Summary: Ellipsoid Method

$\min\limits_{x \in Q \subseteq \mathbb{R}^n} f(x)$, V - variation of f

$\frac{R}{r}$ - asphericity of Q

n - dimension

$O\left(n^2 \ln \frac{RV}{r\varepsilon}\right)$ separation oracle calls to solve the problem.

The cost of each iteration: $O(n^2)$ + cost of separation.

<u>Example</u> $\min\limits_{Ax \le b} c^T x$ separation oracle: $O(n \cdot m)$

m - number of inequalities, $A \in \mathbb{R}^{m \times n}$

Total number of arithmetic operation:

$2^{|L|}$

linear dependence on m

$$O\left(n^2(n^2 + nm) \ln \boxed{\frac{RV}{r\varepsilon}} \right)$$

$n \approx 10\text{-}20$

<u>Theoretical importance</u>

Nemirovski 1976

Khachiyan 1979

Polynomial solvability of LP

Instance: (A, b, c) - integer numbers

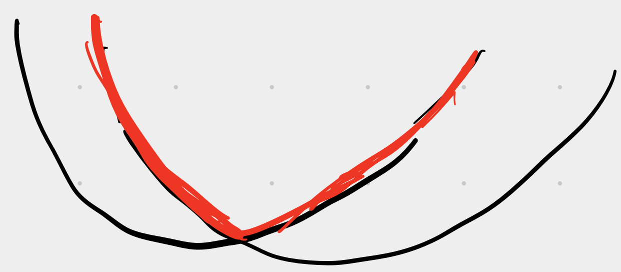
$L = [\ldots | \ldots | \ldots | \ldots]$ $\approx \sum\limits_{i,j} \log_2 A_{ij} + \sum\limits_i \log_2 b_i + \sum\limits_j \log_2 c_j$

Feasibility Problem: <u>$Ax \le b$</u>. $\to$ $\min\limits_x \max\limits_{1 \le i \le m} (a_i^T x - b_i)$

Polynomial: $O(\text{poly}\, |L|)$

Composite Optimization

$$\min_{y\in\mathbb{R}^n}\left[\varphi(y) = \max_{1\le i\le m} f_i(y)\right], \quad f_i \text{ - smooth, convex.}$$



Gradient Steps:

$$\min_{y\in\mathbb{R}^n}\left[\max_{1\le i\le m}\left[f_i(x_k) + \langle f_i'(x_k), y - x_k\rangle\right] + \frac{\alpha}{2}\|y - x_k\|^2\right]$$

m is small $\qquad m = 2 \qquad \min_{y} \max\{f_1(y), f_2(y)\}$

- Multi-obj. optimization
- Mixture of experts

$$\min_{y\in\mathbb{R}^n} \max_{1\le i\le m}\left[\langle a_i, y\rangle - b_i\right] + \frac{\alpha}{2}\|y\|^2 \overset{\text{Dualization}}{=}$$

$$= \min_{y\in\mathbb{R}^n} \max_{\lambda\in\Delta^m} \sum_{i=1}^m \lambda_i\left[\langle a_i, y\rangle - b_i\right] + \frac{\alpha}{2}\|y\|^2$$

$$\lambda \in \mathbb{R}^m_+, \quad \langle e, \lambda\rangle = 1$$

Strong Duality

$$= \max_{\lambda\in\Delta^m} \underbrace{\min_{y\in\mathbb{R}^n} \sum_{i=1}^m \lambda_i\left[\langle a_i, y\rangle - b_i\right] + \frac{\alpha}{2}\|y\|^2}_{g(\lambda)}$$

$$g(\lambda) = -\langle \lambda, b\rangle - \frac{1}{2\alpha}\left\|\sum_{i=1}^m \lambda_i a_i\right\|^2 \text{ - quadratic in } \lambda$$

$m = 2 \Rightarrow$ univariate maximiz. of a concave quadratic

Ellipsoid: $O(n^2 \ln \frac{1}{\varepsilon})$

The Center of Gravity method: $\underline{O(n \ln \frac{1}{\varepsilon})}$ - optimal.

$$G_k = \{ y \in \mathbb{R}^n \mid \langle g_k, x_k - y \rangle \geq 0, \ 0 \leq k \leq k-1 \}$$

$$Vol(E_{k+1}) \leq \exp\left(-\frac{1}{2n}\right) Vol(E_k)$$

$$\underline{x_{k+1} = \frac{1}{Vol(G_k)} \int_{G_k} y \, dy} \ : \ Vol(G_{k+1}) \leq \left(1 - \frac{1}{e}\right) Vol(G_k)$$

impossible to compute in practice.

Large-Scale Optimizations

$n \to +\infty$

$\min_{x \in Q} f(x)$, $Q \subseteq \mathbb{R}^n$ — convex set (simple)

We can project onto it.

f - convex function. $f: \mathbb{R}^n \to \mathbb{R}$.

Stationary Conditions

Directional Derivative:

$$Df(x)[h] = \lim_{\alpha \to 0+} \frac{f(x+\alpha h) - f(x)}{\alpha}$$

- For f differentiable: $Df(x)[h] = \langle f'(x), h \rangle$
- For f convex:

$$Df(x)[h] = \max\{ \langle g, h \rangle : g \in \partial f(x) \}.$$

Proposition Let x^* be a minimum of $\min_{x \in Q} f(x)$.

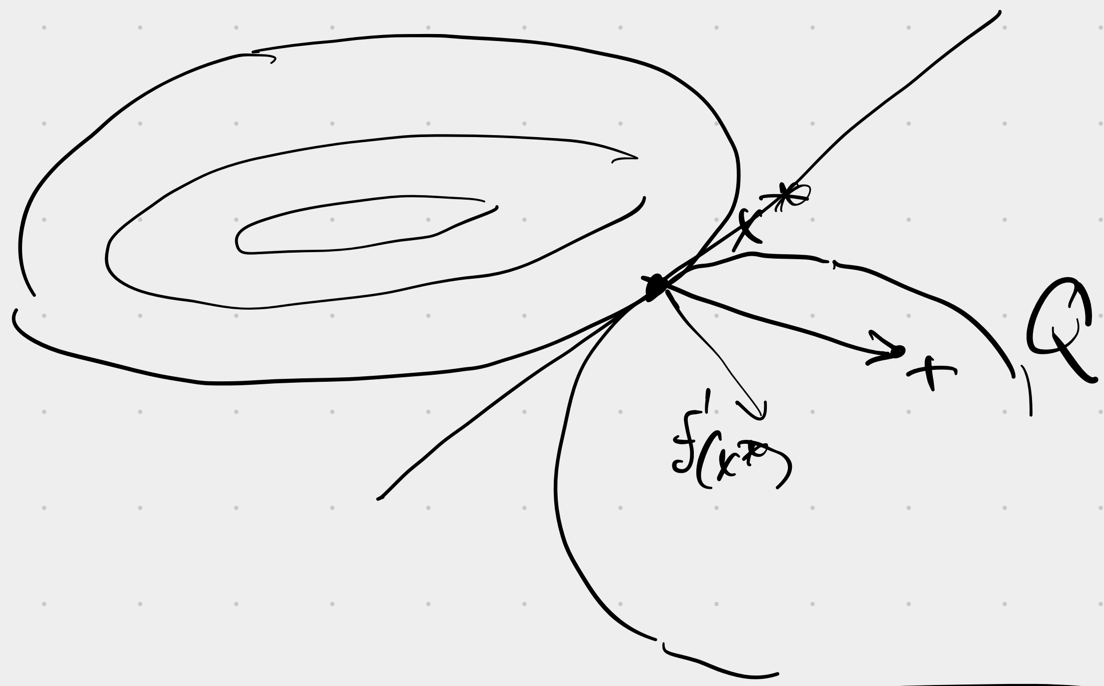
Then $Df(x^*)[x - x^*] \geq 0 \quad \forall x \in Q.$

Proof. For any small $\alpha > 0$:

$$Df(x^*)[x - x^*] = \frac{1}{\alpha}\Big[\underbrace{f(x^* + \alpha(x - x^*)) - f(x^*)}_{\geq 0} \Big] + \bar{o}(1)$$

$\geq o(1)$. Take the limit $\alpha \to 0$. □

Corollary f is differentiable: $\langle f'(x^*), x - x^* \rangle \geq 0 \quad \forall x \in Q$



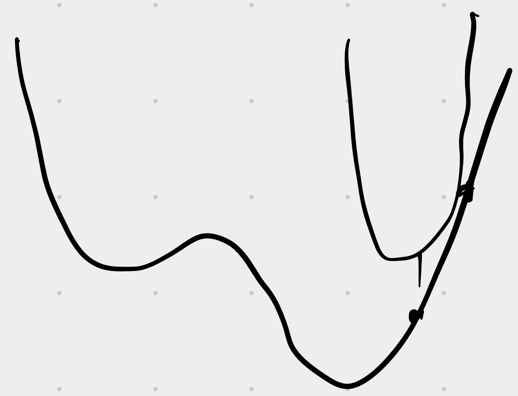
Subgradient step

$x \in Q$

$$x^+ = \arg\min_{y \in Q} \left[m(y) = \eta \langle f'(x), y - x \rangle + \frac{1}{2}\|y - x\|^2 \right] =$$

$\eta > 0$ stepsize. $\qquad f'(x) \in \partial f(x)$

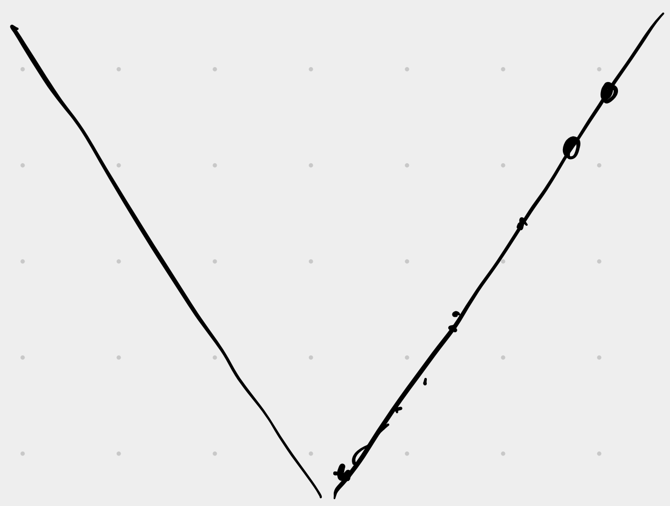
smooth



$$= \pi_Q\left(x - \eta f'(x)\right).$$

How to choose η - ?

Example $\quad f(x) = |x|, \quad x \in \mathbb{R}$



$|f'(x)| = 1 \quad \forall x \neq 0.$

Normalized Stepsizes

Algorithm: Subgradient Method

Init $x_0 \in Q$. Fix a sequence $\{\gamma_k\}_{k\geq 0}$.

Iterate $k \geq 0$:

1. Compute $f'(x_k) \in \partial f(x_k)$
2. $x_{k+1} = \operatorname{argmin}\left\{ m_k(y) = \frac{\gamma_k}{\|f'(x_k)\|}\langle f'(x_k), y - x_k\rangle + \frac{1}{2}\|y - x_k\|^2 \right\}$

Return $\bar{x}_k = \operatorname{argmin}\{ f(y) : y \in \{x_0, \ldots, x_k\}\}$.

Strong convexity of $m_k(y)$.

$$m_k(y) \geq m_k(x_{k+1}) + \underbrace{\langle m_k'(x_{k+1}), y - x_{k+1}\rangle}_{\geq 0} + \frac{1}{2}\|y - x_{k+1}\|^2$$

$$\geq m_k(x_{k+1}) + \frac{1}{2}\|y - x_{k+1}\|^2$$

$$\frac{1}{2}\|y - x_k\|^2 + \frac{\gamma_k}{\|f'(x_k)\|}\langle f'(x_k), y - x_k\rangle \geq \frac{1}{2}\|y - x_{k+1}\|^2 + m_k^*$$

$$m_k^* = \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\gamma_k}{\|f'(x_k)\|}\langle f'(x_k), x_{k+1} - x_k\rangle$$

$$\geq \min_{h \in \mathbb{R}^n} \frac{1}{2}\|h\|^2 + \frac{\gamma_k}{\|f'(x_k)\|}\langle f'(x_k), h\rangle = -\frac{1}{2}\gamma_k^2$$

For every $k \geq 0$:

$$\frac{\gamma_k^2}{2} + \frac{1}{2}\|x_k - y\|^2 \geq \frac{1}{2}\|x_{k+1} - y\|^2 + \gamma_k \Delta_k \quad \forall y$$

where $\Delta_k(y) = \dfrac{\langle f'(x_k), x_k - y\rangle}{\|f'(x_k)\|}$, $\Delta_k = \dfrac{\langle f'(x_k), x_k - x^*\rangle}{\|f'(x_k)\|}$

"measure of optimality"

By convexity: $\geq \dfrac{f(x_k) - f^*}{\|f'(x_k)\|} > 0$

$y := x^*$:

$$\frac{\gamma_k^2}{2} + \frac{1}{2}\|x_k - x^*\|^2 \geq \frac{1}{2}\|x_{k+1} - x^*\|^2 + \gamma_k \Delta_k$$

Telescope for $k \geq 0$ iterations.

Theorem. For any selection of $\{\gamma_k\}_{k\geq 0}$:

$$\underbrace{\sum_{i=0}^{K-1} \gamma_i \Delta_i}_{\text{progress} \to 0} \leq \frac{1}{2}\|x_0 - x^*\|^2 + \underbrace{\frac{1}{2}\sum_{i=0}^{K-1} \gamma_i^2}_{\text{error}}$$

Corollary Constant choice: $\gamma_k \equiv \gamma > 0$

Denote $\bar{\Delta}_K = \frac{1}{K}\sum_{i=0}^{K-1} \Delta_i$. We get:

$$\bar{\Delta}_K \leq \frac{\|x_0 - x^*\|^2}{2\gamma K} + \frac{1}{2}\gamma$$

$$\leq \frac{R^2}{2\gamma K} + \frac{1}{2}\gamma \qquad \text{where } \underline{R \geq \|x_0 - x^*\|}$$

$\xrightarrow{\min_\gamma}$ $\dfrac{R^2}{\gamma^2 K} = 1 \Rightarrow \boxed{\gamma = \dfrac{R}{\sqrt{K}}}$

K is a fixed number of iterations.

We get: $\bar{\Delta}_k \leq \frac{R}{\sqrt{K}}$.

<u>Simple reasoning</u>: $\Delta_k = \frac{\langle f'(x_k), x_k - x^* \rangle}{\|f'(x_k)\|} \geq \frac{f(x_k) - f^*}{\|f'(x_k)\|}$

Assume: $\|f'(x)\| \leq M$ on $x \in Q$. $\Rightarrow \Delta_k \geq \frac{f(x_k) - f^*}{M}$

We get the rate of the Subgradient method:

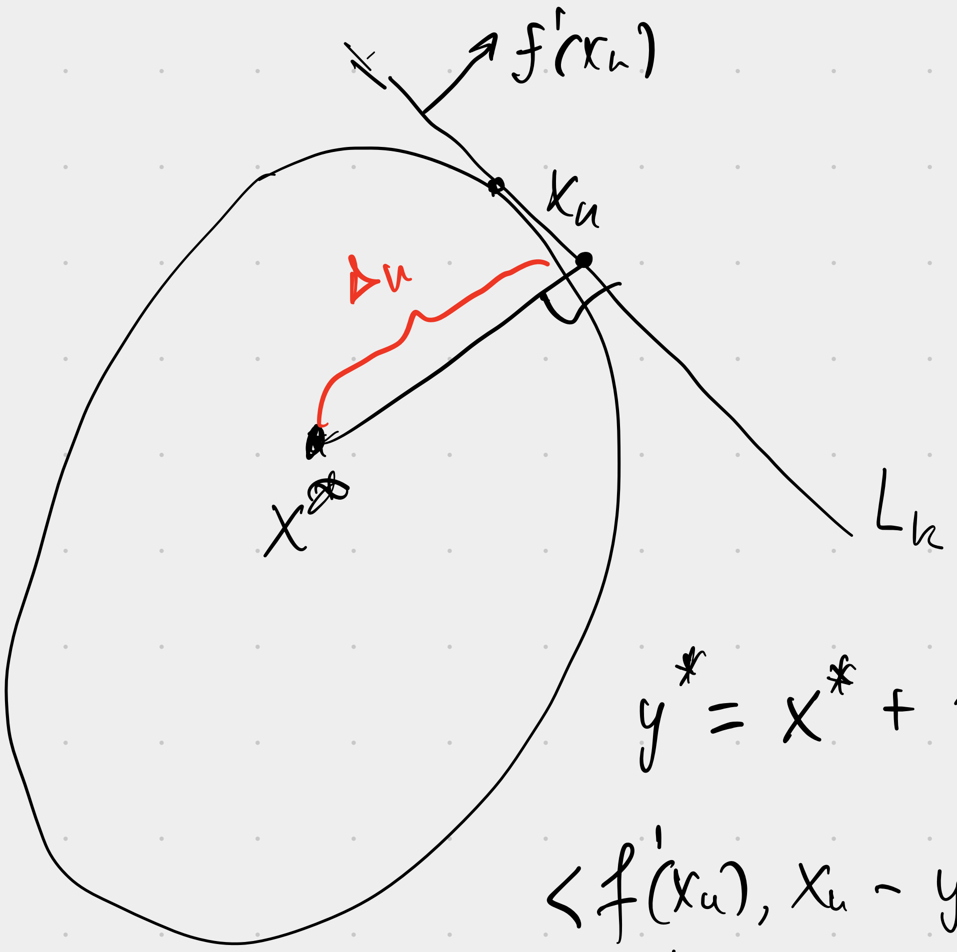
$$f(\bar{x}_k) - f^* \leq \frac{MR}{\sqrt{K}}. \quad \Rightarrow \underline{\text{optimal}}$$

To get $f(\bar{x}_k) - f^* \leq \varepsilon$ $\quad K = \left[\frac{MR}{\varepsilon}\right]^2$.

<u>Geometric Meaning of $\Delta_k \to 0$.</u>

$Q = \mathbb{R}^n$

$L_k = \{ y \in \mathbb{R}^n \mid \langle f'(x_k), x_k - y \rangle \geq 0 \}$



$y^* = x^* + h, \quad h = \frac{f'(x_k)}{\|f'(x_k)\|} \cdot \Delta_k$

$\min_{y \in L_k} \|y - x^*\|$

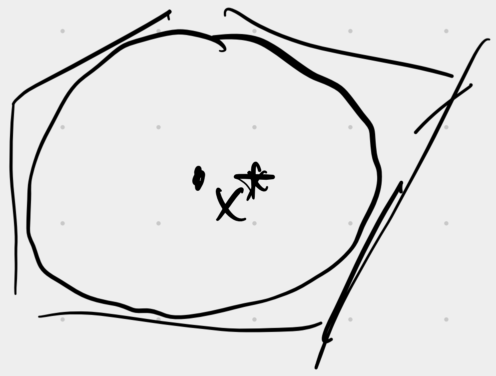
$$\langle f'(x_k), x_k - y^* \rangle =$$
$$= \langle f'(x_k), x_k - x^* \rangle - \langle f'(x_k), h \rangle$$
$$= \|f'(x_k)\| \cdot \Delta_k - \|f'(x_k)\| \cdot \Delta_k = 0$$

Corollary Δ_k is the distance from x^* to the L_k.

Corollary

$$G_{k+1} = \{ y \in \mathbb{R}^n : \langle f'(x_0), x_0 - y \rangle \geq 0, \dots, \langle f'(x_k), x_k - y \rangle \geq 0 \}$$

$x^* \in G_{k+1}$



$$\Delta_k^* = \max \{ r \geq 0 : B_r(x^*) \subseteq G_{k+1} \} \to 0.$$

$\Rightarrow$ Δ_k^* is the maximal radius of the Euclidean ball contained in G_{k+1}.