$$\min_{x \in Q} f(x)$$

$Q \subset \mathbb{R}^n$ - compact convex set

We denote by $\|\cdot\|$ - an arbitrary norm in $\mathbb{R}^n$

Dual norm: $\|\cdot\|_*$

Main complexity parameter:

$$\|f'(x)\|_* \leq M \quad \forall x \in Q \quad f'(x) \in \partial f(x).$$

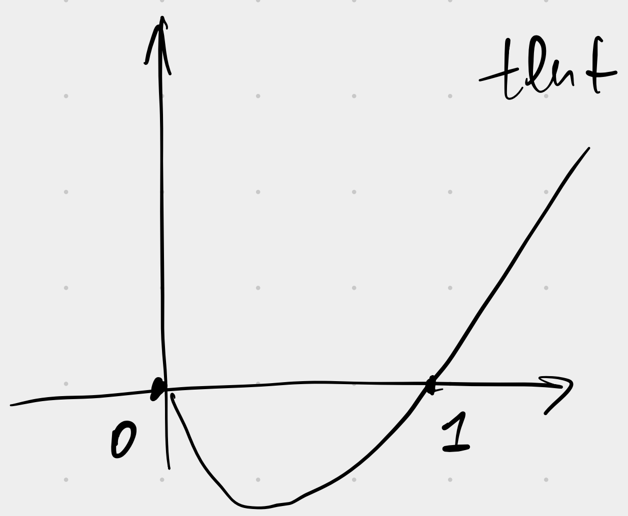
Distance function:

$$d : \operatorname{int} Q \to \mathbb{R}$$

- "simple function"
- d - strictly convex, d is differentiable (ideally, strongly convex w.r.t. $\|\cdot\|$)

$$x_0 = \arg\min_{x \in Q} d(x) \quad \text{- "a center of } Q\text{"}$$

Example Negative Entropy: $d(y) = \sum_{i=1}^n y^{(i)} \ln y^{(i)} + \ln n$



$x_0 = \left[\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right]$

$d(x_0) = 0$

$d(y)$ mesures a distance from $y \in \text{int}\, Q$ to x_0

Bregman Divergence:

$$\beta_d(x;y) = d(y) - d(x) - \langle d'(x), y-x \rangle \geq 0$$

L "a distance" from y to x.

$$\beta_d(x;y) \neq \beta_d(y;x)$$

Strictly convex: $\beta_d(x;y) > 0 \quad , \forall y \neq x.$

Strong convexity: $\beta_d(x;y) \geq \frac{1}{2}\|y-x\|^2$

Main Lemma

Let $g(y) = f(y) + d(y)$, f and d are convex. differentiable

$$x_g^* = \underset{y \in Q}{\operatorname{argmin}}\, g(y)$$

Then: $$g(y) \geq g(x_g^*) + \underbrace{\beta_d(x_g^*;y)}_{\geq \frac{1}{2}\|y-x_g^*\|^2} \quad \forall y \in \text{int}\, Q.$$

when d is strongly convex.

Mirror Descent

x_0; iterate $k \geq 0$:

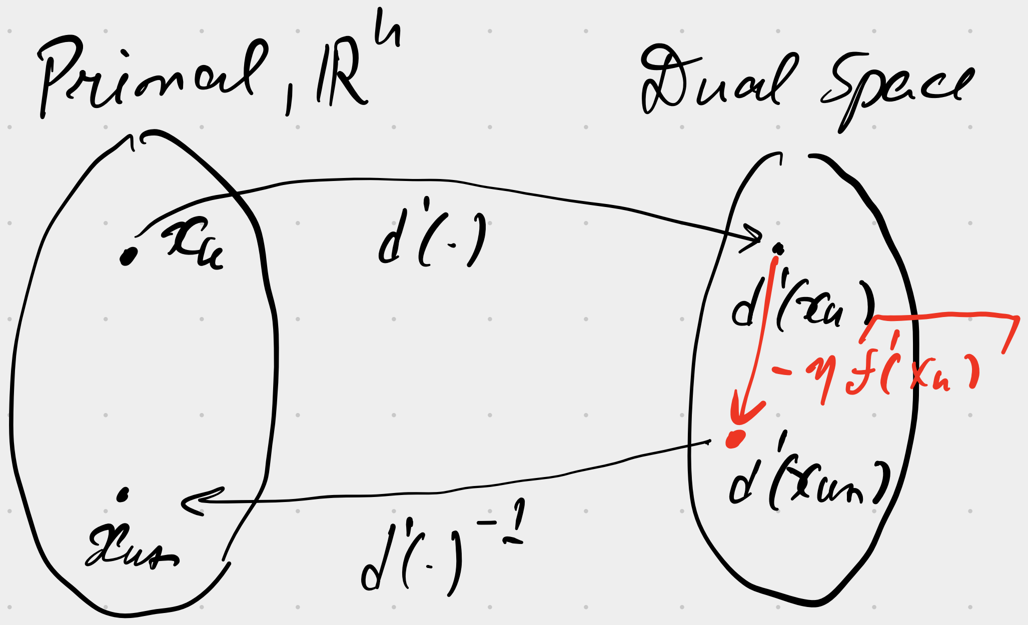
$$x_{k+1} = \operatorname*{argmin}_{y \in Q} \Big[\eta \langle f'(x_k), y - x_k \rangle + \underbrace{\beta_d(x_k; y)}_{\substack{\| \\ d(y) - d(x_k) - \langle d'(x_k), y - x_k \rangle}} \Big],$$

$\eta > 0$ - step-size

Assume for a moment, $Q = \mathbb{R}^n$.

$$d'(x_{k+1}) - d'(x_k) + \eta f'(x_k) = 0$$

$$d'(x_{k+1}) = d'(x_k) - \eta f'(x_k)$$



Euclidean:

$d(x) = \frac{1}{2}\|x\|^2$

$d'(x) = x$

"Dual averaging method"

Exercise Let $Q = \Delta_n$, $d(x) = \sum_{i=1}^{n} x^{(i)} \ln x^{(i)}$.

$$x_{k+1} = \operatorname*{argmin}_{y \in Q} \Big[\eta \langle g_k, y - x_k \rangle + \sum_{i=1}^{n} y^{(i)} \ln \frac{y^{(i)}}{x_k^{(i)}} \Big]$$

We have:

"gradient step" in dual space.

$$x_{k+1}^{(i)} = \frac{x_k^{(i)} \cdot \exp(-\eta g_k^{(i)})}{\sum_{j=1}^{n} x_k^{(j)} \exp(-\eta g_k^{(j)})}.$$

– Multiplicative weight update.

Analysis

$$x_{k+1} = \underset{y \in Q}{\operatorname{argmin}}\ m_k(y) := \underbrace{\eta\langle g_k, y - x_k\rangle + d(y) - d(x_k) - \langle d'(x_k), y - x_k\rangle}$$

Main lemma $\Rightarrow$ $\forall y \in Q$

$$m_k(y) \geq m_k(x_{k+1}) + \beta_d(x_{k+1}; y)$$

$$\begin{aligned} m_k(x_{k+1}) &= \eta\langle g_k, x_{k+1} - x_k\rangle + \beta_d(x_k; x_{k+1}) \\ &\geq -\eta \|g_k\|_* \cdot \|x_{k+1} - x_k\| + \tfrac{1}{2}\|x_k - x_{k+1}\|^2 \\ &\geq \min_{t>0} -\eta\|g_k\|_* \cdot t + \tfrac{1}{2}t^2 = -\frac{\eta^2\|g_k\|_*^2}{2} \geq -\frac{\eta^2 M^2}{2} \end{aligned}$$

Therefore,

$$\frac{\eta^2 M^2}{2} + \beta_d(x_k; y) - \beta_d(x_{k+1}; y) \geq \eta\langle g_k, x_k - y\rangle$$

Telescoping for $k \geq 1$ iterations:

$$\begin{aligned} \eta k \cdot \underbrace{\frac{1}{k}\sum_{i=0}^{k-1}\langle g_i, x_i - y\rangle} &\leq \beta_d(x_0; y) - \beta_d(x_{k+1}; y) + \frac{k \cdot \eta^2 M^2}{2} \\ &\leq \beta_d(x_0; y) + \frac{k\eta^2 M^2}{2}. \end{aligned}$$

Denote

$$D^2 := 2 \cdot \max_{y \in Q} \underbrace{\overset{d(y)-d(x_0)}{\beta_d(x_0; y)}}_{\geq 0}$$

For Simplex

$D^2 = 2\ln n$.

Denote $\mathrm{Gap}_K := \max_{y \in Q} \frac{1}{K} \sum_{i=1}^{K-1} \langle g_i, x_i - y \rangle$.

Theorem For $\eta > 0$:

$$\mathrm{Gap}_K \leq \frac{D^2}{2\eta K} + \frac{\eta M^2}{2} \longrightarrow \min_{\eta > 0}.$$

By choosing $\boxed{\eta := \frac{D}{M\sqrt{K}}}$ we get:

$$\mathrm{Gap}_K \leqslant \frac{MD}{\sqrt{K}} \overset{?}{\leq} \varepsilon \quad \Rightarrow \quad K = \left[\frac{MD}{\varepsilon}\right]^2$$

Accuracy Certificates

Gap_K – "accuracy certificate"!

Convex Functions

, x^* – a solution

$$\mathrm{Gap}_K \geqslant \frac{1}{K} \sum_{i=0}^{K-1} \langle \underset{\partial f(x_i)}{\underset{\cap}{g_i}}, x_i - x^* \rangle \geqslant \frac{1}{K} \sum_{i=0}^{K-1} [f(x_i) - f^*]$$

$$\geqslant f(\bar{x}_K) - f^*, \qquad \bar{x}_K = \frac{1}{K} \sum_{i=0}^{K-1} x_i$$

Max Structure

$$\min_{x \in Q} \left[f(x) := \max_{u \in \Omega} F(x,u) \right]$$

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

$$f(x) = F(x, u(x))$$
$$f'(x) = F'_x(x, u(x))$$

$$\mathrm{Gap}_K = \max_{y \in Q} \frac{1}{K} \sum_{i=0}^{K-1} \langle f'(x_i), x_i - y \rangle$$

Our Problem :

$$f^* = \min_{x \in Q} \overbrace{\max_{u \in \Omega} F(x,u)}^{f(x)} \geq \max_{u \in \Omega} \overbrace{\min_{x \in \Omega} F(x,u)}^{\varphi(u)} = \varphi^*$$

Fixed reasoning:

Note that

$$\mathrm{Gap}_K = \max_{y \in Q} \frac{1}{K} \sum_{i=0}^{K-1} \langle F'_x(x_i, u(x_i)), x_i - y \rangle \geq$$

$$\geq \max_{y \in Q} \frac{1}{K} \sum_{i=0}^{K-1} \Big[\underbrace{F(x_i, u(x_i))}_{f(x_i)} - F(y, u(x_i)) \Big]$$

Missing step

using convexity of $f(x)$

$$\geq \max_{y \in Q} \left[f(\bar{x}_K) - F(y, \bar{u}_K) \right] = f(\bar{x}_K) - \varphi(\bar{u}_K)$$

and concavity of $F(y, \cdot)$.

where $\bar{x}_K = \frac{1}{K} \sum_{i=0}^{K-1} x_i$ and $\bar{u}_K = \frac{1}{K} \sum_{i=0}^{K-1} u(x_i)$

Hence,

$$\varepsilon \geq \mathrm{Gap}_K \geq f(\bar{x}_n) - \varphi(\bar{u}_n) =$$

$$= f(\bar{x}_n) - f^* + f^* - \varphi(\bar{u}_n)$$

$$\geq \underbrace{f(\bar{x}_n) - f^*}_{\geq 0} + \underbrace{\varphi^* - \varphi(\bar{u}_n)}_{\geq 0}$$

Primal Residual   Dual Residual

$\varepsilon \to 0 \Rightarrow$ Strong Duality! $f^* = \varphi^*$.