The Projection Algorithm

Recall
$$F(x) = (1/d^6)\max\{d^5||Ax||_\infty - 1, \max_{i\in[N]} v_i^T x - \gamma_{v_i}\}$$
is the objective. Let $a_i \sim \mathcal{H}_d, i = 1, \ldots, \lfloor \frac{d}{2} \rfloor$ denote the rows of A.
Let $\bar{x} = x - \text{Proj}(x)$, where $\text{Proj}(x)$ is projection of x to $Ker(A)$.
Then we have that for all $a \in \mathcal{H}_d$,

$$(I - \frac{aa^T}{d})x = (I - \frac{aa^T}{d})(\text{Proj}(x) + \bar{x}) = \text{Proj}(x) - \frac{aa^T}{d}\text{Proj}(x) + (I - \frac{aa^T}{d})\bar{x}$$
$$= \text{Proj}(x) + (I - \frac{aa^T}{d})\bar{x}$$

The last equality holds because $\text{Proj}(x)$ is orthogonal to the rows of A. Therefore

$$\prod_{i=1}^{m}(I - \frac{a_{k_i}a_{k_i}^T}{d})x_0 = \text{Proj}(x_0) + \prod_{i=1}^{m}(I - \frac{a_{k_i}a_{k_i}^T}{d})\bar{x_0}$$

The intuition is that every step the following algorithm proceeds equals to multiplying $I - \frac{a_{k_i}a_{k_i}^T}{d}$ to x for some a_{k_i}, which equals to multiplying $I - \frac{a_{k_i}a_{k_i}^T}{d}$ to $\bar{x}$ and leave $\text{Proj}(x)$ unchanged.

Algorithm 1 The projection algorithm

Input: x
$(u, f) \leftarrow Query(x)$ ▷ u and f are gradient and function value respectively.
while $\|u\|_2 > d^{-1}$ **do** ▷ $\|u\|_2 > d^{-1}$ indicates that $u = d^{-1}a_i$ for some i.
  $x \leftarrow x - \langle x, u\rangle / \|u\|_2$
  $(v, f) \leftarrow Query(x)$
end while
Return x

Convergence Analysis

For any x, $\max_{i\le\lfloor\frac{d}{2}\rfloor} |a_i^T x| = \max_{i\le\lfloor\frac{d}{2}\rfloor} |a_i^T \bar{x}|$ holds. And

$$\max_{i\le\lfloor\frac{d}{2}\rfloor} |a_i^T\bar{x}| = ||A\bar{x}||_\infty \ge \frac{||A\bar{x}||_2}{\sqrt{d}} \ge \frac{\sigma_{\min}|\bar{x}|}{\sqrt{d}}$$

Where $\sigma_{\min}$ is the smallest singular value of A. (Decompose $A = U\Sigma V^T$ and given that $\bar{x} \in Ker(A)^\perp$, the diagonal coefficients acting on $V^T\bar{x}$ can only be nonzero, thus lower bounded by the minimum singular value.) With the following theorem from Terence Tao's matrix book:

Theorem 2.7.1 (Lower bound)**.** *Let $M = (\xi_{ij})_{1\le i\le p; 1\le j\le n}$ be an $n\times p$ Bernoulli matrix, where $1p \le (1-\delta)n$ for some $\delta > 0$ (independent of n). Then with exponentially high probability (i.e. $1 - O(e^{-cn})$ for some $c > 0$), one has $\sigma_p(M) \ge c\sqrt{n}$, where $c > 0$ depends only on δ.*

Take $\delta = \frac{1}{2}$, we have that $\sigma_{\min}$ is greater than $C\sqrt{d}$ for some constant C with high probability. Therefore we have that

$$|\frac{a_{k_i}a_{k_i}^T}{d}\bar{x}| = \frac{|a_{k_i}||a_{k_i}^T\bar{x}|}{d} = \frac{|a_{k_i}|}{d}\max_{i\le\lfloor\frac{d}{2}\rfloor}|a_i^T x| \ge \frac{|a_{k_i}|}{d}\frac{\sigma_{\min}|\bar{x}|}{\sqrt{d}} = \sigma_{\min}\frac{|\bar{x}|}{d} \ge \frac{C}{\sqrt{d}}|\bar{x}|$$

The last inequality holds w.h.p. Therefore

$$|(I - \frac{a_{k_i}a_{k_i}^T}{d})\bar{x}| \le \sqrt{1-\frac{C^2}{d}}|\bar{x}|$$

And the iterations it takes to convergent given ϵ and $|x| = 1$ is

$$\log_{\sqrt{1-\frac{C^2}{d}}}\epsilon = O(d\log(\frac{1}{\epsilon}))$$

Let $R(v_i) := \{x|F(x) = v_i^T x - i\gamma\}$ be the "Realm" of v_i. However, note that once it reaches $\bigcup_{i=1}^{N} R(v_i)$, the projection algorithm cannot proceed anymore. Though an arbitrarily close projection into Ker(A) is not achievable, the output of this algorithm is already close enough for some further applications. The following paragraph is cited from a document that I have recently been working on, which attempts to view all $v_i, i = 1\ldots N$ in $O(N)$ iterations. Once we can see all gradients efficiently under constrained memory, we can optimize $F(x)$ efficiently.

Example of usage in my recent work

"Let $f_v(x) = \langle v, x\rangle - \gamma_v$, and

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

where $A \in \mathbb{R}^{p\times d}$. Let l_F and u_F denote the lower and upper bound respectively, of $\max_{i=1\ldots N} f_{v_i}(x)$. 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

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

Recall that $R(v_{k_1}, v_{k_2}, \ldots, v_{k_n}) := \{x|F(x) = v_{k_1}^T x - k_1\gamma = \cdots = v_{k_n}^T x - k_n\gamma\}$.

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:

$$RP(x) = P\left(P(x)/\left\|P(x)\right\|\right) \cdot \left\|P(x)\right\|,$$

where $P(x)$ is the result of running the projection algorithm starting from x. It is easy to see that for any vector $x \in \mathbb{B}$, $P\left(P(x)/\left\|P(x)\right\|\right) \in R(V) \subseteq U$, and

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

(This result might be further improved, but it seems enough for now.)"