

(3.9)

$$\begin{aligned}&\left|f_{v_{i_l}}(x_l) - f_{v_{i_l}}(x_0)\right| \\ &\le \sum_{k\in S_m} \frac{\|v_{i_l}\|\sqrt{p}}{\sigma_{\min}(A)}\left(\frac{2d^{-5}|\Delta_k|}{\|v_{i_k}\|^2}\cdot(1+u_F) + 2d^{-5}(1+u_F)\right) + \sum_{k=0}^{l-1}|\Delta_k|\cdot 2\beta/\|v_{i_k}\|^2 \\ &\le \left(1 + \frac{d^{-5}(1+u_F)V_{\max}\sqrt{p}}{\beta\sigma_{\min}(A)}\right)\sum_{k=0}^{N_m-1}|\Delta_k|\cdot 2\beta/\|v_{i_k}\|^2 + |S_m|\tau,\end{aligned}$$

with $\tau \coloneqq \frac{V_{\max}\sqrt{p}}{\sigma_{\min}(A)} 2d^{-5}(1+u_F)$ that turns out to be an important threshold. Let $\delta_l \coloneqq f_{v_{i_l}}(x_0) - T$, we can now assert that

$$\Delta_l \le \delta_l + \left|f_{v_{i_l}}(x_l) - f_{v_{i_l}}(x_0)\right|. \tag{3.10}$$

Combining this with (3.9), we have

$$\begin{aligned}\sum_{k=0}^{N_m-1}|\Delta_k| \le \sum_{k=0}^{N_m-1}\delta_k + \frac{2N_m\beta}{V_{\min}^2}\sum_{k=0}^{N_m-1}|\Delta_k| \\ +N_m\cdot\sum_{k\in S_m}\frac{V_{\max}\sqrt{p}}{\sigma_{\min}(A)}\left(\frac{2d^{-5}|\Delta_k|}{V_{\min}^2}\cdot(1+u_F) + 2d^{-5}(1+u_F)\right).\end{aligned}$$

With $\sigma_{\min}(A) = \Omega(1)$, $d^{-5}N^2\sqrt{p}/\sigma_{\min}(A) = o(1)$ and $N\beta = o(1)$, this immediately yields

$$\begin{aligned}\sum_{k=0}^{N_m-1}|\Delta_k| &\le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k\right) + N_m\cdot|S_m|\frac{V_{\max}\sqrt{p}}{\sigma_{\min}(A)}2d^{-5}(1+u_F)\right) \\ &= (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k\right) + N_m\cdot|S_m|\tau\right).\end{aligned} \tag{3.11}$$

(3.11) indicates the following crude bound that is also useful

$$\sum_{k=0}^{N_m-1}|\Delta_k| \le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k\right) + N_m^2\cdot\tau\right) \quad . \tag{3.12}$$

Similar to (3.9), at the end of the whole round the total perturbation to $f_{v_{i_l}}(x)$ is:
(3.13)

$$
\begin{aligned}
&\left|f_{v_{i_l}}(x_{N_m}) - T\right| \le \sum_{k=l+1}^{N_m-1} \left|f_{v_{i_l}}(x_k) - f_{v_{i_l}}(x_k)\right| \\
&\le \sum_{k>l}^{N_m-1} \left|f_{v_{i_l}}(x_{k+1}) - f_{v_{i_l}}(\widetilde{x_{k+1}})\right| + \sum_{k>l}^{N_m-1} \left|f_{v_{i_l}}(\widetilde{x_{k+1}}) - f_{v_{i_l}}(x_k)\right| \\
&\le \sum_{k\in S_m} \frac{\|v_{i_l}\|\sqrt{p}}{\sigma_{\min}(A)} \left(\frac{2d^{-5}\left|\Delta_k\right|}{\|v_{i_k}\|^2} \cdot (1+u_F) + 2d^{-5}(1+u_F) \right) + \sum_{k=l+1}^{N_m-1} \left|\Delta_k\right| \cdot 2\beta / \|v_{i_k}\|^2 \\
&\le \left(1 + \frac{d^{-5}(1+u_F)V_{\max}\sqrt{p}}{\beta\sigma_{\min}(A)}\right) \sum_{k=0}^{N_m-1} \left|\Delta_k\right| \cdot 2\beta / \|v_{i_k}\|^2 + |S_m|\,\tau
\end{aligned}
$$

The first inequality holds because by construction of the algorithm, $f_{v_{i_l}}(x_{l+1}) = T$.

In the above analysis, we have discussed the iterations in round m using $v_{i_k}, x_k, \widetilde{x_k}, \delta_k, \Delta_k$ and T. Now we need to do induction over m and use $v_{i_k}^{(m)}, x_k^{(m)}, \widetilde{x_k}^{(m)}, \delta_k^{(m)}, \Delta_k^{(m)}$ and $T^{(m)}$ to denote the corresponding variables in round m.

Briefly recall from algorithm 1: $T^{(m)} = T^{(m-1)} - (2\sqrt{N}\beta/V_{\min}^2)(7N\beta/V_{\min}^2)^{m-2}$.

Theorem 3. *Suppose that Assumption 1 holds. Then for the m'th round ($1 \le m \le m_1 < m_0$ with $m_0 = \lfloor -5\log N/\log(7N\beta/V_{\min}^2)\rfloor = O(\log d)$) of Algorithm 1, the following statements hold:*

1. For all $l \in [N_m]$,

$$
\begin{aligned}
\max\left\{\left\|\widetilde{x_l}^{(m)}\right\|, \left\|x_l^{(m)}\right\|\right\} &\le \left(\frac{2+o(1)}{V_{\min}}\right) r + \left(\frac{2+o(1)}{V_{\min}^3}\right) \cdot \sum_{k=2}^{m} (7N\beta/V_{\min}^2)^{k-1} + (m-1)(2+o(1))N^{3/2}\tau/V_{\min} \\
&< r \cdot (2+o'(1))/V_{\min}
\end{aligned}
$$

with some $o(1)$ and $o'(1)$ that do not vary for different m.

2. For all $k \in [N_m]$, $\left|f_{v_{i_k}}(x_{N_m}^{(m)}) - T^{(m)}\right| < \left(\sqrt{N}\beta/V_{\min}^2\right) \cdot (7N\beta/V_{\min}^2)^{m-1}$

3. $N_m = N$ i.e. all gradients will be queried in round m;

Theorem 3 and Theorem 4 corresponds to the two "for loops" as illustrated in algorithm 1. Between them lies the mandatory projection step, i.e. no matter $x \notin R(V)$ or not, we update $x \leftarrow P\left(P(x)/\|P(x)\|\right) \cdot \|P(x)\|$. This is important because the distance from $\mathbb{B}\backslash R(V)$ will always be lower bounded by a certain number ever since the mandatory projection, which indicates that x will not get out of $R(V)$ for the rest of the algorithm.

Theorem 4. *Suppose that Assumption 1 holds. Then for the m 'th round ($m \ge m_1 + 1$) of Algorithm 1, the following statements hold:*

1. $\sum_{j=m_1+1}^{m} \sum_{k=0}^{N_j-1} \left|\Delta_k^{(j)}\right| \le \frac{3+o(1)}{7\sqrt{N}} \cdot \sum_{i=1}^{m-m_1} \left(7N\beta/V_{\min}^2\right)^{i-1} < \frac{1}{2\sqrt{N}}$ with some $o(1)$ that does not vary for different m.

2. $|S_m| = 0$, i.e. the Refined Projection algorithm will not be called in round m.

3. For all $l \in [N_m]$,

$$
\max\left\{\left\|\widetilde{x_l}^{(m)}\right\|, \left\|x_l^{(m)}\right\|\right\} \le \left(\frac{2+o_1(1)}{V_{\min}}\right) r + \left(\frac{2+o(1)}{V_{\min}^3}\right) \cdot \sum_{k=2}^{m} (7N\beta/V_{\min}^2)^{k-1} < r \cdot (2+o_2(1))/V_{\min}
$$

with some $o(1)$, $o_1(1)$ and $o_2(1)$ that do not vary for different m.

4. *For all $k \in [N_m]$, $\left|f_{v_{i_k}}(x_{N_m}^{(m)}) - T^{(m)}\right| < \left(\sqrt{N}\beta/V_{\min}^2\right) \cdot (7N\beta/V_{\min}^2)^{m-1}$*
5. *$N_m = N$ i.e. all gradients will be queried in round m;*

From statement 2 of Theorem 4 we can see that the Refined Projection algorithm will not be called once $m \geq m_1 + 1$, so $\widetilde{x_l}^{(m)} = x_l^{(m)}$ for $m \geq m_1 + 1$ and $l \in [N]$. Further, from the proof of Theorem 4 we will see that for all $l \in [N]$

$$\begin{aligned}\left\|A(\widetilde{x_l}^{(m)})\right\|_\infty &\leq \left\|A(x_0^{(m_1+1)})\right\|_\infty + \sum_{j=m_1+1}^{m}\sum_{k=0}^{N_j-1}\left\|A(\widetilde{x_{k+1}}^{(j)} - x_k^{(j)})\right\|_\infty \\ &< d^{-5}(1+l_F)(1-\alpha) + \sum_{j=m_1+1}^{m}\sum_{k=0}^{N_j-1} 2\left|\Delta_k^{(j)}\right| \cdot d^{-5}(1+u_F)/V_{\min}^2 \quad , \\ &< d^{-5}(1+l_F)(1-\alpha+\frac{1+u_F}{\sqrt{N}\,(1+l_F)\,V_{\min}^2}) < d^{-5}(1+l_F)(1-\alpha/2)\end{aligned}$$

Note that we need $\left\|A(\widetilde{x_l}^{(m)})\right\|_\infty \geq d^{-5}(1+l_F)$ if $\widetilde{x_l}^{(m)} \in \mathbb{B}\backslash R(V)$. This gives the lower bound of the distance rom $\mathbb{B}\backslash R(V)$ uniformly as $m \to \infty$.

What we are trying to explain here is that, by statement 4 of Theorem 4, algorithm 1 will get x_s arbitrarily close to $I(V) \coloneqq \{x : f_{v_1}(x) = f_{v_2}(x) = \cdots = f_{v_N}(x)\}$ while keeping a constant distance from $\mathbb{B}\backslash R(V)$. Within $O(\log d)$ rounds, x_s will end up in a "scenario" that is similar to the case without A because the distance from $I(V)$ is very small in comparison with the distance from $\mathbb{B}\backslash R(V)$. At this point, we can just perform the moves in the direction of v_i instead of $RP(v_i)$ as shown in the following algorithm 2 and do not need to worry about getting out of $R(V)$. Since it reduces to the case without A, within N iterations, every gradient v_i will be seen. We omit the proof of this claim, which is very similar to the proof of statement 5 in Theorem 4.

Algorithm 2: Access all gradients

```
Input: N, x_s, m_2, T^(m_2);
Initialize x = x_s, T = T^(m_2) − (2√N β/V_min^2)(7Nβ/V_min^2)^(m_2−1);
for k = 1, 2, ..., N do
    (v, f) ← Query(x);
    x = x − (f − T) · v/⟨v, v⟩;
    Output v;
end
```

Now we turn to analyze the query complexities. The Projection algorithm takes $O\left(d^2 \log d/\sigma_{\min}^2(A)\right)$ iterations and there are $O(N)$ calls for the Projection algorithm in each round. With $O\left(Nd^2 \log d/\sigma_{\min}^2(A)\right)$ iterations in each round and a total of $O(\log d)$ rounds, algorithm 1 takes $O\left(Nd^2 \log^2 d/\sigma_{\min}^2(A)\right)$ iterations. Though the computation of x_s is expensive, we only need to do it once. After that, we can just do algorithm 2 and see all gradients within only N iterations.

3.2. **Verification of Assumption 1 for the function class given in Marsden et.al.** We would like to point out that by (3.3),

$$\left\| \text{Proj}_{\text{Ker(A)}}(x) - RP(x) \right\| \leq \frac{\sqrt{p}}{\sigma_{\min}(A)} \cdot d^{-5}(1+u_F) \|x\|,$$

and that verifying Assumption 1 with all $RP(\cdot)$ replaced by $\text{Proj}_{\text{Ker(A)}}(\cdot)$ is enough.

Among all the assumptions $(a) \sim (e)$, the hardest one to be verified is (a), so we will only verify (a) here for brevity. (This draft has exceeded 20 pages...)

Lemma 5. *$\mathbb{E}\left[\langle Proj_{Ker(A)}(v_i), v_i\rangle\right] = \|v_i\|^2/2$ and $\langle Proj_{Ker(A)}(v_i), v_i\rangle \geq \|v_i\|^2/3$ w.h.p.*

Proof. Recall that $\|v_i\| = 1$ and $p = d/2$ in Marsden et.al. Let $B \in \mathbb{R}^{\frac{d}{2}\times d}$ be a matrix with its rows $\{b_i\}_{i=1}^{d/2}$ to be an orthonormal basis for $\text{Ker}(A)$. Write v as the shorthand of v_i. Then we have

$$\langle \text{Proj}_{\text{Ker(A)}}(v), v\rangle = v^T B^T B v = \sum_{i=1}^{d/2} v^T b_i b_i^T v \geq 0$$

with strict inequality when $v \in \text{Ker}A^{\perp}$. And

$$\mathbb{E}\langle \text{Proj}_{\text{Ker(A)}}(v), v\rangle = \sum_{i=1}^{d/2} \mathbb{E}\left(v^T b_i\right)^2 = \sum_{i=1}^{d/2} \|b_i\|^2/d = 1/2.$$

We turn to analyze the variance:

$$\text{Var}\langle \text{Proj}_{\text{Ker(A)}}(v), v\rangle = \sum_{i=1}^{d/2} \text{Var}\left[\left(v^T b_i\right)^2\right] + \sum_{i\neq j} \text{Cov}\left[\left(v^T b_i\right)^2, \left(v^T b_j\right)^2\right] \tag{3.14}$$

For the first term in (3.14),

$$\begin{aligned}\text{Var}\left[\left(v^T b_i\right)^2\right] &= \mathbb{E}\left[\left(v^T b_i\right)^4\right] - 1/d^2 \\ &= (1/d^2)\mathbb{E}\left(\sum_{k=1}^{d} s_k b_{ik}\right)^4 - 1/d^2,\end{aligned}$$

where $s_k \in \{1, -1\}$ is a binomial random variable with $\mathbb{P}[s_k = 1] = 1/2$. Further,

$$\begin{aligned}\mathbb{E}\left(\sum_{k=1}^{d} s_k b_{ik}\right)^4 &= \mathbb{E}\left(\sum_{k=1}^{d} s_k^4 b_{ik}^4\right) + \mathbb{E}\sum_{k<l} \binom{4}{2} s_k^2 b_{ik}^2 s_l^4 b_{il}^2 \\ &= \sum_{k=1}^{d} b_{ik}^4 + \sum_{k<l} 6 b_{ik}^2 b_{il}^2 \\ &< 3\left(\sum_{k=1}^{d} b_{ik}^4 + 2\sum_{k<l} b_{ik}^2 b_{il}^2\right) = 3\|b_i\|^4 = 3\end{aligned}$$

We have obtained $\mathrm{Var}\left[\left(v^T b_i\right)^2\right] < 2/d^2$. Now we turn to bound the second term in (3.14). We will show that the covariance term is always nonpositive. Compute

$$\begin{aligned}\mathrm{Cov}\left[\left(v^T b_i\right)^2, \left(v^T b_j\right)^2\right] &= \mathbb{E}\left[\left(v^T b_i\right)^2 \left(v^T b_j\right)^2\right] - \mathbb{E}\left(v^T b_i\right)^2 \mathbb{E}\left(v^T b_j\right)^2 \\ &= (1/d^2)\mathbb{E}\left[\left(\sum_{k=1}^d s_k b_{ik}\right)^2 \left(\sum_{l=1}^d s_l b_{jl}\right)^2\right] - 1/d^2\end{aligned}$$

Further,

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{k=1}^d s_k b_{ik}\right)^2 \left(\sum_{k=1}^d s_k b_{jk}\right)^2\right] \\ &= \mathbb{E}\left[\sum_{k=1}^d b_{ik}^2 b_{jk}^2 s_k^4 + \sum_{k<l} \left(2b_{ik}b_{il}s_k s_l\right)\left(2b_{jk}b_{jl}s_k s_l\right) + \sum_{k\neq l} \left(s_k b_{ik}\right)^2 \left(s_l b_{jl}\right)^2\right] \\ &= \sum_{k=1}^d b_{ik}^2 b_{jk}^2 + 4\sum_{k<l} b_{ik}b_{il}b_{jk}b_{jl} + \sum_{k\neq l} b_{ik}^2 b_{jl}^2 \\ &= \|b_i\|^2 \|b_j\|^2 + 2\sum_{k\neq l} \left(b_{ik}b_{jk}\right)\left(b_{il}b_{jl}\right) \\ &= 1 + 2\sum_{k,l} \left(b_{ik}b_{jk}\right)\left(b_{il}b_{jl}\right) - 2\sum_k b_{ik}^2 b_{jk}^2 \\ &= 1 + 2\langle b_i, b_j\rangle^2 - 2\sum_k b_{ik}^2 b_{jk}^2\end{aligned}$$

we already know that $\{b_i\}_{i=1}^{d/2}$ is an orthonormal basis, so $\langle b_i, b_j\rangle = 0$, and

$$\mathbb{E}\left[\left(\sum_{k=1}^d s_k b_{ik}\right)^2 \left(\sum_{k=1}^d s_k b_{jk}\right)^2\right] = 1 - 2\sum_k b_{ik}^2 b_{jk}^2 < 1$$

which results in $\mathrm{Cov}\left[\left(v^T b_i\right)^2, \left(v^T b_j\right)^2\right] < 0$. Consequently,

$$\mathrm{Var}\langle \mathrm{Proj}_{\mathrm{Ker(A)}}\left(v\right), v\rangle < 2/d^2.$$

With Chebyshev's inequality,

$$\mathbb{P}\left[\langle \mathrm{Proj}_{\mathrm{Ker(A)}}\left(v\right), v\rangle \geq 1/3\right] \leq \frac{2/d^2}{(1/2-1/3)^2} = 72/d^2.$$

With union bound,

$$\mathbb{P}\left[\max_{i=1\ldots N} \langle \mathrm{Proj}_{\mathrm{Ker(A)}}\left(v_i\right), v_i\rangle \geq 1/3\right] \leq 72N/d^2 = 1/\mathrm{poly}(d).$$

□

3.3. **Proof of Theorems.**

Proof of Theorem 3.

Proof. We first do the case for $m = 1$.

We start with the first statement. By Lemma 2 and with $x_0^{(1)} = \mathbf{0}$,

$$\begin{aligned}\left\|x_l^{(1)}\right\| &\le \left\|\sum_{k=1}^{N_1}\left(\widetilde{x_k}^{(1)} - x_{k-1}^{(1)}\right)\right\| + \sum_{k\in S_1}\left\|\widetilde{x_k}^{(1)} - x_k^{(1)}\right\| \\ &\le (2+o(1))\cdot\left(\sum_{k=1}^{N_1}(\Delta_k^{(1)})^2/\left\|v_{i_k}^{(1)}\right\|^2\right)^{1/2} + \sum_{k=1}^{N_1}\left\|\widetilde{x_k}^{(1)} - x_k^{(1)}\right\|\end{aligned} \tag{3.15}$$

Combining (3.7) and (3.12) yields

$$\begin{aligned}&\sum_{k=1}^{N_1}\left\|\widetilde{x_k}^{(1)} - x_k^{(1)}\right\| \\ &\le \frac{\sqrt{p}}{\sigma_{\min}(A)}\left(\left(\sum_{k=1}^{N_1}\frac{2d^{-5}\left|\Delta_k^{(1)}\right|}{\left\|v_{i_k}^{(1)}\right\|^2}\cdot(1+u_F)\right) + 2N_1d^{-5}(1+u_F)\right) \\ &\le \frac{\sqrt{p}}{\sigma_{\min}(A)}\left(2d^{-5}/V_{\min}^2\right)\cdot(1+u_F)(1+o(1))\left(\left(\sum_{k=0}^{N_1-1}\delta_k^{(1)}\right) + N_1^2\tau\right) \\ &\qquad + \frac{\sqrt{p}}{\sigma_{\min}(A)}\cdot 2N_1d^{-5}(1+u_F) \\ &= \frac{\sqrt{p}}{\sigma_{\min}(A)}\left(2d^{-5}/V_{\min}^2\right)\cdot(1+u_F)(1+o(1))\left(\sum_{k=0}^{N_1-1}\delta_k^{(1)}\right) + \frac{\sqrt{p}}{\sigma_{\min}(A)}\cdot 2N_1d^{-5}(1+u_F)(1+o(1)) \\ &= \frac{\sqrt{p}}{\sigma_{\min}(A)}\left(2d^{-5}\right)\cdot(1+u_F)(1+o(1))\left(N_1 + \sum_{k=0}^{N_1-1}\delta_k^{(1)}/V_{\min}^2\right)\end{aligned} \tag{3.16}$$

Further, given that $\delta_k^{(1)} = -\gamma_{v_{i_k}^{(1)}} - T^{(1)}$ and $T^{(1)} = -\left(\gamma_{\max} + 2\sqrt{N}\beta/V_{\min}^2\right)$,

$$\left(\sum_{k=0}^{N_1-1}(\delta_k^{(1)})^2\right)^{1/2} \le \left(\sum_{k=0}^{N_1-1}\left(\gamma_{v_{i_k}^{(1)}} - \gamma_{\max}\right)\right)^{1/2} + 2\sqrt{N_1N}\beta/V_{\min}^2 \le r + 2N\beta/V_{\min}^2, \tag{3.17}$$

where the last inequality follows from $N_1 \le N$ and (d) in Assumption 1. Provided that $\sum_{k=0}^{N_1-1}\delta_k^{(1)} \le \sqrt{N}\left(\sum_{k=0}^{N_1-1}(\delta_k^{(1)})^2\right)^{1/2} < \sqrt{N}$, from (3.16) we have

$$\sum_{k=1}^{N_1}\left\|\widetilde{x_k}^{(1)} - x_k^{(1)}\right\| = O(\frac{Nd^{-5}\sqrt{p}}{\sigma_{\min}(A)}) = o(1).$$

Then we turn to bound the first term in (3.15). By (3.9), we have

$$
\begin{aligned}
&\left| f_{v_{i_l}^{(1)}}(x_l^{(1)}) - f_{v_{i_l}^{(1)}}(x_0^{(1)}) \right| \\
&\le (1+o(1)) \sum_{k=0}^{N_1-1} \left|\Delta_k^{(1)}\right| \cdot 2\beta / \left\| v_{i_k}^{(1)} \right\|^2 + |S_1|\,\tau \\
&\le (1+o(1)) \left(\left(\sum_{k=0}^{N_1-1} \delta_k^{(1)} \right) + N_1^2 \tau \right) \cdot 2\beta / V_{\min}^2 + |S_1|\,\tau \\
&\le (1+o(1)) \left(\sum_{k=0}^{N_1-1} \left(\delta_k^{(1)}\right)^2 \right)^{1/2} \sqrt{N} \cdot 2\beta / V_{\min}^2 + o(\sqrt{N}\beta)
\end{aligned} \tag{3.18}
$$

where the second inequality follows from (3.12) and the last inequality follows from Cauchy's inequality and $N_1 \le N$. Combining this with (3.10) yields

$$
\begin{aligned}
\left(\sum_{k=1}^{N_1} (\Delta_k^{(1)})^2 / \left\| v_{i_k}^{(1)} \right\|^2 \right)^{1/2} &\le (1/V_{\min}) \left(\left(\sum_{k=0}^{N_1-1} (\delta_k^{(1)})^2 \right)^{1/2} + \left(\sum_{l=0}^{N_1-1} \left| f_{v_{i_l}^{(1)}}(x_l^{(1)}) - f_{v_{i_l}^{(1)}}(x_0^{(1)}) \right|^2 \right)^{1/2} \right), \\
&= (1/V_{\min}) \left(\sum_{k=0}^{N_1-1} (\delta_k^{(1)})^2 \right)^{1/2} \cdot (1+o(1)) = r \cdot (1+o(1)) / V_{\min}
\end{aligned}
$$

where the first inequality follows from the triangular inequality, and the last equality follows from (3.17). Therefore, we finally have

$$
\left\| x_l^{(1)} \right\| \le 2 \cdot r \cdot (1+o(1)) / V_{\min}
$$

for all l. It is easy to see that this holds for $\widetilde{x_l}^{(1)}$ as well. So the first statement for $m = 1$ is proved.

Now we turn to prove the second statement for $m = 1$. By (3.13), we have

$$
\begin{aligned}
&\left| f_{v_{i_l}^{(1)}}(x_{N_1}^{(1)}) - T^{(1)} \right| \\
&\le (1+o(1)) \sum_{k=0}^{N_1-1} \left|\Delta_k^{(1)}\right| \cdot 2\beta / \left\| v_{i_k}^{(1)} \right\|^2 + |S_1|\,\tau \\
&\le (1+o(1)) \left(\left(\sum_{k=0}^{N_1-1} \delta_k^{(1)} \right) + N_1^2 \tau \right) \cdot 2\beta / V_{\min}^2 + |S_1|\,\tau \\
&\le (1+o(1)) \left(\sum_{k=0}^{N_1-1} \left(\delta_k^{(1)}\right)^2 \right)^{1/2} \sqrt{N} \cdot 2\beta / V_{\min}^2 + o(\sqrt{N}\beta) < \sqrt{N}\beta / V_{\min}^2
\end{aligned} \tag{3.19}
$$

where the second inequality follows from (3.12), the third inequality follows from Cauchy's inequality and the last inequality follows from $\left(\sum_{k=0}^{N_1-1} \left(\delta_k^{(1)}\right)^2 \right)^{1/2} \le r < 0.5$.

Finally, we begin proving the third statement for $m = 1$. For the sake of contradictory, assume $N_1 < N$, i.e. some v_i is not queried in round 1. Through similar analysis as we did in (3.9),

$$\begin{aligned}&\left|f_{v_i}(x_{N_1}^{(1)}) - f_{v_i}(x_0^{(1)})\right| \\ &\le \sum_{k\in S_1} \frac{\|v_i\|\sqrt{p}}{\sigma_{\min}(A)} \left(\frac{2d^{-5}\left|\Delta_k^{(1)}\right|}{\left\|v_{i_k}^{(1)}\right\|^2} \cdot (1+u_F) + 2d^{-5}(1+u_F) \right) + \sum_{k=0}^{N_1-1} \left|\Delta_k^{(1)}\right| \cdot 2\beta / \left\|v_{i_k}^{(1)}\right\|^2. \\ &= (1+o(1)) \sum_{k=0}^{N_1-1} \left|\Delta_k^{(1)}\right| \cdot 2\beta/V_{\min}^2 + |S_1|\,\tau\end{aligned}$$

We continue to assert that by similar process in (3.19), we have

$$\begin{aligned}\left|f_{v_i}(x_{N_1}^{(1)}) - f_{v_i}(x_0^{(1)})\right| &\le (1+o(1)) \sum_{k=0}^{N_1-1} \left|\Delta_k^{(1)}\right| \cdot 2\beta/V_{\min}^2 + |S_1|\,\tau \\ &\le (1+o(1)) \left(\sum_{k=0}^{N_1-1} \left(\delta_k^{(1)}\right)^2 \right)^{1/2} \sqrt{N} \cdot 2\beta/V_{\min}^2 + o(\sqrt{N}\beta) < \sqrt{N}\beta/V_{\min}^2\end{aligned}$$

where the last inequality follows from $\left(\sum_{k=0}^{N_1-1}\left(\delta_k^{(1)}\right)^2\right)^{1/2} \le r < 0.5$. Since $x_0^{(1)} = \mathbf{0}$, we have $f_{v_i}(x_0^{(1)}) = -\gamma_{v_i} > -\gamma_{\max}$, and

$$f_{v_i}(x_{N_1}^{(1)}) > f_{v_i}(x_0^{(1)}) - \sqrt{N}\beta/V_{\min}^2 > -\gamma_{\max} - \sqrt{N}\beta/V_{\min}^2$$

With $T^{(1)} = -\left(\gamma_{\max} + 2\sqrt{N}\beta/V_{\min}^2\right)$, consider (3.19) and yield

$$f_{v_{i_l}^{(1)}}(x_{N_1}^{(1)}) < T^{(1)} + \sqrt{N}\beta/V_{\min}^2 = -\gamma_{\max} - \sqrt{N}\beta/V_{\min}^2$$

for all $0 \le l \le N_1 - 1$. However, given that v_i is not queried, we must have $\max_{l=0,1\ldots N_1-1} f_{v_{i_l}^{(1)}}(x_{N_1}^{(1)}) \ge f_{v_i}(x_{N_1}^{(1)})$, contradictory! Now proof of the third statement is complete.

Now we begin the induction step: Let $2 \le m \le m_0 = -5\log N/\log(7N\beta/V_{\min}^2)$. Assume the three statements are true for $m-1$, we will prove that they are still true for m. We start with the first statement. For any $l \in [N_m]$, by Lemma 2,

(3.20)

$$\begin{aligned}\max\left\{\left\|x_l^{(m)} - x_0^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)} - x_0^{(m)}\right\|\right\} &\le \left\|\sum_{k=1}^{N_m}\left(\widetilde{x_k}^{(m)} - x_{k-1}^{(m)}\right)\right\| + \sum_{k\in S_m}\left\|\widetilde{x_k}^{(m)} - x_k^{(m)}\right\| \\ &\le (2+o(1)) \cdot \left(\sum_{k=1}^{N_m} (\Delta_k^{(m)})^2 / \left\|v_{i_k}^{(m)}\right\|^2\right)^{1/2} + \sum_{k\in S_m}\left\|\widetilde{x_k}^{(m)} - x_k^{(m)}\right\|\end{aligned}$$

The equation (3.11) will be useful in the following analysis, so we paste it here for reference:

$$\sum_{k=0}^{N_m-1} |\Delta_k| \le (1+o(1)) \left(\left(\sum_{k=0}^{N_m-1} \delta_k \right) + N_m \cdot |S_m|\,\tau \right).$$

Combining (3.7) and (3.11) yields
(3.21)

$$\begin{aligned}
&\sum_{k\in S_m}\left\|\widetilde{x_k}^{(m)}-x_k^{(m)}\right\| \\
&\le \frac{\sqrt{p}}{\sigma_{\min}(A)}\left(\left(2d^{-5}/V_{\min}^2\right)\sum_{k\in S_m}\left|\Delta_k^{(m)}\right|\cdot(1+u_F)+2\left|S_m\right|d^{-5}(1+u_F)\right) \\
&\le \frac{\sqrt{p}}{\sigma_{\min}(A)}\left(\left(2d^{-5}/V_{\min}^2\right)(1+u_F)(1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+N_m\cdot\left|S_m\right|\tau\right)+2\left|S_m\right|d^{-5}(1+u_F)\right) \\
&\le \frac{\sqrt{p}}{\sigma_{\min}(A)}(1+o(1))\left(\left(2d^{-5}/V_{\min}^2\right)(1+u_F)\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+2\left|S_m\right|d^{-5}(1+u_F)\right) \\
&\le \frac{2d^{-5}\sqrt{p}}{\sigma_{\min}(A)}(1+o(1))(1+u_F)\left(\left(1/V_{\min}^2\right)\sqrt{N_m}\left(\sum_{k=0}^{N_m-1}\left(\delta_k^{(m)}\right)^2\right)^{1/2}+\left|S_m\right|\right)
\end{aligned}$$

From (3.9) we obtain

$$\begin{aligned}
&\left|f_{v_{i_l}^{(m)}}(x_l^{(m)})-f_{v_{i_l}^{(1)}}(x_0^{(m)})\right| \\
&\le \sum_{k\in S_m}\frac{\left\|v_{i_l}^{(m)}\right\|\sqrt{p}}{\sigma_{\min}(A)}\left(\frac{2d^{-5}\left|\Delta_k^{(m)}\right|}{\left\|v_{i_k}^{(m)}\right\|^2}\cdot(1+u_F)+2d^{-5}(1+u_F)\right)+\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2 \\
&= (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2+\left|S_m\right|\tau \qquad , \\
&\le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+N_m\cdot\left|S_m\right|\tau\right)\cdot 2\beta/V_{\min}^2+\left|S_m\right|\tau \\
&= (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)\cdot 2\beta/V_{\min}^2+\left|S_m\right|\tau\right)
\end{aligned}$$

where the second inequality follows from (3.11) and the last equality holds because $N\beta = o(1)$. Note that every "$o(1)$" does not vary for different m.

Combining this with (3.10) yields

(3.22)

$$\begin{aligned}\left(\sum_{k=1}^{N_m}(\Delta_k^{(m)})^2/\left\|v_{i_k}^{(m)}\right\|^2\right)^{1/2} &\le (1/V_{\min})\left(\left(\sum_{k=0}^{N_m-1}(\delta_k^{(m)})^2\right)^{1/2}+\left(\sum_{l=0}^{N_m-1}\left|f_{v_{i_l}^{(m)}}(x_l^{(m)})-f_{v_{i_l}^{(m)}}(x_0^{(m)})\right|^2\right)^{1/2}\right)\\
&= (1/V_{\min})\left(\sum_{k=0}^{N_m-1}(\delta_k^{(m)})^2\right)^{1/2}+(1/V_{\min})\sqrt{N_m}(1+o(1))\\
&\quad\cdot\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)\cdot 2\beta/V_{\min}^2+|S_m|\,\tau\right)\\
&= (1/V_{\min})\left(\sum_{k=0}^{N_m-1}(\delta_k^{(m)})^2\right)^{1/2}+(1/V_{\min})\sqrt{N_m}(1+o(1))\\
&\quad\cdot\left(\left(\sum_{k=0}^{N_m-1}(\delta_k^{(m)})^2\right)^{1/2}\cdot 2\sqrt{N_m}\beta/V_{\min}^2+|S_m|\,\tau\right)\\
&= (1/V_{\min})(1+o(1))\left(\left(\sum_{k=0}^{N_m-1}(\delta_k^{(m)})^2\right)^{1/2}+|S_m|\cdot\sqrt{N_m}\tau\right).\end{aligned}$$

Still, each "$o(1)$" does not vary for different m. Combining (3.22) with (3.20) and (3.21) yields

(3.23)

$$\begin{aligned}&\max\left\{\left\|x_l^{(m)}-x_0^{(m)}\right\|,\left\|\widetilde{x}_l^{(m)}-x_0^{(m)}\right\|\right\}\\
&\le (1/V_{\min})(2+o(1))\left(\left(\sum_{k=0}^{N_m-1}(\delta_k^{(m)})^2\right)^{1/2}+|S_m|\cdot\sqrt{N_m}\tau\right)\\
&= (1/V_{\min})(2+o(1))\left(\left(\sum_{k=0}^{N_m-1}\left(3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-2}/V_{\min}^2\right)^2\right)^{1/2}+|S_m|\cdot\sqrt{N_m}\tau\right).\\
&\le (1/V_{\min})(2+o(1))\left((3/7)\cdot(7N\beta/V_{\min}^2)^{m-1}+|S_m|\cdot\sqrt{N_m}\tau\right)\end{aligned}$$

where the first equality holds because, with $T^{(m)}=T^{(m-1)}-\left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-2}$ and the statement 2 for round $m-1$,

(3.24)

$$\begin{aligned}\delta_k^{(m)} &\le\left|T^{(m-1)}-T^{(m)}\right|+\left|f_{v_{i_k}^{(m)}}(x_{N_{m-1}}^{(m-1)})-T^{(m-1)}\right|\\
&\le\left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-2}+\left(\sqrt{N}\beta/V_{\min}^2\right)\cdot\left(7N\beta/V_{\min}^2\right)^{m-2}\\
&=3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-2}/V_{\min}^2\end{aligned}$$

for all $0\le k\le N_m-1$. Note that in (3.24), the bound on $\left|f_{v_{i_k}^{(m)}}(x_0^{(m)})-T^{(m-1)}\right|$ also relies on statement 3 for round $m-1$, which guarantees that $v_{i_k}^{(m)}$ will be queried in round $m-1$.

Given the first statement for $m-1$ hold,

$$\begin{aligned}&\left\|x_0^{(m)}\right\| = \left\|x_{N_{m-1}}^{(m-1)}\right\|\\ &\le (\frac{2+o(1)}{V_{\min}})r + (\frac{2+o(1)}{V_{\min}^3})\cdot\sum_{k=2}^{m-1}(7N\beta/V_{\min}^2)^{k-1} + (m-2)(2+o(1))N^{3/2}\tau/V_{\min}\end{aligned}$$

Note that for $m=2$, take $\sum_{k=2}^{m-1}(7N\beta)^{k-1} := 0$. And

(3.25)

$$\begin{aligned}\max\left\{\left\|x_l^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)}\right\|\right\} &\le \left\|x_0^{(m)}\right\| + \max\left\{\left\|x_l^{(m)} - x_0^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)} - x_0^{(m)}\right\|\right\}\\ &\le (\frac{2+o(1)}{V_{\min}})r + (\frac{2+o(1)}{V_{\min}^3})\cdot\sum_{k=2}^{m-1}(7N\beta/V_{\min}^2)^{k-1} + (\frac{2+o(1)}{V_{\min}^3})\cdot\frac{3}{7}(7N\beta/V_{\min}^2)^{m-1}\\ &\quad + (2+o(1))(m-2+1)\cdot N^{3/2}\tau/V_{\min}\\ &< (\frac{2+o(1)}{V_{\min}})r + (\frac{2+o(1)}{V_{\min}^3})\cdot\sum_{k=2}^{m}(7N\beta/V_{\min}^2)^{k-1}\\ &\quad + (2+o(1))(m-1)N^{3/2}\tau/V_{\min}.\end{aligned}$$

With $m \le -5\log N/\log(7N\beta/V_{\min}^2) = O(\log d)$, we have

$$(2+o(1))(m-2)N^{3/2}\tau/V_{\min} \le (2+o(1))\log d\cdot N^{3/2}\tau/V_{\min} := o''(1).$$

Since $o(1)$ does not vary for different m, $o''(1)$ does not vary for different m, either. Therefore, statement 1 is proved for round m:

$$\begin{aligned}\max\left\{\left\|x_l^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)}\right\|\right\} &< (\frac{2+o(1)}{V_{\min}})r + (\frac{2+o(1)}{V_{\min}^3})\cdot\sum_{k=2}^{m}(7N\beta/V_{\min}^2)^{k-1} + o''(1)\\ &< r\cdot(2+o'(1))/V_{\min}\end{aligned}$$

Before proving statement 2, we need to estimate $|S_m|$. Since (d) in Assumption 1, for d sufficiently large (uniform in m),

$$\max\left\{\left\|x_l^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)}\right\|\right\} < (1-\alpha)\frac{1+l_F}{1+u_F}.$$

If $x_l^{(m)} \ne \widetilde{x}_l^{(m)}$, then $x_l^{(m)} \notin R(V)$. And by (3.3), we have

$$\left\|Ax_l^{(m)}\right\|_\infty \le d^{-5}(1+u_F)\left\|\widetilde{x}_l^{(m)}\right\| < d^{-5}(1+l_F)(1-\alpha), \tag{3.26}$$

which shows that once finished running the Refined Projection Algorithm, the current point will be of at least some distance away from $R(V)$. And the projection will not be called until this distance is covered up. To be more specific, given (3.6),

$$\sum_{k=l}^{l+l'-1}\left\|A(\widetilde{x_{k+1}}^{(m)} - x_k^{(m)})\right\|_\infty \le \sum_{k=l}^{l+l'-1} 2\left|\Delta_k^{(m)}\right|\cdot d^{-5}(1+u_F)/V_{\min}^2.$$

Therefore,

$$\begin{aligned}\left\|A(\widetilde{x_{l+l'}}^{(m)})\right\|_\infty &\le \left\|A(x_l^{(m)})\right\|_\infty + \sum_{k=l}^{l+l'-1} 2\left|\Delta_k^{(m)}\right| \cdot d^{-5}(1+u_F)/V_{\min}^2 \\ &< d^{-5}(1+l_F)(1-\alpha) + \sum_{k=l}^{l+l'-1} 2\left|\Delta_k^{(m)}\right| \cdot d^{-5}(1+u_F)/V_{\min}^2\end{aligned},$$

which implies that if $\widetilde{x_{l+l'}}^{(m)} \notin R(V)$ and has to be projected, then

$$\sum_{k=l}^{l+l'-1} 2\left|\Delta_k^{(m)}\right| \cdot (1+u_F)/V_{\min}^2 > (1+l_F)\alpha. \tag{3.27}$$

Recall that $S_m \coloneqq \left\{0 \le k \le N_m - 1 : \widetilde{x_{k+1}}^{(m)} \notin R(V)\right\}$ denotes the number of call for the Refined Projection Algorithm in round m. With the intuition demonstrated above, the following holds:

$$|S_m| \le 1 + \left(\sum_{k=0}^{N_m-1} 2\left|\Delta_k^{(m)}\right| \cdot (1+u_F)/V_{\min}^2\right) / \left((1+l_F)\alpha\right). \tag{3.28}$$

Now we can elaborate our analysis with this bound on $|S_m|$. Recall (3.11), we can easily obtain

$$\begin{aligned}\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| &\le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right) + N_m \cdot |S_m|\,\tau\right) \\ &\le (1+o(1))\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)} + N_m\tau\right) + o(1)\cdot\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\end{aligned},$$

which immediately implies

$$\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| \le (1+o(1))\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)} + N_m\tau\right).$$

Now we turn to prove the second statement for m. By (3.13), we have

$$\begin{aligned}&\left|f_{v_{i_l}^{(m)}}(x_{N_m}^{(m)}) - T^{(m)}\right| \\ &\le (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| \cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2 + |S_m|\,\tau \\ &\le (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| \cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2 + o(\beta)\cdot\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| + \tau \\ &\le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right) + N_m\tau\right)\cdot 2\beta/V_{\min}^2 + \tau \\ &\le (1+o(1))\left(N_m \cdot 3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-2}/V_{\min}^2 + N_m\tau\right)\cdot 2\beta/V_{\min}^2 + \tau \\ &\le (1+o(1))\cdot(6/7)\cdot\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-1}/V_{\min}^2 + (1+o(1))\tau \\ &< (13/14)\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-1}/V_{\min}^2 < \sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-1}/V_{\min}^2\end{aligned} \tag{3.29}$$

where the second inequality uses (3.28), the third inequality follows from (3.11), the fourth inequality follows from (3.24), the fifth inequality holds given $N\beta = o(1)$ and the second last inequality follows given $m \le m_0$.

Finally, we begin proving the third statement. For the sake of contradictory, assume $N_m < N$, i.e. some v_i is not queried in round m. Through similar analysis as we did in (3.9),

$$\begin{aligned}&\left|f_{v_i}(x_{N_m}^{(m)}) - f_{v_i}(x_0^{(m)})\right|\\ &\le \sum_{k\in S_m} \frac{\|v_i\|\sqrt{p}}{\sigma_{\min}(A)}\left(\frac{2d^{-5}\left|\Delta_k^{(m)}\right|}{\left\|v_{i_k}^{(m)}\right\|^2}\cdot(1+u_F) + 2d^{-5}(1+u_F)\right) + \sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2\\ &= (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/V_{\min}^2 + |S_m|\,\tau.\end{aligned}$$

We continue to assert that by similar process in (3.29), we have

$$\begin{aligned}\left|f_{v_i}(x_{N_m}^{(m)}) - f_{v_i}(x_0^{(m)})\right| &\le (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/V_{\min}^2 + |S_m|\,\tau\\ &< \sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-1}/V_{\min}^2\end{aligned}$$

Since statement 3 hold for $m-1$, v_i must have been queried in round $m-1$. Then with statement 2 for $m-1$, $f_{v_i}(x_0^{(m)}) = f_{v_i}(x_{N_{m-1}}^{(m-1)}) > T^{(m-1)} - \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot (7N\beta/V_{\min}^2)^{m-2}$, and

$$\begin{aligned}f_{v_i}(x_{N_m}^{(m)}) &> f_{v_i}(x_0^{(m)}) - \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-1}\\ &> T^{(m-1)} - (1+7N\beta/V_{\min}^2)\left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-2}\end{aligned}$$

With $T^{(m)} = T^{(m-1)} - \left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-2}$, consider (3.29) and yield

$$\begin{aligned}f_{v_{i_l}^{(m)}}(x_{N_m}^{(m)}) &< T^{(m)} + \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-1}\\ &= T^{(m-1)} - (2-7N\beta/V_{\min}^2)\left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-2} < f_{v_i}(x_{N_m}^{(m)})\end{aligned}$$

for all $0 \le l \le N_m - 1$. However, given that v_i is not queried, we must have $\max_{l=0,1\ldots N_m-1} f_{v_{i_l}^{(m)}}(x_{N_m}^{(m)}) \ge f_{v_i}(x_{N_m}^{(m)})$, contradictory! Now proof of the third statement in round m is complete. $\square$

Let $m_1 \coloneqq \lceil -\log N/\log(7N\beta/V_{\min}^2)\rceil$. We would like to recall that the proposed algorithm 1 is designed to do a mandatory Projection at the end of round m_1 no matter $x_N^{(m_1)} \notin R(V)$ or not, i.e. $x_0^{(m_1+1)} = P\left(P(x_N^{(m_1)})/\left\|P(x_N^{(m_1)})\right\|\right)\cdot \left\|P(x_N^{(m_1)})\right\|$. And by (3.7) and (3.8), for all $j \in [N]$, the perturbation caused by

this mandatory projection is
(3.30)

$$\left|f_{v_j}\left(x_N^{(m_1)}\right) - f_{v_j}\left(x_0^{(m_1+1)}\right)\right| \le \frac{V_{\max}\sqrt{p}}{\sigma_{\min}(A)}\left(\frac{2d^{-5}\left|\Delta_{N-1}^{(m_1)}\right|}{V_{\min}^2}\cdot(1+u_F) + 2d^{-5}(1+u_F)\right)$$
$$= \tau\left(1 + \left|\Delta_{N-1}^{(m_1)}\right|/V_{\min}^2\right)$$

To estimate $\Delta_{N-1}^{(m_1)}$, combining (3.12), (3.9) and (3.10) yields

$$\left|\Delta_{N-1}^{(m_1)}\right| \le \delta_{N-1}^{(m_1)} + (1+o(1))\cdot 2\beta/V_{\min}^2\cdot\left(\sum_{k=0}^{N-1}\left|\delta_k^{(m_1)}\right| + N^2\tau\right) + N\tau$$

Recall the bound established in (3.24) we have

$$\left|\Delta_{N-1}^{(m_1)}\right| \le (1+o(1))3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m_1-2}/V_{\min}^2 + (1+o(1))\cdot 2\beta/V_{\min}^2\cdot N^2\tau + N\tau = o(1).$$

Plugging this into (3.30) yields

$$\left|f_{v_j}\left(x_N^{(m_1)}\right) - f_{v_j}\left(x_0^{(m_1+1)}\right)\right| \le (1+o(1))\tau \tag{3.31}$$

Since $m_1 < m_0$, with Theorem 3, we have for all $i \in [N]$:

$$\left|f_{v_i}(x_N^{(m_1)}) - T^{(m_1)}\right| < \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1} \asymp N^{-3/2}.$$

Moreover, if we take a closer look, we will see that in the development of (3.29),

$$\left|f_{v_i}(x_N^{(m_1)}) - T^{(m_1)}\right| < (13/14)\cdot\left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1},$$

we already know that $\tau = o\left(N^{-3/2}\right)$, which, combined with (3.31), results in

$$\left|f_{v_j}\left(x_0^{(m_1+1)}\right) - T^{(m_1)}\right| < \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1}. \tag{3.32}$$

Now we turn to bound $\left\|x_0^{(m_1+1)}\right\|$:

$$\begin{aligned}\left\|x_0^{(m_1+1)}\right\| &= \left\|P\left(P(x_N^{(m_1)})/\left\|P(x_N^{(m_1)})\right\|\right)\right\|\cdot\left\|P(x_N^{(m_1)})\right\| \\ &\le \left\|P(x_N^{(m_1)})/\left\|P(x_N^{(m_1)})\right\|\right\|\cdot\left\|P(x_N^{(m_1)})\right\| \le \left\|x_N^{(m_1)}\right\|,\end{aligned} \tag{3.33}$$

where the last two inequalities hold by the contraction property of the Projection Algorithm.

Proof of Theorem 4.

Proof. We will first prove statement $1 \sim 5$ in order for $m = m_1 + 1$. By (3.3), since $x_0^{(m_1+1)} = P\left(P(x_N^{(m_1)})/\left\|P(x_N^{(m_1)})\right\|\right)\cdot\left\|P(x_N^{(m_1)})\right\|$, we have

$$\left\|Ax_0^{(m_1+1)}\right\|_\infty \le d^{-5}(1+u_F)\left\|x_N^{(m_1)}\right\| < d^{-5}(1+l_F)(1-\alpha) \tag{3.34}$$

Given (3.34) and with the intuition demonstrated in (3.27), we can get a much better bound than (3.28), namely,

$$|S_{m_1+1}| \le \left\lfloor\left(\sum_{k=0}^{N_{m_1+1}-1} 2\left|\Delta_k^{(m_1+1)}\right|\cdot(1+u_F)/V_{\min}^2\right)/\left((1+l_F)\alpha\right)\right\rfloor. \tag{3.35}$$

Now we can elaborate our analysis with this bound on $|S_{m_1+1}|$. Recall (3.11), we can easily obtain

$$\begin{aligned}\sum_{k=0}^{N_{m_1+1}-1}\left|\Delta_k^{(m_1+1)}\right| &\le (1+o(1))\left(\left(\sum_{k=0}^{N_{m_1+1}-1}\delta_k^{(m_1+1)}\right)+N_{m_1+1}\cdot|S_{m_1+1}|\,\tau\right)\\ &\le (1+o(1))\left(\sum_{k=0}^{N_{m_1+1}-1}\delta_k^{(m_1+1)}\right)+o(1)\cdot\sum_{k=0}^{N_{m_1+1}-1}\left|\Delta_k^{(m_1+1)}\right|\end{aligned},$$

which immediately implies

$$\sum_{k=0}^{N_{m_1+1}-1}\left|\Delta_k^{(m_1+1)}\right| \le (1+o(1))\sum_{k=0}^{N_{m_1+1}-1}\delta_k^{(m_1+1)}$$

with some $o(1)$ that doesn't vary for different m. Also, by (3.32),

$$\begin{aligned}\delta_k^{(m_1+1)} &\le \left|T^{(m_1)}-T^{(m_1+1)}\right|+\left|f_{v_{i_k}^{(m_1+1)}}(x_{N_{m_1}}^{(m_1)})-T^{(m_1)}\right|\\ &\le \left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1}+\left(\sqrt{N}\beta/V_{\min}^2\right)\cdot\left(7N\beta/V_{\min}^2\right)^{m_1-1}\\ &=3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m_1-1}/V_{\min}^2.\end{aligned}$$

Therefore,

(3.36)
$$\sum_{k=0}^{N_{m_1+1}-1}\left|\Delta_k^{(m_1+1)}\right| \le (1+o(1))N\cdot 3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m_1-1}/V_{\min}^2=\frac{(3+o(1))}{7\sqrt{N}}$$

and statement 1 for $m=m_1+1$ is proved. Since $\frac{(3+o(1))}{7\sqrt{N}} <= o(1)$, (3.35) implies that $|S_{m_1+1}|=0$, which proves statement 2.

By (3.23), for all $l\in[N_{m_1+1}]$

$$\begin{aligned}&\max\left\{\left\|x_l^{(m_1+1)}-x_0^{(m_1+1)}\right\|,\left\|\widetilde{x}_l^{(m_1+1)}-x_0^{(m_1+1)}\right\|\right\}\\ &\le (1/V_{\min})(2+o(1))\left((3/7)\cdot(7N\beta/V_{\min}^2)^{m_1}+|S_{m_1+1}|\cdot\sqrt{N_{m_1+1}}\tau\right)\\ &=(1/V_{\min})(2+o(1))\cdot(3/7)\cdot(7N\beta/V_{\min}^2)^{m_1}\end{aligned}$$

Also, by (3.33) and statement 1 of Theorem 3,

$$\begin{aligned}\left\|x_0^{(m_1+1)}\right\| &\le \left\|x_N^{(m_1)}\right\|\\ &\le \left(\frac{2+o(1)}{V_{\min}}\right)r+\left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot\sum_{k=2}^{m_1}(7N\beta/V_{\min}^2)^{k-1}+(m_1-1)(2+o(1))N^{3/2}\tau/V_{\min}\\ &=\left(\frac{2+o_{m_1}(1)}{V_{\min}}\right)r+\left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot\sum_{k=2}^{m_1}(7N\beta/V_{\min}^2)^{k-1}\end{aligned}$$

The last equality holds because $m_1 = O(\log d)$. Therefore,

$$\begin{aligned}
&\max\left\{\left\|x_l^{(m_1+1)}\right\|, \left\|\widetilde{x}_l^{(m_1+1)}\right\|\right\} \\
&\leq \left(\frac{2+o_1(1)}{V_{\min}}\right) r + \left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot \sum_{k=2}^{m_1}(7N\beta/V_{\min}^2)^{k-1} + (1/V_{\min})(2+o(1))\cdot(3/7)\cdot(7N\beta/V_{\min}^2)^{m_1} \\
&< \left(\frac{2+o_1(1)}{V_{\min}}\right) r + \left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot \sum_{k=2}^{m_1+1}(7N\beta/V_{\min}^2)^{k-1}
\end{aligned}$$

which proves statement 3.

By (3.13), (3.36) and $|S_{m_1+1}| = 0$:

(3.37)

$$\begin{aligned}
\left|f_{v_{i_l}^{(m_1+1)}}(x_{N_{m_1+1}}^{(m_1+1)}) - T^{(m_1+1)}\right| &\leq (1+o(1)) \sum_{k=0}^{N_{m_1+1}-1} \left|\Delta_k^{(m_1+1)}\right| \cdot 2\beta/\left\|v_{i_k}^{(m_1+1)}\right\|^2 + |S_{m_1+1}|\,\tau \\
&= (1+o(1)) \sum_{k=0}^{N_{m_1+1}-1} \left|\Delta_k^{(m_1+1)}\right| \cdot 2\beta/\left\|v_{i_k}^{(m_1+1)}\right\|^2 \\
&\leq (1+o(1))N\cdot 3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m_1-1}/V_{\min}^2\cdot 2\beta/V_{\min}^2 \\
&< \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1}
\end{aligned}$$

which proves statement 4.

Finally, statement 5 in Theorem 4 follows with exactly the same reasoning as in the proof of statement 3 in Theorem 3 and we will only state the steps without detailed explanation: Assume some v_j is not queried in round m_1+1, then

$$\left|f_{v_j}(x_{N_{m_1+1}}^{(m_1+1)}) - T^{(m_1)}\right| < \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1} + \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1}$$

where the second term is caused by (3.32). With $T^{(m_1+1)} = T^{(m_1)} - \left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1}$ and (3.37),

$$f_{v_{i_l}^{(m_1+1)}}(x_{N_{m_1+1}}^{(m_1+1)}) < T^{(m_1)} - \left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1} + \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1}$$

which contradicts with

$$f_{v_j}(x_{N_{m_1+1}}^{(m_1+1)}) > T^{(m_1)} - \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1-1} - \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m_1}.$$

Now we begin the induction step: Let $m \geq m_1+2$. Assume all five statements are true for $m_1+1, m_1+2, \ldots, m-1$, we will prove that they will still be true for m. Since we already know statement 1 holds for $m-1$,

$$\sum_{j=m_1+1}^{m-1}\sum_{k=0}^{N_j-1}\left|\Delta_k^{(j)}\right| \leq \frac{3+o(1)}{7\sqrt{N}}\cdot\sum_{i=1}^{m-1-m_1}\left(7N\beta/V_{\min}^2\right)^{i-1} < \frac{1}{2\sqrt{N}}.$$

Then, similar to the intuition of (3.35), we have

$$|S_m| \leq \left\lfloor\left(\left(\frac{1}{2\sqrt{N}} + \sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\right)\cdot(1+u_F)/V_{\min}^2\right)/\left((1+l_F)\alpha\right)\right\rfloor. \tag{3.38}$$

Now we can elaborate our analysis with this bound on $|S_m|$. Recall (3.11), we can easily obtain

$$\begin{aligned}\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| &\le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+N_m\cdot|S_m|\,\tau\right)\\ &\le (1+o(1))\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+o(1)\cdot\left(\frac{1}{2\sqrt{N}}+\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\right),\end{aligned}$$

which immediately implies

$$\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| \le (1+o(1))\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+o(1)$$

with some $o(1)$ that doesn't vary for different m. Also, by (3.32),

$$\begin{aligned}\delta_k^{(m)} &\le \left|T^{(m-1)}-T^{(m)}\right|+\left|f_{v_{i_k}^{(m)}}(x_{N_{m-1}}^{(m-1)})-T^{(m-1)}\right|\\ &\le \left(2\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-2}+\left(\sqrt{N}\beta/V_{\min}^2\right)\cdot\left(7N\beta/V_{\min}^2\right)^{m-2}\\ &=3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-2}/V_{\min}^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| &\le (1+o(1))N\cdot 3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-2}/V_{\min}^2+\frac{o(1)}{2\sqrt{N}}\\ &=\frac{(3+o(1))}{7}\sqrt{N}\cdot(7N\beta/V_{\min}^2)^{m-m_1-1}\cdot(7N\beta/V_{\min}^2)^{m_1}+\frac{o(1)}{2\sqrt{N}}\\ &<\frac{(3+o(1))}{7\sqrt{N}}\cdot(7N\beta/V_{\min}^2)^{m-m_1-1}+\frac{o(1)}{2\sqrt{N}}\end{aligned}$$

where the last inequality follows because $m_1 \ge -\log N/\log(7N\beta/V_{\min}^2)$. Combining this with (3.38) yields

$$|S_m| \le \left(\frac{1+o(1)}{2\sqrt{N}}+\frac{3+o(1)}{7\sqrt{N}}\cdot(7N\beta/V_{\min}^2)^{m-m_1-1}\right)\cdot\frac{1+u_F}{(1+l_F)\alpha V_{\min}^2}=o(1)<1. \tag{3.39}$$

Here we still would like to point out that every "$o(1)$" does not vary for different m, so the last inequality holds uniformly for all $m \ge m_1+2$ with sufficiently large d. (3.39) immediately implies statement 2.

Again, with (3.11),

$$\begin{aligned}\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| &\le (1+o(1))\left(\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)+N_m\cdot|S_m|\,\tau\right)\\ &=(1+o(1))\left(\sum_{k=0}^{N_m-1}\delta_k^{(m)}\right)\le\frac{3+o(1)}{7\sqrt{N}}\cdot(7N\beta/V_{\min}^2)^{m-m_1-1}.\end{aligned} \tag{3.40}$$

Then,

$$\begin{aligned}
\sum_{j=m_1+1}^{m}\sum_{k=0}^{N_j-1}\left|\Delta_k^{(j)}\right| &= \sum_{j=m_1+1}^{m-1}\sum_{k=0}^{N_j-1}\left|\Delta_k^{(j)}\right| + \sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right| \\
&\leq \frac{3+o(1)}{7\sqrt{N}}\cdot\sum_{i=1}^{m-1-m_1}\left(7N\beta/V_{\min}^2\right)^{i-1} + \frac{3+o(1)}{7\sqrt{N}}\cdot(7N\beta/V_{\min}^2)^{m-m_1-1} \\
&= \frac{3+o(1)}{7\sqrt{N}}\cdot\sum_{i=1}^{m-m_1}\left(7N\beta/V_{\min}^2\right)^{i-1}.
\end{aligned}$$

which proves statement 1 for round m.

By (3.23), for all $l \in [N_m]$

$$\begin{aligned}
&\max\left\{\left\|x_l^{(m)} - x_0^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)} - x_0^{(m)}\right\|\right\} \\
&\leq (1/V_{\min})(2+o(1))\left((3/7)\cdot(7N\beta/V_{\min}^2)^{m-1} + |S_m|\cdot\sqrt{N_m}\tau\right) \\
&= (1/V_{\min})(2+o(1))\cdot(3/7)\cdot(7N\beta/V_{\min}^2)^{m-1}
\end{aligned}$$

Also, by statement 3 of round $m-1$,

$$\left\|x_0^{(m)}\right\| = \left\|x_N^{(m-1)}\right\| \leq \left(\frac{2+o_1(1)}{V_{\min}}\right)r + \left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot\sum_{k=2}^{m-1}(7N\beta/V_{\min}^2)^{k-1}$$

Therefore,

$$\begin{aligned}
&\max\left\{\left\|x_l^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)}\right\|\right\} \\
&\leq \left\|x_0^{(m)}\right\| + \max\left\{\left\|x_l^{(m)} - x_0^{(m)}\right\|, \left\|\widetilde{x}_l^{(m)} - x_0^{(m)}\right\|\right\} \\
&\leq \left(\frac{2+o_1(1)}{V_{\min}}\right)r + \left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot\sum_{k=2}^{m-1}(7N\beta/V_{\min}^2)^{k-1} + (1/V_{\min})(2+o(1))\cdot(3/7)\cdot(7N\beta/V_{\min}^2)^{m-1} \\
&< \left(\frac{2+o_1(1)}{V_{\min}}\right)r + \left(\frac{2+o(1)}{V_{\min}^3}\right)\cdot\sum_{k=2}^{m}(7N\beta/V_{\min}^2)^{k-1}
\end{aligned}$$

which proves statement 3 of round m. By (3.13), (3.40) and $|S_m| = 0$:

$$\begin{aligned}
\left|f_{v_{i_l}^{(m)}}(x_{N_m}^{(m)}) - T^{(m)}\right| &\leq (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2 + |S_m|\,\tau \\
&= (1+o(1))\sum_{k=0}^{N_m-1}\left|\Delta_k^{(m)}\right|\cdot 2\beta/\left\|v_{i_k}^{(m)}\right\|^2 \\
&\leq (1+o(1))N\cdot 3\sqrt{N}\beta\cdot(7N\beta/V_{\min}^2)^{m-2}/V_{\min}^2\cdot 2\beta/V_{\min}^2 \\
&< \left(\sqrt{N}\beta/V_{\min}^2\right)\cdot(7N\beta/V_{\min}^2)^{m-1}
\end{aligned}$$

which proves statement 4 for round m.

Finally, again, statement 5 for round m in Theorem 4 follows with exactly the same reasoning as in the proof of statement 3 in Theorem 3 and we omit it for brevity. □