Application Example:

Min-Max Problems

$$\min_{x \in Q} f(x) \qquad Q \subset \mathbb{R}^n \text{ - bounded set}$$

Assume

$$f(x) = \max_{u \in \Omega} F(x,u)$$

$\Omega \subset \mathbb{R}^m$ - bounded convex set

- $F(\cdot, u)$ - convex for any $u \in \Omega$
- $F(x, \cdot)$ - concave for any $x \in Q$

Primal Problem

$$\min_{x \in Q} f(x) = \min_{x \in Q} \max_{u \in \Omega} F(x,u)$$

$$\geq \max_{u \in \Omega} \min_{x \in Q} F(x,u) =: \max_{u \in \Omega} \varphi(u)$$

weak duality

where $\varphi(u) := \min_{x \in Q} F(x,u)$ — adjoint / dual problem

$$\min_{x \in Q} f(x) = \max_{u \in \Omega} \varphi(u)$$

strong duality

First-order oracle for $f(x) = \max_{u \in \Omega} F(x,u)$

- We can compute

$$u(x) = \operatorname{argmax}_{u \in \Omega} F(x,u)$$

- Function value: $f(x) = F(x, u(x))$
- Subgradient $f'(x) = F_x'(x, u(x))$ - "partial subgradient of F w.r.t. x"

Min-Max Matrix Games

$$F(x,u) = \langle Ax, u\rangle + \langle b, u\rangle + \langle c, x\rangle$$

$A \in \mathbb{R}^{m \times n}$ matrix, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$

$$x \in \Delta_n = \left\{ x \in \mathbb{R}^n_{\geq 0} : \sum_{i=1}^{n} x^{(i)} = 1 \right\}$$

$$u \in \Delta_m = \left\{ u \in \mathbb{R}^m_{\geq 0} : \sum_{i=1}^{m} u^{(i)} = 1 \right\}$$

Our Problem:

$$\min_{x \in \Delta_n} \underbrace{\max_{u \in \Delta_m} \left[\langle Ax, u\rangle + \langle b, u\rangle + \langle c, x\rangle \right]}_{f(x)}$$

$$f(x) = \max_{u \in \Delta_m} \sum_{i=1}^{m} u^{(i)} \left[\langle a_i, x\rangle + b_i \right] + \langle c, x\rangle =$$

$$= \max_{1 \leq i \leq m} \left[\langle a_i, x\rangle + b_i \right] + \langle c, x\rangle$$

$$A = \begin{bmatrix} a_1^T \\ \hline \dots \\ \hline a_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}$$

$$x \mapsto \quad n(x) = e_i \in \Delta_m \quad \text{s.t. } 1 \le i \le m:$$

$$f(x) = \langle a_i, x \rangle + b_i + \langle c, x \rangle$$

$$f'(x) = a_i + c$$

<u>Subgradient method.</u>

$$x_{k+1} = \pi_{\Delta_n}\left(x_k - \eta_k f'(x_k)\right)$$

Normalized stepsizes: $\eta_k = \frac{\gamma}{\|f'(x_k)\|}$, $\gamma := \frac{D}{\sqrt{K}}$

or <u>adaptive stepsizes</u>.

$$f(\bar{x}_k) - f^* \le \frac{MD}{\sqrt{K}}.$$

where $D \ge \|x_0 - x^*\|_2$, $\|f'(x)\|_2 \le M$, $\forall x \in \Delta_n$.

- When to stop? $\longrightarrow$ "Accuracy certificates"
- What about dual/adjoint problem?
- Better norms?

<u>Example</u> $Q = \Delta_n$

$$\operatorname{diam}_{\|\cdot\|_2}(\Delta_n) = \|e_1 - e_2\|_2 = \sqrt{2}.$$

$$f'(x) = a_i = (1, 1, \dots, 1)^T \in \mathbb{R}^n : \quad \|f'(x)\|_2 = \|a_i\|_2 = \sqrt{n}.$$

$M_{\|\cdot\|_2} \geq \sqrt{n}$.

Example $\|\cdot\|_1$-norm for Δ_n.

$$\operatorname{diam}_{\|\cdot\|_1}(\Delta_n) = \|e_1 - e_2\|_1 = 2.$$

For subgradients, we use the dual norm, $\|\cdot\|_\infty \leq \|\cdot\|_2 \leq \|\cdot\|_1$

$$\|(1, \dots, 1)\|_\infty = 1 \quad , \quad M_{\|\cdot\|_\infty} \approx \text{const}$$

$$M_{\|\cdot\|_\infty} \leq M_{\|\cdot\|_2}$$

Motivation

Primal norm $\|\cdot\|$ in $\mathbb{R}^n$

Dual norm: $s \in \mathbb{R}^n$

$$\|s\|_* := \max_{\substack{h \in \mathbb{R}^n \\ \|h\| \leq 1}} \langle s, h \rangle.$$

Gradient Method (for Smooth Functions)

$$x_{k+1} = \arg\min_{y \in Q} \left\{ f(x_k) + \langle f'(x_k), y - x_k \rangle + \overbrace{\frac{L}{2}\|y - x_k\|^2}^{\beta_d(x_k; y)} \right\}$$

- How to solve it? – Difficult for an arbitrary $\|\cdot\|$
- f is non-smooth ($L = +\infty$?)

$$f_k - f_{k+1} \geq \frac{1}{2L} D f_k^2 \to 0$$

Strong convexity of $\frac{1}{2}\|x\|_2^2$. ← Key property analyzing subgradient method

Exercise $d(x) = \|x\|_1^2$, $x \in \mathbb{R}^n$ – not strongly convex for any $n > 1$.

Arbitrary Regularizers.

A distance Function $d : \operatorname{int} Q \to \mathbb{R}_{\geq 0}$

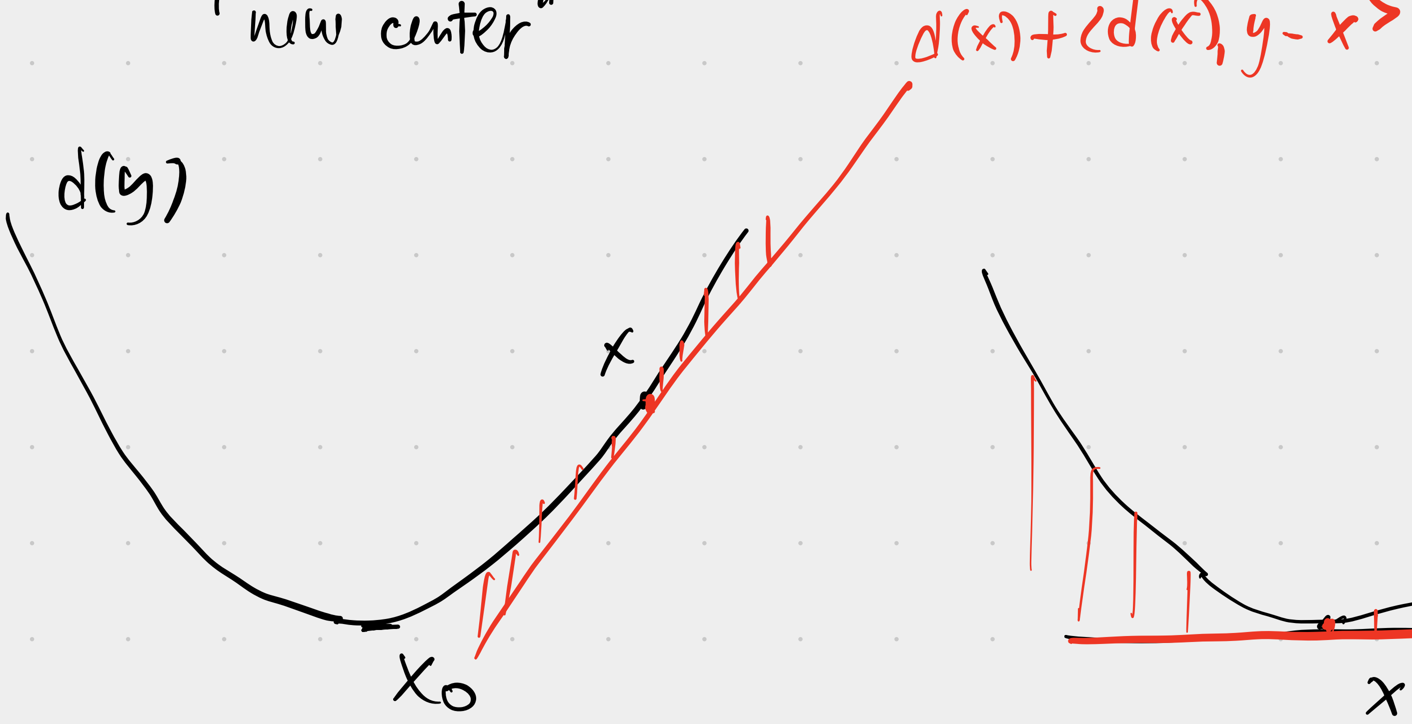
Assumption: d is simple and strictly convex differentiable funct.

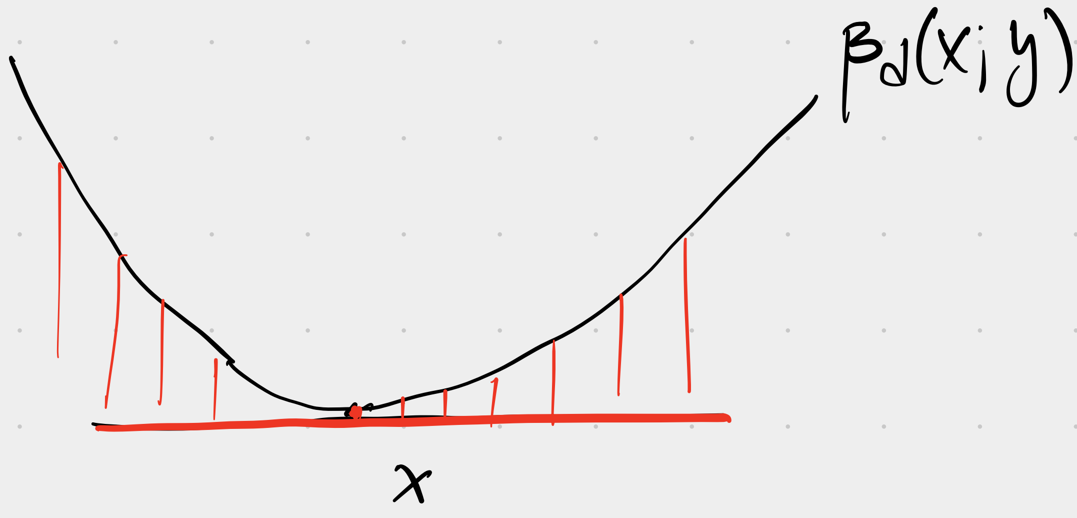
Bregman Divergence

strictly convex

$$\beta_d(x; y) = d(y) - d(x) - \langle d'(x), y - x \rangle > 0 \quad \forall x \neq y$$

"new center"





$x_0 = \arg\min_{y \in Q} d(y)$.

Example $d(x) = \frac{1}{2}\|x\|_2^2$, $x_0 = 0$

$$\beta_d(x;y) = \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|x\|_2^2 - \langle x, y-x\rangle = \frac{1}{2}\|x-y\|^2$$

Example Negative Entropy. $x \in \Delta_n$

$$d(x) = \sum_{i=1}^n x^{(i)} \ln x^{(i)} \quad \text{- convex}$$

$$x_0 = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \in \Delta_n$$

$$d(x_0) = -\ln n.$$

$$\beta_d(x;y) = \sum_{i=1}^n y^{(i)} \ln y^{(i)} - \sum_{i=1}^n x^{(i)} \ln x^{(i)} - \sum_{i=1}^n (1 + \ln x^{(i)})(y^{(i)} - x^{(i)})$$

$$[d'(x)]^{(i)} = \ln x^{(i)} + 1$$

$$= \sum_{i=1}^n y^{(i)} \ln y^{(i)} - \sum_{i=1}^n x^{(i)} \ln x^{(i)} - \sum_{i=1}^n \ln x^{(i)} \cdot (y^{(i)} - x^{(i)})$$

$$= \sum_{i=1}^n y^{(i)} \ln \frac{y^{(i)}}{x^{(i)}} \quad \text{- Kullback-Leibler divergence}$$

$$[d''(x)]^{(i,i)} = \frac{1}{x^{(i)}} > 0 \Rightarrow \text{strictly convex}$$

$$\langle d''(x)h, h\rangle = \sum_{i=1}^{n} \frac{[h^{(i)}]^2}{x^{(i)}}, \quad x \in \operatorname{int} \Delta_n, \; h \in \mathbb{R}^n$$

From Cauchy-Schwarz:

$$\|h\|_1 = \sum_{i=1}^{n} |h^{(i)}| = \sum_{i=1}^{n} \boxed{\frac{|h^{(i)}|}{\sqrt{x^{(i)}}}} \cdot \sqrt{x^{(i)}} \leq$$

$$\leq \left(\sum_{i=1}^{n} \frac{|h^{(i)}|^2}{x^{(i)}} \right)^{1/2} \cdot \underbrace{\left(\sum_{i=1}^{n} x^{(i)} \right)^{1/2}}_{= 1} = \sqrt{\langle d''(x)h, h\rangle}$$

$\Rightarrow$ We proved:

$$\langle d''(x)h, h\rangle \geq \|h\|_1^2 \quad \forall x \in \operatorname{int} \Delta_n, \; h \in \mathbb{R}^n$$

$\Rightarrow$ Function d is strongly convex:

$$\beta_d(x; y) \geq \frac{1}{2}\|y - x\|_1^2 \quad \forall x, y \in \operatorname{int} \Delta_n.$$

Lemma Let f ~~be~~ a convex function.

Consider the regularized objective: $g(y) = f(y) + d(y)$.

Denote $x_g^* = \operatorname{argmin}_{y \in Q} g(y)$. Then

$$g(y) \geq g(x_g^*) + \beta_d(x_g^*; y) \quad \forall y \in Q.$$

Proof. $$\beta_g(x;y) \equiv \underbrace{\beta_f(x;y)}_{\geq 0} + \beta_d(x;y)$$
$$\geq \beta_d(x;y)$$

$$g(y) \geq g(x) + \langle g'(x), y-x \rangle + \beta_d(x;y), \qquad g'(x) \in \partial g(x).$$

Take $x := x_g^*$, $g'(x_g^*) = 0$. $\square$