OPT --> MEM

Tuesday, February 11, 2025 4:58 PM

yesterday ~~OPT $\longrightarrow$ SEP~~ ✓

SEP $\longrightarrow$ MEM

$\approx$ GRAD $\longrightarrow$ EVAL

$$\nabla f(x)_i = \lim_{h \to 0} \frac{f(x + h e_i) - f(x)}{h} \quad O(n) \text{ calls.}$$

But what if f is not differentiable, e.g. $f = \delta_K$?

Thm. For an L-Lipschitz convex $f : B_n(0,1) \to \mathbb{R}$
$\nabla^2 f(x)$ exists almost everywhere and
$\mathbb{E}_{B^n}(\|\nabla^2 f\|_F) \le nL$.

Pf. Existence a.e. is a classical theorem of Lebesgue.

$$\int_{B^n} \|\nabla^2 f(x)\|_F \, dx \le \int_{B^n} \mathrm{Tr}(\nabla^2 f(x))\, dx = \int_{B^n} \sum_i \frac{\partial^2 f}{\partial x_i^2} \, dx$$

$$\left(\sqrt{\sum \lambda_i^2} \le \sum \lambda_i, \ \lambda_i \ge 0 \right)$$

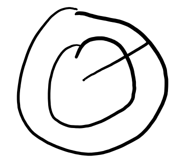
$$= \int_{\partial B^n} \langle \nabla f(x), \vec{n}(x) \rangle \, dx \le \int_{\partial B^n} \|\nabla f(x)\| \, \|\vec{n}(x)\| \, dx$$

$$\le L \, \mathrm{Vol}(\partial B^n)$$

$$\therefore \ \mathbb{E} \ (\|\nabla^2 f(x)\|_F) \le L \cdot \frac{\mathrm{Vol}(\partial B^n)}{} = nL.$$

$$\therefore \; \mathbb{E}_{B^n}\left(\|\nabla^2 f(x)\|_F\right) \le L \cdot \frac{\mathrm{Vol}(\partial B^n)}{\mathrm{Vol}(B^n)} = nL.$$

$$\left(\mathrm{vol}(B^n) = \int_0^1 r^{n-1}\,\mathrm{vol}(\partial B^n)\,dr = \frac{\mathrm{vol}(\partial B^n)}{n}\right).$$



We use this idea to estimate the gradient of any Lipschitz convex function.

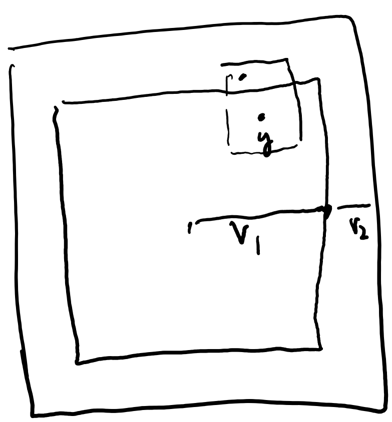
$$B_\infty(x, r) = \{y : \|x-y\|_\infty \le r\}.$$

<u>Lemma.</u> $r_1 \ge r_2 > 0$. f is convex, L-Lipschitz $f : B_\infty(x, r_1+r_2) \to \mathbb{R}$

$$g(y) = \mathbb{E}_{x \sim B_\infty(y, r_2)}\left(\nabla f(x)\right). \qquad \|\nabla f(x)\|_\infty \le L.$$

$$\mathbb{E}_{y \sim B_\infty(x, r_1)} \mathbb{E}_{z \sim B_\infty(y, r_2)} \|\nabla f(z) - g(y)\|_1 \le n^{3/2} \frac{r_2}{r_1} L$$

<u>Pf.</u>



$$w_i(z) = \langle \nabla f(z) - g(y), e_i \rangle$$

$$\int_{B_\infty(y, r_2)} \|\nabla f(z) - g(y)\|_1 = \sum_i \int_{B_\infty(y, r_2)} |w_i(z)|\,dz$$

$$\int_{B_\infty(y, r_2)} w_i(z) = 0.$$

<u>Thm</u> (L_1-Poincaré) $f : \Omega \to \mathbb{R}$

$$\left\| f - \frac{1}{\Omega}\int_\Omega f \right\|_{L^1} \le \sup_{S \subseteq \Omega} \frac{2|S|\,|\Omega \setminus S|}{|\partial S|\,|\Omega|} \cdot \|\nabla f\|_{L^1}$$

$$\left(\|g\|_{L_1} = \int_\Omega \|g\| \right)$$

Applying this to w_i, $\Omega = B_\infty(y, r_2)$

$$\int |w_i(z)| \le r_2 \int \| \nabla^2 f(z) e_i \|_2$$

$$\sum_i \int |w_i(z)| \le r_2 \sqrt{n} \int \|\nabla^2 f\|_F$$

$$\le r_2 \sqrt{n} \int Tr(\nabla^2 f)$$

$$\mathbb{E}_{B_\infty(y,r_2)} \|\nabla f(z) - g(y)\|_1 \le r_2 \sqrt{n} \; \mathbb{E}_{B_\infty(y,r_2)} \sum_i \frac{\partial^2 f(x)}{\partial x_i^2} = r_2 \sqrt{n}\, \mathbb{E}_{B_\infty(y,r_2)} \Delta f$$

Let $h = \frac{1}{(2r_2)^n} \cdot f * \mathbb{1}_{B_\infty(0,r_2)}$. Then $\mathbb{E}_{B_\infty(y,r_2)} \Delta f = \Delta h$.

So, $$\mathbb{E}_{y \sim B_\infty(x,r_1)} \mathbb{E}_{B_\infty(y,r_2)} \|\nabla f(z) - g(y)\|_1 \le r_2 \sqrt{n} \cdot \mathbb{E}_{B_\infty(x,r_1)} \Delta h$$

$$= r_2 \sqrt{n} \; \frac{1}{(2r_1)^n} \cdot \int_{\partial B_\infty(x,r_1)} \langle \nabla h(y), \vec{n}(y) \rangle \, dy$$

$$\le r_2 \sqrt{n} \cdot \frac{1}{(2r_1)^n} \cdot \int_{\partial B_\infty(x,r_1)} \|\nabla h(y)\|_\infty \, \|\vec{n}(y)\|_1 \, dy$$

$$\le r_2 \sqrt{n} \; \frac{1}{(2r_1)^n} L \cdot 1 \cdot (2n)\,(2r_1)^{n-1}$$

$$= n^{3/2} \cdot \frac{r_2}{r_1} \cdot L .$$

Lemma. For $\Omega = \square$ ($\sqrt{2}$)

$$\sup_{S \subseteq \Omega} \frac{2\,|S|\,|\Omega \setminus S|}{|\partial S|\,|\Omega|} = r_2 \ldots$$

Algo: SubGrad(f, x, r_1, ε)

INPUT $r_1 > 0$, $\|\partial f(z)\|_\infty \le L$ $\forall z \in B_\infty(x, 2r_1)$.

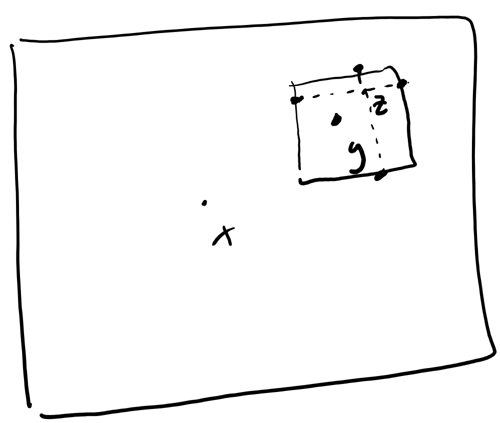
set $r_2 = \sqrt{\dfrac{\varepsilon r_1}{\sqrt{n} L}}$

pick random $y \in B_\infty(x, r_1)$ and $z \in B_\infty(y, r_2)$.

For $i = 1 \ldots n$

- Let α_i, β_i be endpoints of $B_\infty(y, r_2) \cap \{z + s e_i : s \in \mathbb{R}\}$
- set $\tilde{g}_i = \dfrac{f(\beta_i) - f(\alpha_i)}{2 r_2}$

Output $\tilde{g}$.



Lemma. f convex, $\|\nabla f(z)\|_\infty \le L$ $\forall z \in B_\infty(x, 2r_1)$.

EVAL(c) with error $\varepsilon < r_1 \sqrt{n} L$. $\tilde{g} = \text{SubGrad}(f, x, r_1, \varepsilon)$.

Lemma: f convex, ...

EVAL(f) with error $\varepsilon \le r_1 \sqrt{n} L$. $\tilde{g} = \text{SubGrad}(f, x, r_1, \varepsilon)$.

$\exists$ random variable $\xi \ge 0$ with $\mathbb{E}\,\xi \le 2\sqrt{\frac{L\varepsilon}{r_1}}\, n^{5/4}$ s.t.

$\forall q$

$$f(q) \ge f(x) + \langle \tilde{g}, q - x \rangle - \xi \|q - x\|_\infty - 4 n r_1 L$$

Approximate Gradient Oracle!

Pf. Assume f twice-differentiable by viewing as limit of twice-differentiable functions.

$$g(y) = \mathbb{E}_{B_\infty(y, r_2)} (\nabla f)$$

$$\mathbb{E}_z \left| \tilde{g}_i - g(y)_i \right| = \mathbb{E}_z \left| \frac{f(\beta_i) - f(\alpha_i)}{2 r_2} - g(y)_i \right|$$

$$\le \mathbb{E}_z \frac{1}{2 r_2} \int_{z + s e_i \in B_\infty(y, r_2)} \left| \frac{df}{dx_i}(z + s e_i) - g(y)_i \right|$$

$$= \mathbb{E}_z \left| \frac{df}{dx_i}(z) - g(y)_i \right|$$

z, $z + s e_i$ are both uniform in $B_\infty(y, r_2)$.

$$\therefore \mathbb{E}_z \|\nabla f(z) - \tilde{g}\|_1 \le \mathbb{E} \|\nabla f(z) - g(y)\|_1 + \mathbb{E} \|\tilde{g} - g(y)\|_1$$

$$\le 2\, \mathbb{E} \|\nabla f(z) - g(y)\|_1 \left(\le 2 n^{3/2} \cdot \frac{r_2}{r_1} L \quad \text{taking } \mathbb{E}_y \text{ as well} \right)$$

Hence ... of f

(taking $\mathbb{E}_y$ as well)

Hence, by convexity of f

$\forall q, z$

$$f(q) \geq f(z) + \langle \nabla f(z), q - z \rangle$$
$$\geq f(x) - \|\nabla f(z)\|_\infty \|x - z\|_1 + \langle \tilde{g}, q - x \rangle$$
$$+ \langle \nabla f(z) - \tilde{g}, q - x \rangle + \langle \nabla f(z), x - z \rangle$$

$$f(q) \geq f(x) + \langle \tilde{g}, q - x \rangle - \xi \|q - x\|_\infty - 4 r_1 n L$$

$(\|x - z\|_1 \leq (r_1 + r_2) n \leq 2 r_1 n)$

Where $\mathbb{E}\xi \leq 2 n^{3/2} \frac{r_2}{r_1} L$.

If f is evaluated only to $+\varepsilon$ error, then error in $\tilde{g}_i$ can be $+\frac{\varepsilon}{2r_2}$, so $\mathbb{E}\xi \leq 2n^{3/2}\frac{r_2}{r_1}L + \frac{\varepsilon n}{2 r_2}$

$$r_2 = \frac{1}{2}\sqrt{\frac{\varepsilon r_1}{\sqrt{n} L}} < r_1$$

Using this we will design an (approximate) SEP oracle for K, given an (approximate) MEM oracle for K.

MEM_K^δ :
- YES $\quad x \in B(K, \delta) \quad \exists y \in K: \|x - y\| \leq \delta$
- NO $\quad x \notin B(K, -\delta) \quad \exists y \notin K \quad \|x - y\| \leq \delta$

SEP_K^δ :
- YES $\quad$ "
- $\quad a^T x \leq a^T y + \delta\|a\| \quad \forall x \in B(K, -\delta)$.

SEP_K : ...

NO, θ : $\theta^T x \le \theta^T y + \delta\|\theta\| \;\forall\; x \in B(K, -\delta)$.

<u>Thm</u> $B(0, r) \subseteq K \subseteq B(0, R)$. For any $\eta \in [0, \frac{1}{2})$

$$SEP_K^{\eta} = O\left(n \log\left(\frac{nR}{\eta r}\right)\right) MEM_K^{(\eta r / nR)^c}.$$

<u>Cor.</u> $B(0, r) \subseteq K \subseteq B(0, R)$. Convex f, $\varepsilon > 0$.

MEM EVAL.

Then we can find $z \in B(K, \varepsilon)$ s.t.

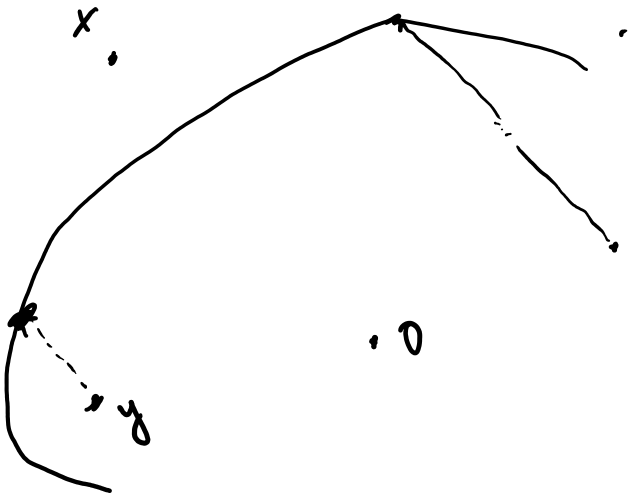
$$f(z) \le \min_K f(x) + \varepsilon \left(\max_K f(x) - \min_K f(x)\right)$$

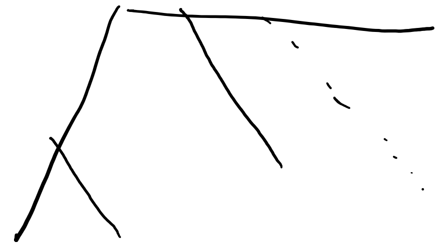
with prob 0.99, using $O\left(n^2 \log^2 \frac{nR}{\varepsilon r}\right)$ calls to MEM/EVAL and $O\left(n^3 \log^{O(1)}\left(\frac{nR}{\varepsilon r}\right)\right)$ arithmetic operations in total.

We need a <u>Lipschitz</u> convex f whose GRAD is SEP for K. δ_K doesn't work, not Lipschitz!

Given $x \notin K$, define

$$h_x : K \to \mathbb{R} \qquad h_x(y) = -\max\{\alpha \mid y + \alpha x \in K\}$$





Lema. 1. h_x is convex.

Pf. y_1, y_2

$y_1 + \alpha_1 x \in K$

$y_2 + \alpha_2 x \in K$

$t \in [0,1]$

$h_x(ty_1 + (1-t)y_2)$ - max α :

$$ty_1 + (1-t)y_2 + \alpha x \in K$$

$$\alpha \geq t\alpha_1 + (1-t)\alpha_2$$

So $h_x(ty_1 + (1-t)y_2) \leq t\, h_x(y_1) + (1-t)\, h_x(y_2)$

Lema 2. $\forall\, \delta < r$ h_x is $\left(\frac{R+\delta}{r-\delta}\right)$-Lipschitz over $B(0,\delta)$.

Pf. Need to show that for $y_1, y_2 \in B(0,\delta)$

$$|h_x(y_1) - h_x(y_2)| \leq L\, \|y_1 - y_2\|.$$

asme $\alpha(y_1) > \alpha(y_2)$.

So $\quad \left| h_x(y_1) - h_x(y_2) \right| = \left(\alpha(y_1) - \alpha(y_2) \right) \|x\|_2$

(1) if $\|y_1 - y_2\| \geq r - \delta$, and $0 \geq h_x \geq -R - \delta$

$$\left| h_x(y_1) - h_x(y_2) \right| \leq R + \delta \leq \left(\frac{R+\delta}{r-\delta} \right) \cdot \|y_1 - y_2\|.$$

(2) Else $\|y_1 - y_2\| < r - \delta$

Let $y_3 = y_1 + \dfrac{y_2 - y_1}{\lambda} \qquad \lambda = \dfrac{\|y_2 - y_1\|}{r - \delta}$

$$\|y_3\| \leq \|y_1\| + r - \delta \leq r \implies y_3 \in K.$$

$\lambda y_3 + (1-\lambda)\left(y_1 + \alpha(y_1) x\right) \quad \in K$

$$y_2 + (1-\lambda)\,\alpha(y_1)\, x \in K \implies \alpha(y_2) \geq (1-\lambda)\,\alpha(y_1)$$
$$\geq \left(1 - \frac{\|y_2 - y_1\|}{r-\delta}\right) \alpha(y_1)$$

$$\left| h_x(y_1) - h_x(y_2) \right| = \left(\alpha(y_1) - \alpha(y_2) \right) \|x\|_2$$
$$\leq \alpha(y_1) \frac{\|y_1 - y_2\|}{r - \delta} \|x\|_2$$
$$\leq \frac{R+\delta}{r-\delta} \cdot \|y_1 - y_2\|.$$

Algo. $\quad x \in B(K, \varepsilon) \longrightarrow$ YES ✓

Algo. $x \in B(K, \varepsilon) \to$ YES ✓

$x \notin B(0, R)$ — NO , $\langle y - x, x \rangle \le 0$.

else $\text{SubGrad}(h_x, 0, r_1, 4\varepsilon) \to \tilde{g}$

report H_x: $\langle \tilde{g}, y - x \rangle \le C \cdot n^{7/6} \varepsilon^{1/3} \dfrac{r}{R^{1/3}}$.

Lemma. $K \subseteq H_x$.

Pf. Set $\delta = r/2$ h_x is $\dfrac{R + r/2}{r/2} \le \dfrac{3R}{r}$ - Lipschitz.

$3K \quad \left(\dfrac{R}{r} = K\right)$.

over $B(0, \frac{r}{2})$.

We will set r_1 small enough st. $B_\infty(0, 2r_1) \subseteq B(0, \frac{r}{2})$.

Then $h_x(y) \ge h_x(0) + \langle \tilde{g}, y \rangle - \xi \|y\|_\infty - 12 n r_1 K \quad (*)$

$\forall y \in K$.

Since $\|x\| \le R$, $\dfrac{-x}{K} \in B(0, r) \subseteq K$

and $h_x\left(-\dfrac{x}{K}\right) = h_x(0) - \dfrac{1}{K}\|x\|_2$

$\ge h_x(0) + \langle \tilde{g}, -\dfrac{x}{K} \rangle - \dfrac{1}{K} \xi \|x\|_\infty - 12 n r_1 K$.

$\therefore \quad \langle \tilde{g}, x \rangle \ge \|x\|_2 - \xi \|x\|_\infty - 24 n r_1 K^2$.

$\therefore \quad \langle \tilde{g}, x \rangle \geq \|x\|_2 - \xi \|x\|_\infty - 24 n r_1 K .$

$x \notin B(K, -\varepsilon)$, $B(0, r) \subseteq K \Rightarrow \left(1 - \frac{\varepsilon}{r}\right) K \subset B(K, -\varepsilon)$

$$h_x(0) \geq -\left(1 - \frac{\varepsilon}{r}\right)\|x\|_2 \geq -\|x\|_2 .$$

So $h_x(0) + \langle \tilde{g}, x \rangle \geq -\xi \|x\|_\infty - 12 n r_1 K^2$

$(*) \Rightarrow 0 \geq h_x(y) \geq \langle \tilde{g}, y - x \rangle - 2 \xi R - 24 n r_1 K^2$

$\therefore \quad \langle \tilde{g}, y - x \rangle \leq 2 \xi R + 24 n r_1 K^2$

$\mathbb{E}(\xi) \leq 2 \sqrt{\frac{3K\varepsilon}{r_1}}\, n^{5/4}$. $\qquad r_1 = n^{1/6} \frac{\varepsilon^{1/3} R^{2/3}}{K}$, $\varepsilon \leq r$.

$$\mathbb{E}(\text{RHS}) \leq 31\, n^{7/6} R^{2/3} \varepsilon^{1/3} K . \tag{1}$$

(OP). Deterministic Algorithm?