LEC 4 DSA5103

→ Non-smooth optimization

APG → Accelerated Proximal Gradient

- convex analysis basis
- proximal operator
- PG method

① Definition

a) norm function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$

1. $\|x\| \geq 0$ & $\|x\| = 0 \Leftrightarrow x = 0$
2. $\|\alpha x\| = \alpha \|x\|$
3. $\|x+y\| \leq \|x\| + \|y\|$

b) inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

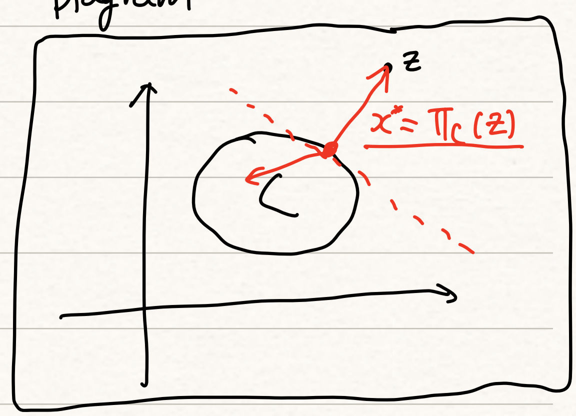
[e.g.] $\langle A, B \rangle = tr(A^T B) = \sum_{i,j} a_{ij} b_{ij}$

$\langle x, y \rangle = x^T y = \sum_i x_i y_i$

② Projection theroem

- a) exists & unique
- b) sufficient & necessary condition

Diagram



C → closed convex set

a) $\Pi_C(z) = \underset{x \in C}{\operatorname{argmin}} \|x - z\|_2^2$

$\boxed{\text{exist \& unique}}$

b) $x^* = \Pi_C(z) \Leftrightarrow \langle z - x^*, x - x^* \rangle \leq 0 \quad \forall x \in C$

(i)

[e.g.] (i) $\boxed{C = \mathbb{R}^n_+ = \{ x \in \mathbb{R}^n : x_i \geq 0,\ i \in [n] \}}$ Positive orthant.

$$\Pi_C(z) = \underline{\max\{z, 0\}}$$

$\hookrightarrow$ n-dimension

(ii)

$$\boxed{C = B_1 := \{ x \in \mathbb{R}^n : \|x\|_2 \leq 1 \}}$$ norm-1 Ball

$$\Pi_C(z) = \frac{z}{\max\{\|z\|_2, 1\}} = \begin{cases} z & \|z\|_2 \leq 1 \\ \frac{z}{\|z\|_2} & \|z\|_2 > 1 \end{cases}$$

(iii)

$\boxed{q_i \to \text{eigenvector}}$ $\to$ $\boxed{C = S^n_+ = \{ X \in \text{Symmetric Matrix} : X \in \text{PSD} \}}$

$\downarrow$

$Q = (q_1, \dots, q_n)$

$$\Pi_C(A) = Q \begin{pmatrix} \max\{\lambda_1, 0\} & & \\ & \ddots & \\ & & \max\{\lambda_n, 0\} \end{pmatrix} Q^T$$

where $A \in S^n$

$\boxed{A = Q \Lambda Q^T}$ $\Leftarrow$ $\boxed{\text{Symmetric matrix}}$ $\boxed{\text{eigenvalue decomposition matrix}}$

③ Defn:

$\boxed{\text{avoid } -\infty}$ since we care about $\min f(x)$

$f : \mathcal{X} \to (-\infty, +\infty]$ extended real-valued function

☆ $\downarrow$ $\boxed{\text{can achieve } +\infty}$

a) $\text{dom}(f) := \{ x : f(x) < +\infty \}$.

b) f is proper $\Leftrightarrow$ $\text{dom}(f) \neq \phi$

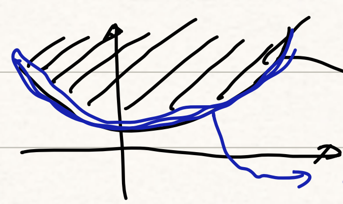
$\downarrow$

kick out $\boxed{f(x) \equiv +\infty}$ (improper case)

c) $\text{epi}(f) := \{ (x, \alpha) \in \mathcal{X} \times \mathbb{R} : f(x) \leq \alpha \}$

c-extra) f is closed

$\Leftrightarrow$ $\text{epi}(f)$ is closed



$\boxed{\text{epi}(f)}$

$f(x)$

d) f is convex $\Leftrightarrow$ epi(f) is convex
(w.r.t function) (w.r.t set)

not extended $\mathbb{R}$-value function

e) convex function $\boxed{f : \mathcal{D} \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}}$

$\downarrow$ extended to $\boxed{\text{convex function } \tilde{f}}$ on $\mathbb{R}^n$

$\tilde{f} : \mathbb{R}^n \longrightarrow \mathbb{R}$

Meaning $\rightarrow$ $\boxed{\text{No need specify domain}}$ where $\tilde{f}(x) = \begin{cases} f(x), & x \in \mathcal{D} \\ +\infty & \text{o/w}. \end{cases}$

$\Rightarrow$ in the following discussion, we can only consider extended real-valued function defined on $\mathbb{R}^n$

④ **Indicator function** $\xrightarrow{\text{duality}}$ $\boxed{\text{Support Function}}$

Indicator function $\longleftrightarrow$ non-empty set C

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

a) $\text{dom}(\delta_C) = C$

b) $\boxed{\text{epi}(f) = C \times [0, +\infty)}$ $\rightarrow$ like a bar

epi(f) closed $\Leftrightarrow$ C closed

$\Leftrightarrow$ $\delta_C(\cdot)$ is closed

(through definition of function closedness)

c) $\text{epi}(f) = C \times [0, +\infty)$

epi(f) convex $\Leftrightarrow$ C convex

$\Leftrightarrow$ $\delta_C(\cdot)$ convex

d) Support function of $C \rightarrow \boxed{\delta_C^*}$

$$\boxed{\delta_C^*(x) = \max_{y \in C} \langle x, y \rangle}$$

$\rightarrow$ Actually it is $\boxed{\text{dual}}$ w.r.t indicator func

$$\delta_C^*(y) = \sup_x \{\langle x, y\rangle - \delta_C(x)\}$$
$$= \max_{x \in C} \langle x, y \rangle$$

⑤ **Cone** $C \subseteq \mathcal{X} \Leftrightarrow x \in C \Rightarrow \lambda x \in C$ for $\forall x \in C$ & $\lambda \geq 0$

Dual Cone C^* of set C (any non-empty one)

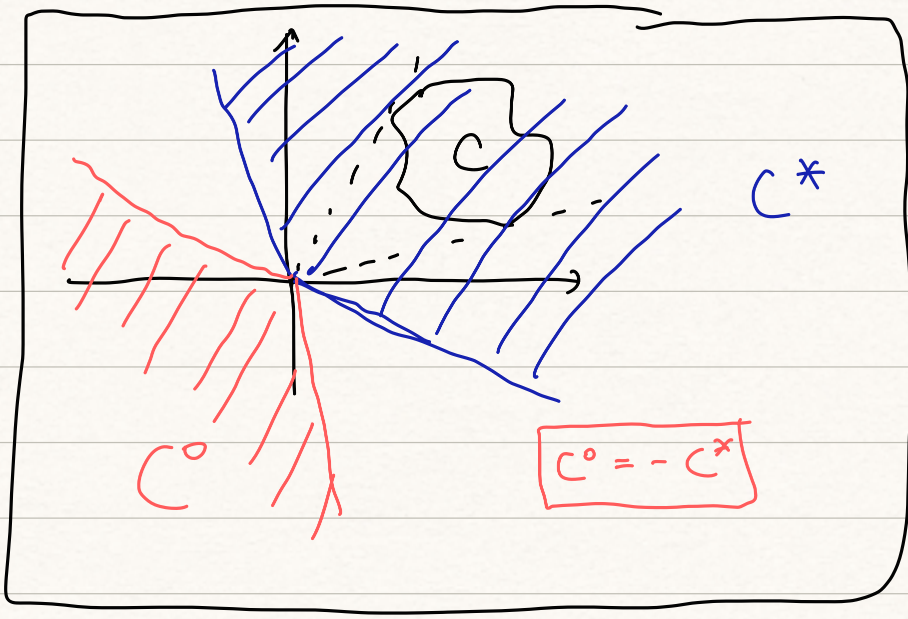
$$C^* = \{ y \in \mathcal{X} : \langle x, y \rangle \geq 0 \quad \forall x \in C \}$$

$\boxed{\text{Remark: } C^* \text{ is alway convex}}$

polar cone $C^\circ := -C^*$

$\boxed{\boxed{\mathbb{R}^n_+} \text{ self-dual cone}}$

if $C^* = C$, then C is self-dual cone



example

[e.g.] 1. $\mathcal{X} = \mathbb{R}^n$, $C = \mathbb{R}^n_+$ is a self-dual closed convex cone

大前提

$\boxed{\text{PSD cone}}$

2. $\mathcal{X} = S^n$, $C = S^n_+$ is a self-dual closed convex cone

大前提

$\boxed{\text{epi}(\|\cdot\|)}$

$\boxed{\text{ice-cream cone}}$

3. second-order cone → self-dual closed convex cone

$\mathcal{X} = \mathbb{R}^n \times \mathbb{R}$ $C = \{ (x, t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_2 \leq t \}$

⑥ Normal Cone for one non-empty set C

$$N_C(\bar{x}) := \{ z \in \mathcal{X} : \langle z, x - \bar{x} \rangle \le 0 \ \ \forall x \in C \}$$

$\boxed{\bar{x} \in C}$ (must on the boundary)

By convention, $N_C(\bar{x}) = \phi$ if $\bar{x} \notin C$

[Prop] if C is non-empty & convex set & $\bar{x} \in C$

a) $N_C(\bar{x})$ closed & convex

b) $\bar{x} \in \text{int}$ then $N_C(\bar{x}) = \{0\}$

c) if C is a cone, then $N_C(\bar{x}) \subseteq C^\circ$

if is non-empty & convex & closed set.

then: $u \in N_C(y) \Leftrightarrow y = \Pi_C(y+u)$

⑦ Sub-differentiable / Sub-gradient

$\downarrow$

$\boxed{\text{deal with non-smoothness}}$

$$v \in \partial f(x) \Leftrightarrow f(z) \geq \underbrace{f(x) + v^T(z-x)}_{\text{linear approximation}} \quad \forall z \in \mathcal{X}$$

By convention, $\partial f(x) = \phi$ for $\forall x \notin \text{dom}(f)$

$\updownarrow$

we only consider $\partial f(x)$ at $x \in \text{dom}(f)$

↓
we don't want $f(x) = +\infty$

[prop]: a) if $f \in C^2$ (differentiable), then $\partial f(x) = \{ \nabla f(x) \}$

b) $\bar{x} \in \underset{x \in X}{\operatorname{argmin}} f(x) \iff 0 \in \partial f(\bar{x})$

if f is convex & proper

⑧ Fenchel Duality

$$f^*(y) = \sup_{x \in X} \{ \langle y, x \rangle - f(x) \} \quad y \in X$$

a) f^* is closed & convex

b) if f closed & proper & convex, then $(f^*)^* = f$

[Prop] $\langle x, y \rangle = f(x) + f^*(y)$

$\iff y \in \partial f(x)$

$\iff x \in \partial f^*(y)$

$$\begin{tikzcd}
 & & \delta_C^* \arrow[dl, leftrightarrow] \\
\Pi_C \arrow[rr, leftrightarrow, "P_{\delta_C} = \Pi_C"] \arrow[d, leftrightarrow, "{u \in N_C(y) \Updownarrow y = \Pi_C(y+u)}"'] & & \delta_C \\
N_C \arrow[urr, leftrightarrow, "\partial \delta_C = N_C"'] & &
\end{tikzcd}$$

Proximal operator

differentiable

→ Moreau- Yosida Regularization

$$\varphi_f(x) = \min_y \{ f(y) + \tfrac{1}{2}\|y - x\|_2^2 \}$$

↓

exist & unique

$$\boxed{\varphi_f(x) \le f(x)}$$

$$\Rightarrow P_f(x) = \arg\min_y \{ y : f(y) + \tfrac{1}{2}\|y - x\|_2^2 \}$$

[Prop] 1. $\nabla \varphi_f(x) = x - P_f(x)$

$$= Q_f(x)$$

2. $$\begin{cases} \arg\min_x f(x) = \arg\min_x \varphi_f(x) \\ \min_x f(x) = \arg\min_x \varphi_f(x) \end{cases}$$

3 $\Pi_C = P_{\delta_C}$

$$\boxed{\varphi_{\delta_C}(x) = \tfrac{1}{2}\|x - \Pi_C(x)\|_2^2} \longrightarrow \text{smooth the } \delta_C(\cdot)$$

[e.g.] $f(x) = \lambda|x|$

$$\varphi_f(x) = \min_y \{ \lambda|y| + \tfrac{1}{2}\|y - x\|_2^2 \}$$

smooth out $\lambda|x|$

$$= \begin{cases} \tfrac{1}{2}x^2 & |x| \le \lambda \\ \lambda|x| - \tfrac{\lambda^2}{2}, & |x| > \lambda \end{cases}$$

$$\boxed{P_f(x) = \begin{cases} x+\lambda & x < -\lambda \\ 0 & |x| \le \lambda \\ x-\lambda & x > \lambda \end{cases}}$$ → proximal operator ↓ $\boxed{\text{soft-threshold}}$

→ M-Y Decomposition

$$x = P_f(x) + P_{f^*}(x)$$

$\boxed{C \rightarrow \text{closed convex cone}}$

[e.g.] $f(x) = \delta_C(x)$

$$f^*(x) = \delta_C^*(x) = \delta_{C^\circ}(x)$$

$$\Rightarrow x = P_{\delta_C}(x) + P_{\delta_{C^\circ}}(x)$$

$$= \Pi_C(x) + \Pi_{C^\circ}(x)$$

$\boxed{\text{PG Method}}$

Recap: Gradient Descent

$$\beta^{(k+1)} = \beta^{(k)} - \alpha \nabla f(\beta^{(k)})$$

$$\Leftrightarrow \tilde{f}_{\beta^{(k)}}(\beta) = \underbrace{f(\beta^{(k)}) + \nabla f(\beta^{(k)})^T(\beta - \beta^{(k)})}_{\text{Taylor term}}$$

$$+ \underbrace{\frac{1}{2} \cdot \frac{1}{\alpha_k} \|\beta - \beta^{(k)}\|_2^2}_{\text{proximal term}}$$

$$\& \quad \beta^{(k+1)} = \underset{\beta}{\operatorname{argmin}}\ \tilde{f}_{\beta^{(k)}}(\beta)$$

$$= \underset{\beta}{\arg\min} \quad \text{Tayor tem}(\beta) + \frac{1}{2\alpha_k}\|\beta - \beta^{(k)}\|_2^2$$

on $\beta^{(k)}$

$$= P_{\alpha_k \cdot (\text{Taylor approx.})}(\beta^{(k)})$$

trick $\Rightarrow$ linearization

$\boxed{\text{PG Framework}}$

$$\min_{\beta} \quad h(\beta) = \boxed{f(\beta)} + g(\beta)$$

smooth

non-smooth

$\boxed{\text{absorb}}$

$$\beta^{(k+1)} = \underset{\beta}{\arg\min} \left\{ \underbrace{\hat{f}_{\beta^{(k)}}(\beta)}_{\text{linear on } (\beta - \beta^{(k)})} + g(\beta) + \frac{1}{2\alpha_k}\|\beta - \beta^{(k)}\|_2^2 \right\}$$

$$= P_{\alpha_k g(\cdot)}\left(\beta^{(k)} - \alpha_k \nabla f(\beta^{(k)})\right)$$

① Calculation of Normal Cone

$$N_C(\bar{x}) := \{ z \in \mathbb{R}^n : \langle z, x - \bar{x} \rangle \leq 0 \quad \forall x \in C \}$$

a) $C = \mathbb{R}^n_+$ $\quad z \in N_C(\bar{x})$

$$\Leftrightarrow \langle z, x - \bar{x} \rangle \leq 0 \quad \forall x \geq 0$$

consider a) if $\bar{x}_i > 0$,

then we can choose $\tilde{x}$ & $\hat{x} \geq 0$

s.t $\tilde{x}_i - \bar{x}_i > 0$, $\hat{x}_i - \bar{x}_i < 0$

$\tilde{x}_j = \bar{x}_j$, $\hat{x}_j = \bar{x}_j$

then $\langle z, \tilde{x} - \bar{x} \rangle \leq 0$

$$\Leftrightarrow z_i \cdot (\hat{x}_i - \bar{x}_i) \leq 0 \Rightarrow z_i \leq 0$$

$$\langle z, \hat{x} - \bar{x} \rangle \leq 0$$

$$\Leftrightarrow z_i (\hat{x}_i - \bar{x}) \leq 0 \Rightarrow z_i \geq 0$$

$\Rightarrow$ $z_i = 0$!

b) if $\bar{x}_i = 0$,

then for $\forall x \geq 0$, we should make sure

$$\langle z, x - \bar{x} \rangle \leq 0$$

$$\Leftrightarrow \sum_i z_i (x_i - \bar{x}_i) \leq 0$$

$$\Leftrightarrow \sum_{\bar{x}_i = 0} z_i (x_i - \bar{x}_i) \leq 0$$

$$\Leftrightarrow \sum_{\bar{x}_i = 0} z_i x_i \leq 0 \qquad x_i \geq 0$$

$$\Leftrightarrow z_i \geq 0$$

This means, $z \in N_C(\bar{x}) \Leftrightarrow z_i = \begin{cases} 0 & \bar{x}_i \neq 0 \\ [0, +\infty) & \bar{x}_i = 0 \end{cases}$

$\Rightarrow$ if $\bar{x} = (1, 1, 0 \dots, 0)$

then $\underline{\underline{N_C(\bar{x}) = \{0\} \times \{0\} \times \mathbb{R}_+^{n-2}}}$

② Calculation of sub-gradient

$f(x) = \|x\|_1 \longrightarrow$ consider $\underline{\underline{\partial f(0)}}$

$z \in \partial f(0)$

$\Leftrightarrow \quad f(x) \geq f(0) + z^T(x - 0) \qquad \forall x \in \mathbb{R}^n$

$\Leftrightarrow \quad \|x\|_1 \geq z^T x \qquad \forall x \in \mathbb{R}^n$

$\Leftrightarrow \quad z^T u \leq 1 \qquad \forall u \in \mathbb{R}^n$ & $\underline{\underline{\|u\|_1 = 1}}$

$\Leftrightarrow \quad \|z\|_\infty \leq 1$

The last "equivalence" comes from:

a) if z satisfies $z^T u \leq 1$ for $\forall \|u\|_1 = 1$

then choose $\underline{u = e_i} \Rightarrow z_i \leq 1 \quad \forall i \in [n]$

$\Rightarrow \|z\|_\infty \leq 1$

b) if z satisfies $\|z\|_\infty \leq 1$.

then $\langle z, u \rangle = \sum_{i=1}^n z_i u_i$

$$\leq \sum_{i=1}^n \|z\|_\infty u_i$$

$$\leq \sum_{i=1}^n |u_i| = \|u\|_1 = 1.$$