Computing proximal maps and projections

7:19 PM

List of projections on simple convex sets

*** Computing proximal maps:**

let $h(z) = I_{\mathcal{X}}(z) = \begin{cases} \infty & z \notin \mathcal{X} \\ 0 & z \in \mathcal{X} \end{cases}$

$\therefore p^* = \begin{pmatrix} \min f_0(x) \\ \text{s.t. } x \in \mathcal{X} \end{pmatrix} = \min f_0(x) + \underbrace{I_{\mathcal{X}}(z)}_{h(z)}$

$\{[\,],\square\}$

So any general convex optimization problem can be solved by taking $x_{k+1} = \text{prox}_{s_k I_{\mathcal{X}}(x)}(x_k - s_k \nabla f_0(x_k)) = [x_k - s_k \nabla f_0(x_k)]_{\mathcal{X}}$ {Eucledian projection of $x_k - s_k \nabla f_0(x_k)$ on $\mathcal{X}$ will give the next point}

now $\text{prox}_{sI_{\mathcal{X}}}(x) = \underset{z}{\text{argmin}}\left(s\,h(z) + \frac{1}{2}\|z-x\|_2^2\right) = \underset{z}{\text{argmin}} \begin{pmatrix} \frac{1}{2}\|z-x\|_2^2, & \text{if } z \in \mathcal{X} \\ \infty, & \text{if } z \notin \mathcal{X} \end{pmatrix} = \underset{z\in\mathcal{X}}{\text{argmin}}\left(\frac{1}{2}\|z-x\|_2^2\right) = \underset{z\in\mathcal{X}}{\text{argmin}}\left(\|z-x\|_2\right) = [x]_{\mathcal{X}}$ {Eucledian projection of x on $\mathcal{X}$}

will result to infeasibility not valid case as we are assuming the optimization problem is feasible.

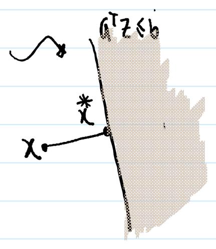
$\therefore \text{prox}_{sI_{\mathcal{X}}}(x) = [x]_{\mathcal{X}} = \text{prox}_{I_{\mathcal{X}}}(x)$ {কারণ s is not present inside the expression}

* Projection onto a halfspace:

$\mathcal{X} = \{x : a^T x \leq b\},\ a \neq 0$

$\therefore \text{prox}_{I_{\mathcal{X}}}(x) = \underset{z\in\mathcal{X}}{\text{argmin}}\ \|z-x\|_2^2 = [x]_{\mathcal{X}}$ now clearly if $x \in \mathcal{X}$ the $[x]_{\mathcal{X}} = x$ কারণ $x, \mathcal{X}$ এ থাকলে x এর উপর তার projection হবে সে নিজেই

now if $x \notin \mathcal{X}$ then $[x]_{\mathcal{X}}$ will be the projection of x on the hyperplane $\{z : a^T z = b\}$



[eq: Eucledian_proj_on_hyperplane] থেকে পাই,

$$[\sigma]_{(a^T\square = b)} = \sigma - \frac{a^T\sigma - b}{\|a\|_2^2}a$$

so for $x \notin \mathcal{X}$, $[x]_{\mathcal{X}} = x - \dfrac{a^T x - b}{\|a\|_2^2}a$

$$\therefore [x]_{(a^T\square \leq b)} = \begin{cases} x, & \text{if } a^T x \leq b \\ x - \dfrac{a^T x - b}{\|a\|_2^2}a, & \text{if } a^T x > b \end{cases}$$

Ans:

• Projection onto the positive orthant:

$\mathcal{X} = \mathbb{R}^n_+ = \{x : \mathbb{R}^n : x \geq 0\}$

$\therefore [x]_{\mathcal{X}} = \underset{z\in\mathcal{X}}{\text{argmin}}\ \|z-x\|_2^2 = \underset{z \geq 0}{\text{argmin}} \sum_{i=1}^{n}(z_i - x_i)^2 = [x]_+$ [eq: Projection on positive orthant]

This can be solved by inspection

if $x_i \geq 0$, we take $z_i^* = x_i$, $z_i^* \geq 0$ ✓

when $x_i < 0$, we cannot take $z_i^* = x_i$, as it would violate $z \geq 0$

so the closest we can get to $x_i < 0$, is if we take $z_i^* = 0$

Combining both: $z_i^* = \max(0, x_i)$ which reflects the decision above

in vector notation: $z^* = \max(0, x) = [x]_+$ //where max is intended elementwise.

Projections onto the standard simplex:

[Projection onto the standard simplex]

$\mathcal{X} = \{x \in \mathbb{R}^n : x \geq 0, \mathbf{1}^T x = 1\}$

$[x]_{\mathcal{X}} = \underset{z}{\text{argmin}}\left\{\frac{1}{2}\|z-x\|_2^2 \;\middle|\; z \geq 0, \mathbf{1}^T z = 1\right\} = z^*$

$\begin{pmatrix} \min \frac{1}{2}\|z-x\|_2^2 \\ \text{s.t. } z \geq 0 \\ \mathbf{1}^T z = 1 \end{pmatrix} = \begin{pmatrix} \min \frac{1}{2}\|z-x\|_2^2 + I_{\mathbb{R}^n_+}(z) \\ \text{s.t. } \mathbf{1}^T z = 1 \end{pmatrix}$

partial lagrangian $L(z,v) = \frac{1}{2}\|z-x\|_2^2 + v(\mathbf{1}^T z - 1) + I_{\mathbb{R}^n_+}(z)$

//remember: for convex optimization problem the dual problem
//associated with a partial lagrangian will result in strong duality
// if Slater's condition holds

$\therefore g(v) = \inf_{z \geq 0} L(z,v)$ // Full Lagrangian হয় তাহলে $\inf_z$ has no constraints

আমরা dual problem টা solve করলে পাবো, $g(v)$ এর argmin $z^*(v)$ কে খুঁজলেই হবে, যেটা this is also the solution of $\nabla_z L(z,v)$

$= \inf_{z \geq 0}\left(\frac{1}{2}\underbrace{\|z-x\|_2^2}_{\sum_{i=1}^n (z_i - x_i)^2} + v\underbrace{(\mathbf{1}^T z - 1)}_{(\sum_{i=1}^n z_i) - 1}\right)$

$= \inf_{z \geq 0}\left(\frac{1}{2}\sum_{i=1}^n (z_i - x_i)^2 + v\left(\sum_{i=1}^n z_i\right) - v\right)$

$= \inf_{z \geq 0}\left[\left(\sum_{i=1}^n \left(\frac{1}{2}(z_i - x_i)^2 + v z_i\right)\right) - v\right]$

$= -v + \inf_{z \geq 0}\left[\underbrace{\left(\sum_{i=1}^n \frac{1}{2}(z_i - x_i)^2 + v z_i\right)}\right]$

note that the objective is seperable in each z_i

$= -v + \sum_{i=1}^n \left(\inf_{z_i \geq 0}\left(\frac{1}{2}(z_i - x_i)^2 + v z_i\right)\right)$ //want to simplify this in z_i

$\frac{1}{2}(z_i - x_i)^2 + v z_i = \frac{1}{2}(z_i^2 + x_i^2 - 2z_i x_i) + v z_i = \frac{x_i^2}{2} + \frac{1}{2}(z_i^2 - 2x_i z_i + 2v z_i) = \frac{x_i^2}{2} + \frac{1}{2}\left(z_i^2 - 2(x_i - v)z_i + (x_i - v)^2\right) - \frac{1}{2}(x_i - v)^2$

$= \underbrace{\frac{1}{2}\left(z_i - (x_i - v)\right)^2} + \underbrace{\frac{1}{2}\left(x_i^2 - (x_i - v)^2\right)}_{\text{constant w.r.t } z}$

$\begin{pmatrix} \min_{z_i} \frac{1}{2}(z_i - x_i)^2 + v z_i \\ \text{s.t. } z_i \geq 0 \end{pmatrix} = \begin{pmatrix} \min_{z_i} \frac{1}{2}\left(z_i - (x_i - v)\right)^2 \\ \text{s.t. } z_i \geq 0 \end{pmatrix}$

now if $x_i - v \geq 0$ then $z_i^*(v) = x_i - v \geq 0$ ✓ (objective square এর minimum পাবো at $z_i = x_i - v \geq 0$ which will make it zero)

if $x_i - v < 0$ then $z_i^*(v) = 0$ (কারণ $z_i^* = x_i - v < 0$ হলে $z_i \geq 0$ violated হবে, in that case the square term will be minimized by setting $z_i = 0$, which will minimize the distance between $\{z_i : z_i \geq 0\}$ and $(x_i - v)$.

Combining both: $\forall_i\ z_i^*(v) = \max\{0, x_i - v\}$ //max is //taken elementwise

$\therefore z^*(v) = \left(\max\{0, x_i - v\}\right) = \max\{0, x - v\mathbf{1}\}$

$= \frac{1}{2}\left(\|z^*(v) - x\|_2^2\right) + v\left(\mathbf{1}^T z^*(v) - 1\right)$ //remember though $g(v)$ is a weird looking function, by construction it is concave (see Boyd)

//now $g(v)$ solve করব which is an unconstrained optimization problem in v

// v^* পেলাম, then $z^* = [x]_{\mathcal{X}} = z^*(v^*)$ //By KKT condition $\nabla_x L(x,v,\lambda) = 0$ solve করে

// প্রাপ্ত $x(v,\lambda)$ এ optimal v^*, λ^* বসালে আমরা primal

//minimizer x^* পেয়ে যাবো.

Now, by KKT condition:

- $\nabla_z L(z,v) = 0 \Rightarrow z(v) = \max\{0, x - v\mathbf{1}\}$ then $\boxed{z^* = [x]_{\mathcal{X}} = \max\{0, x - v^*\mathbf{1}\}}$
- Primal feasibility: $\mathbf{1}^T z(v) = 1$

  $\Leftrightarrow \mathbf{1}^T \max\{0, x - v\mathbf{1}\} = \sum_i \max\{0, x_i - v\} = 1$: এটা solve করে v^* পেয়ে যাবো

  given, given, a scalar equation
- Dual feasibility: নাই
- Complementary slackness: নাই

Similarly when $\mathcal{X} = \{x \mid x \geq 0, \mathbf{1}^T x = b\}$ then $z^* = [x]_{\mathcal{X}} = \max\{0, x - v^*\mathbf{1}\}$

$\sum_{i=1}^n \max\{0, x_i - v_i\} = b$

* Projection onto the Eucledian ball:

$\mathcal{X} = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$

$[x]_{\mathcal{X}} = \underset{z\in\mathcal{X}}{\text{argmin}}\ \|z-x\|_2 = \underset{z\in\mathcal{X}}{\text{argmin}}\ \underbrace{\frac{1}{2}\|z-x\|_2^2}$

optimization problem $\|x\|_2 \leq 1$

optimization problem $\|x\|_2 \le 1$

$$\begin{pmatrix} \min \frac{1}{2}\|z-x\|_2^2 \\ \text{s.t. } \|z\|_2 \le 1 \end{pmatrix}$$

- if $x \in \mathcal{X}$, then $z^* = x$ with 0 objective value which is a norm so obviously that is $z^* = x$ is an argmin. (a norm can have 0 as a absolutely value)
- if $x \notin \mathcal{X} \Leftrightarrow \|x\|_2 > 1$

$\|z\|_2 \le 1 \Leftrightarrow \|z\|_2^2 \le 1$

$$L(z,\lambda) = \frac{1}{2}\|z-x\|_2^2 + \lambda(\|z\|_2^2 - 1)$$

KKT

- $(\nabla_z) \to (z-x) + \lambda 2z = 0 \to (1+2\lambda)z = x \to z = \frac{x}{1+2\lambda}$
- $\lambda \ge 0$
- $\lambda(\|z\|_2^2 - 1) = 0$
- $\|z\|_2^2 \le 1$

$$\frac{\|x\|_2^2}{(1+2\lambda)^2} \le 1 \to \|x\|_2^2 \le (1+2\lambda)^2 \to 1 < \|x\|_2^2 \le (1+2\lambda)^2$$

$$\to (1+2\lambda)^2 > 1 \to \lambda > 0$$

$$\|z\|_2^2 - 1 = 0 \to \|z\|_2^2 = 1 \quad \{z = \frac{x}{1+2\lambda}\}$$

$$\to \frac{\|x\|_2^2}{(1+2\lambda)^2} = 1 \to 1+2\lambda = \|x\|_2 \to \lambda = \frac{\|x\|_2 - 1}{2} = \lambda^*$$

$$\therefore z^* = \frac{x}{1+2\lambda^*} = \frac{x}{1+2\cdot\frac{(\|x\|_2-1)}{2}} = \frac{x}{\|x\|_2} \quad [x > 1]$$

$$[x]_{\mathcal{X}} = \begin{cases} x, & \text{if } \|x\|_2 \le 1 \\ \frac{x}{\|x\|_2}, & \text{if } \|x\|_2 > 1 \end{cases}$$

$\boxed{\text{* Projection onto the } l_1 \text{ norm ball:}}$

$\mathcal{X} = \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}$

$$[x]_{\mathcal{X}} = \underset{\|z\|_1 \le 1}{\operatorname{argmin}} \frac{1}{2}\|z-x\|_2^2$$

(eq: Optimization Problem for l-1 norm projection)

// l_1 norm은 이제 square 처리 안해도 된다!

$$\begin{pmatrix} \min \frac{1}{2}\|z-x\|_2^2 \\ \text{s.t. } \|z\|_1 \le 1 \end{pmatrix} \to L(z,\lambda) = \frac{1}{2}\|z-x\|_2^2 + \lambda(\|z\|_1 - 1)$$

$$= \left(\frac{1}{2}\sum_{i=1}^n (z_i - x_i)^2 + \lambda\sum_{i=1}^n |z_i|\right) - \lambda \quad // \lambda \text{ is a constant w.r.t } z$$

$$= \left(\sum_{i=1}^n \left[\frac{1}{2}(z_i-x_i)^2 + \lambda|z_i|\right]\right) - \lambda$$

\# [illegible]: $\underset{z}{\operatorname{argmin}} L(z,\lambda) = \underset{z}{\operatorname{argmin}} \left(\frac{1}{2}\|z-x\|_2^2 + \lambda\|z\|_1 - \cancel{\lambda}\right) = [\operatorname{shr}_\lambda(x_i)]_{i=1}^n$ #

constant so 빼도 argmin same 하니까

Note that this is what we are trying to find argmin of with \lambda =s_k

$$g(\lambda) = \inf_z \left(\left(\sum_{i=1}^n \left[\frac{1}{2}(z_i-x_i)^2 + \lambda|z_i|\right]\right) - \lambda\right)$$

$$= -\lambda + \inf_z \sum_{i=1}^n \left[\frac{1}{2}(z_i-x_i)^2 + \lambda|z_i|\right]$$

note that this is seperable in z_i

$$= -\lambda + \sum_{i=1}^n \inf_{z_i} \underbrace{\left(\frac{1}{2}(z_i-x_i)^2 + \lambda|z_i|\right)}_{\phi(z_i,\lambda)} = -\lambda + \sum_{i=1}^n \inf_{z_i} \phi(z_i,\lambda) \quad \{\phi(y,\lambda) = \frac{1}{2}(y-x_i)^2 + \lambda|y|\}$$

$$z_i^*(\lambda) = \underset{z_i}{\operatorname{argmin}} \left(\frac{1}{2}(z_i-x_i)^2 + \lambda|z_i|\right) = \underset{z_i}{\operatorname{argmin}}\ \phi(z_i,\lambda) \quad \text{[eq: z\_i^*(\lambda) Orginal]}$$

We will use the identity: $|z| = \max_{|\rho| \le 1} \rho z$

e.g. $|-3| = \max_{-1 \le \rho \le 1} (-3\rho) = \max[-3,3] = 3$; $|3| = \max_{-1\le\rho\le1}(3\rho) = \max[-3,3] = 3$ } so it works!

$\{\forall_{\rho_i}\ \phi(\cdot,\rho_i) = \square + \rho_i\square = \{\text{affine}\}$ $\{D, \square\}$

$\phi(z_i,\rho_i) \to \forall_{z_i}\ \phi(z_i,\cdot) = \underbrace{\frac{1}{2}(z_i-\square)^2}_{D} + \underbrace{\square z_i}_{\text{affine}} = D$ [$\because$ affine + $D = D$] $= \{D, \cup\}\}$

$\Rightarrow$ so ϕ is [concave (affine), $\cup$] in ρ_i the maximizing variable, $[D, \cup]$ in z_i minimizing; maximizing set is compact $\Rightarrow$ Sion's minimax theorem will hold.

$$\min_{z_i} \frac{1}{2}(z_i-x_i)^2 + \lambda|z_i| = \min_{z_i}\left(\frac{1}{2}(z_i-x_i)^2 + \lambda\max_{|\rho_i|\le1}\rho_i z_i\right) = \min_{z_i}\left(\underbrace{\frac{1}{2}(z_i-x_i)^2}_{\text{constant w.r.t } \rho_i} + \max_{|\rho_i|\le1}\lambda\rho_i z_i\right) = \min_{z_i}\max_{|\rho_i|\le1}\left(\frac{1}{2}(z_i-x_i)^2 + \lambda\rho_i z_i\right) = \max_{|\rho_i|\le1}\min_{z_i}\left(\frac{1}{2}(z_i-x_i)^2 + \lambda\rho_i z_i\right)$$

so 밖으로 뺄 수 있다

$\in \mathbb{R}\{D\}$ $\{D, [\le]\}$

\# Sion's minimax theorem:

- $(X \in \{D, [\le], \subseteq \mathbb{R}^n\},\ Y \in \{D, \subseteq \mathbb{R}^m\},\ \phi \in : X \times Y \to \mathbb{R},\ \forall_{y \in Y}\ \phi(\cdot, y) \in \{D_x, \cup_x\},\ \forall_{x\in X}\ \phi(x,\cdot) \in \{\square_y, \cup_y\}) \Rightarrow \sup_{y\in Y}\min_{x\in X}\phi(x,y) = \min_{x\in X}\sup_{y\in Y}\phi(x,y)$
- $(X \in \{D, \subseteq \mathbb{R}^n\},\ Y \in \{D, [\le], \subseteq \mathbb{R}^m\},\ \phi \in : X\times Y \to \mathbb{R},\ \forall_{y\in Y}\ \phi(\cdot,y) \in \{D_x, \cup_x\},\ \forall_{x\in X}\ \phi(x,\cdot) \in \{\square_y, \cup_y\}) \Rightarrow \max_{y\in Y}\inf_{x\in X}\phi(x,y) = \inf_{x\in X}\max_{y\in Y}\phi(x,y)$ # [eq: Sion's mini-max theorem]

function of x (y fixed); maximizer variable; function of y (x fixed); minimizer variable

//mnemonic:
// minimizer variable → function은 convex
// maximizer variable → function은 concave 해야
// 둘 중 one of the set compact 해야 Sion's minimax theorem
// 성립한다.

$$= \max_{|\rho_i|\le1}\left(\min_{z_i}\ \frac{1}{2}(z_i^2 + x_i^2 - 2z_i x_i) + \lambda\rho_i z_i\right)$$

\# $\frac{1}{2}(z_i^2 + x_i^2 - 2z_i x_i + 2\lambda\rho_i z_i)$ constant w.r.t z_i

$$= \frac{1}{2}\left(z_i^2 - 2(x_i - \lambda\rho_i)z_i + (x_i - \lambda\rho_i)^2\right) + \frac{x_i^2}{2} - \frac{1}{2}(x_i - \lambda\rho_i)^2$$

$$= \underbrace{\frac{1}{2}\left(z_i - (x_i - \lambda\rho_i)\right)^2}_{\text{function to be min}} + \underbrace{\left(\frac{x_i^2}{2} - \frac{1}{2}(x_i-\lambda\rho_i)^2\right)}_{\text{constant w.r.t } z_i}$$

so z_i minimum achieve 하려면

$\min = z_i^*(\rho,\lambda) = x_i - \lambda\rho_i$ with 0 norm^2 value ... [eq: z^*(\lambda)]

optimum objective

$$= \max_{|\rho_i|\le1}\left[\frac{x_i^2}{2} - \frac{1}{2}(x_i - \lambda\rho_i)^2\right]$$

$$\frac{x_i^2}{2} - \frac{1}{2}(x_i^2 + \lambda^2\rho_i^2 - 2\lambda x_i\rho_i)$$

$$= -\frac{1}{2}\lambda^2\rho_i^2 + \lambda x_i\rho_i = \lambda\left(x_i\rho_i - \frac{1}{2}\lambda\rho_i^2\right)$$

$$= \lambda \max_{|\rho_i|\le1}\left(x_i\rho_i - \frac{1}{2}\lambda\rho_i^2\right)$$

\# $\max = -\min$ # $-\frac{1}{2}\lambda\left(\rho_i^2 - 2\frac{x_i}{\lambda}\rho_i + \left(\frac{x_i}{\lambda}\right)^2 - \frac{x_i^2}{\lambda^2}\right)$

$$= -\frac{1}{2}\lambda\left(\rho_i - \frac{x_i}{\lambda}\right)^2 + \frac{1}{2}\cdot\lambda\cdot\frac{x_i^2}{\lambda^2}$$

$$= -\lambda \min_{|\rho_i|\le1}\left(\frac{1}{2}\lambda\left(\rho_i - \frac{x_i}{\lambda}\right)^2 - \frac{1}{2}\frac{x_i^2}{\lambda}\right)$$

$$= -\lambda\left[-\frac{1}{2}\frac{x_i^2}{\lambda} + \frac{\lambda}{2}\underbrace{\min_{|\rho_i|\le1}\left(\rho_i - \frac{x_i}{\lambda}\right)^2}\right]$$

$\frac{|x_i|}{\lambda} \in \{\square : |\square| \le 1\}$

[illegible], [$\frac{|x_i|}{\lambda} \le 1$] $\rho_i^*(\lambda) = \frac{x_i}{\lambda}$ with 0 minvalue

[$\frac{x_i}{\lambda} > 1$] $\rho_i^*(\lambda) = 1$ — $+1$, -1, x_i/λ } min distance, } ρ_i

[$-\frac{x_i}{\lambda} > 1$] $\rho_i^*(\lambda) = -1$ — $\{x_i/\lambda\} < -1$ — $+1$, -1 } ρ_i, x_i/λ

in this cases we can impose the following formula:

$\rho_i^*(\lambda) = \operatorname{sgn}(x_i)$ {as when $x_i > \lambda > 0$, $\rho_i^*(\lambda) = \operatorname{sgn}(x_i) = 1$ (by defn of λ)

〃 $x_i < -\lambda < 0$, $\rho_i^*(\lambda) = \operatorname{sgn}(x_i) = -1$ }

combining all of these

$$\min_{|\rho_i|\le1}\left[\left(\rho_i - \frac{x_i}{\lambda}\right)^2\right] = \rho_i^*(\lambda) = \begin{cases} x_i/\lambda, & \text{if } |x_i| \le \lambda \\ \operatorname{sgn}(x_i), & \text{else} \end{cases}$$

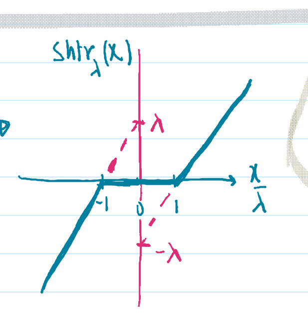
Then, from [eq: z^*(\lambda)], $\rho_i^*(\lambda)$의 value 대입하면, $z_i^*(\rho_i^*,\lambda) = x_i - \rho_i^*(\lambda)\lambda = x_i - \lambda\begin{cases} x_i/\lambda, & \text{if } |x_i| \le \lambda \\ \operatorname{sgn}(x_i), & \text{else} \end{cases}$

$$= \begin{cases} 0, & \text{if } |x_i| \le \lambda \\ x_i - \lambda\operatorname{sgn}(x_i), & \text{else} \end{cases}$$

So, now we have $z_i^*(\lambda) = \begin{cases} 0, & \text{if } |x_i| \le \lambda, \\ x_i - \lambda\operatorname{sgn}(x_i), & \text{else} \end{cases} = \operatorname{shr}_\lambda(x_i) = \operatorname{sgn}(x_i)[|x_i| - \lambda]_+$

$z^*(\lambda) = (z_i^*(\lambda))_{i=1}^n$ is min of the Lagrangian, giving us one piece of the KKT puzzle

this is called soft threshold function / shrinkage operator



now by Sion's minimax theorem $z_i^*(\rho_i^*, \lambda) = z_i^*(\lambda)$ will also be the minimizer of the Lagrangian as in [eq: z_i^*(\lambda) Orginal]

thus giving us information regarding one equation (vanishing (sub) gradient of the Lagrangian) of the KKT condition

\# By Sion's mini-max theorem [eq: Sion's mini-max theorem] $\min_{z_i}\max_{|\rho_i|\le1}\left(\frac{1}{2}(z_i-x_i)^2 + \lambda\rho_i z_i\right) = \max_{|\rho_i|\le1}\min_{z_i}\left(\frac{1}{2}(z_i-x_i)^2 + \lambda\rho_i z_i\right)$

$\rho_i^* = \begin{cases} x_i/\lambda, & \text{if } |x_i| \le \lambda \\ \operatorname{sgn}(x_i), & \text{else} \end{cases}$

$z^*(\lambda)=(z_i^*(\lambda))_{i=1}^n$ is $\inf$ of the Lagrangian, giving us one piece of the KKT puzzle.

this is called soft threshold function / shrinkage operator

(eq: Optimization Problem for l-1 norm projection) যেহেতু বোঝা যাচ্ছে, strong duality holds, as $z=0$ is strictly feasible point + the optimal value is finite because the set is bounded.

$[x]_{\mathcal{X}}$ দ্বারা কোন কোন আমাদের optimal λ পাওয়া যাবে from the KKT condition.

Rest of the KKT condition are:

primal feasibility: $\|z^*(\lambda)\|_1 \le 1$

dual feasibility: $\lambda \ge 0$

complementary slackness: $\lambda(\|z^*(\lambda)\|-1)=0$

now. if $\lambda=0 \Rightarrow$ ($\|z^*(\lambda)\|\le 1$ is an inactive constraint $\wedge$ $z^*(\lambda^*)=[z_i^*(\lambda^*)]_{i=1}^n=[x_i]_{i=1}^n=x$)

$[x]_{\mathcal{X}}=z_i^*(\lambda)=x \Rightarrow \underbrace{\|x\|_1\le 1}_{x\in\mathcal{X}}$ [যেমন then $[x]_{\mathcal{X}}=x$.

$\therefore \lambda=0 \rightarrow [x]_{\mathcal{X}}=x$

অথবা:

$[x]_{\mathcal{X}}=x=z^*(\lambda)=[z_i^*(\lambda)]=\left[\begin{cases}0, & \text{if } \|x_i\|/\lambda\le 1\\ x_i-\lambda\,\mathrm{sgn}(x_i)\end{cases}\right]_{i=1}^n \Rightarrow \lambda=0$

$\lambda=0 \leftrightarrow \|x\|_1\le 1$

• $\lambda=0 \Rightarrow [x]_{\mathcal{X}}=x-\mathrm{sgn}(x)$ {other case does not hold $\frac{|x_i|}{0}=\infty \overset{x}{\lessgtr} 0$} $=x=z^*(\lambda)\Rightarrow \|z^*(\lambda)\|_1=\|x\|_1\le 1$ by feasibility

• $\|x\|_1\le 1 \Rightarrow \left(\inf_{\|z\|_1\le 1}\frac{1}{2}\|z-x\|_2^2\right)=0$ as $x=z$ will yield 0 norm with feasibility constraint $\|z\|_1=\|x\|_1\le 1$ satisfied $\Rightarrow z^*(\lambda)=x=\left[\begin{cases}0, & \|x_i\|/\lambda\le 1\\ x_i-\lambda\,\mathrm{sgn}(x_i), & \text{else}\end{cases}\right]_{i=1}^n \Rightarrow \lambda=0$

// remember this formula // was derived for any λ

$\lambda=0 \leftrightarrow \|x\|_1\le 1$

$\therefore \lambda=0 \Leftrightarrow [x]_{\mathcal{X}}=x \Leftrightarrow \|x\|_1\le 1$

• Now consider, then $\underset{\|x\|_1>1}{\lambda>0} \rightarrow \|z^*(\lambda)\|_1-1=0 \Leftrightarrow \left(\sum_{i=1}^n \max(|x_i|-\lambda,0)=1\right)\xrightarrow[\lambda]{\text{solve}}\lambda^*$, then, $[x]_{\mathcal{X}}=\mathrm{sthr}_{\lambda^*}(x)$ {elementwise}

$z_i^*(\lambda)=\left\|\begin{cases}0, & \frac{|x_i|}{\lambda}\le 1\\ x_i-\lambda\,\mathrm{sgn}(x_i), & \text{else}\end{cases}\right\|_1=\begin{cases}|0|=0, & \frac{|x_i|}{\lambda}\le 1\\ |x_i-\lambda\,\mathrm{sgn}(x_i)|, & \frac{|x_i|}{\lambda}>1\end{cases}=\begin{cases}0, & \text{if } |x_i|-\lambda\le 0\\ |x_i|-\lambda, & \text{if } |x_i|-\lambda>0\end{cases}=\max\{|x_i|-\lambda,0\}$

{ $[x_i>0]$ $|x_i-\lambda\,\underbrace{\mathrm{sgn}(x_i)}_{1}|=|x_i-\lambda|=x_i-\lambda$ { $\because |x_i|=x_i>\lambda \rightarrow (x_i-\lambda)>0 \rightarrow |x_i-\lambda|=(x_i-\lambda)$ }

$[x_i=-\bar{x}_i<0]$ $|-\bar{x}_i-\lambda\,\underbrace{\mathrm{sgn}(-\bar{x}_i)}_{-1}|=|-\bar{x}_i+\lambda|=|-(\bar{x}_i-\lambda)|$ { $|x_i|=\bar{x}_i>\lambda\rightarrow(\bar{x}_i-\lambda)>0$ } $=\bar{x}_i-\lambda=-x_i-\lambda$

combining both $|x_i-\lambda\,\mathrm{sgn}(x_i)|=|x_i|-\lambda$ }

$$[x]_{\mathcal{X}}=\begin{cases}x, & \text{if } \|x\|_1\le 1\\ \mathrm{sthr}_{\lambda^*}(x), & \text{if } \|x\|_1>1\end{cases}$$

λ^* = {solve$_\lambda$ $\left(\sum_{i=1}^n\max(|x_i|-\lambda,0)=1\right)$}

Ans:

Lets try to understand how $\mathrm{sthr}_\lambda(x_i)$ works.

Need to finish

$-\lambda$

now:

$$\inf_{z_i}\left[\frac{1}{2}(z_i-x_i)^2+\lambda|z_i|\right]=\inf_{z_i}\left[\sup_{|\rho_i|\le 1}\left(\frac{1}{2}(z_i-x_i)^2+\lambda\rho_i z_i\right)\right]=\sup_{|\rho_i|\le 1}\left[\inf_{z}\left(\frac{1}{2}(z_i-x_i)^2+\lambda\rho_i z_i\right)\right]=\sup_{|\rho_i|\le 1}\left[\frac{1}{2}(z_i-x_i)^2+\lambda\rho_i z_i\right]_{\substack{z_i=z_i^*(\rho_i,\lambda)\\=x_i-\lambda\rho_i}}=\sup_{|\rho_i|\le 1}\left[\frac{1}{2}(z_i^*(\rho_i,\lambda)-x_i)^2+\lambda\rho_i z_i^*(\rho_i,\lambda)\right]=\frac{1}{2}(z_i^*(\rho_i^*,\lambda)-x_i)^2+\lambda\underbrace{\rho_i^* z_i^*(\rho_i^*,\lambda)}_{|z_i^*(\rho_i^*,\lambda)|}=\frac{1}{2}(z^*(\rho_i^*,\lambda)-x_i)^2+\lambda|z_i^*(\rho_i^*,\lambda)|$$

{ [eq: z^*(\lambda)] থেকে }

$\rho_i^*=\{\ \ldots;\ \mathrm{sgn}(x_i),\ \text{else}$

{ $|z_i^*(\rho_i^*,\lambda)|=\rho_i^* z_i^*(\rho_i^*,\lambda)=\sup_{|\rho_i|\le 1}\rho_i z_i^*(\rho_i,\lambda)$ }

A better explanation needed

$$z_i^*(\lambda)=\inf_{z_i}\left[\frac{1}{2}(z_i-x_i)^2+\lambda|z_i|\right]=(z_i^*(\rho_i^*,\lambda)=\begin{cases}0, & \text{if } |x_i|\le\lambda\\ x_i-\lambda\,\mathrm{sgn}(x_i), & \text{else}\end{cases}\ \}$$

Projection onto the positive semidefinite cone.

$\mathcal{X}=\{X\in S^n: X\succeq 0\}=S_+^n$

Given: X { $\in S^n$ }

?: $[X]_{\mathcal{X}}=\underset{Z\in\mathcal{X}}{\operatorname{argmin}}\|Z-X\|_F^2$ { $\|X\|_F^2=\sum_{i=1}^n \lambda_i^2$ }

// A symmetric matrix can be diagonalized by orthogonal similarity transformation

$X=U\Lambda U^T$ { U {orthogonal matrix}, $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ }

$\|Z-X\|_F^2=\|Z-U\Lambda U^T\|_F^2=\|\underbrace{UU^T}_{I}Z\underbrace{UU^T}_{I}-U\Lambda U^T\|_F^2=\|\overset{\text{orthogonal}}{U}(\underbrace{U^TZU}_{\tilde{Z}}-\Lambda)\overset{\text{orthogonal}}{U^T}\|_F^2$ { norm does not change when you multiply something by orthogonal matrix }

$=\|U^TZU-\Lambda\|_F^2=\|\tilde{Z}-\Lambda\|_F^2$

one-to-one   one-to-one

{ the operator: $U^T\square U: S^n\to S^n$ is a one to one operator ($X\ne Y\rightarrow \underset{\tilde{X}}{U^T\square U}\ne\underset{\tilde{Y}}{U^T\square U}$)

so $[X]_{\mathcal{X}}=\underset{Z\in\mathcal{X}}{\operatorname{argmin}}\|Z-X\|_F^2=\underset{Z\in\mathcal{X}}{\operatorname{argmin}}\|U^TZU-\Lambda\|_F^2$

$=U\left[\underset{\tilde{Z}\in\mathcal{X}}{\operatorname{argmin}}\|\tilde{Z}-\Lambda\|\right]U^T$ # say, Z^* solves $\inf_{Z\in\mathcal{X}}\|U^TZU-\Lambda\|_F^2$

$\tilde{Z}^*$ " $\inf_{\tilde{Z}\in\mathcal{X}}\|\tilde{Z}-\Lambda\|_F^2$

then $\tilde{Z}^*=U^TZ^*U$ { one to one $U\square U^T$ operator বলে }

$\left[\underset{\tilde{Z}\in\mathcal{X}}{\operatorname{argmin}}\|\tilde{Z}-\Lambda\|_F^2\right]=U^T\left[\underset{Z\in\mathcal{X}}{\operatorname{argmin}}\|UZU^T-\Lambda\|_F^2\right]U$

$\rightarrow U\left[\underset{\tilde{Z}\in\mathcal{X}}{\operatorname{argmin}}\|\tilde{Z}-\Lambda\|_F^2\right]U^T=\underset{Z\in\mathcal{X}}{\operatorname{argmin}}\|UZU^T-\Lambda\|_F^2$ [$\because U^TU=UU^T=I$] }

$\therefore \underset{\tilde{Z}\succeq 0}{\operatorname{argmin}}\|\tilde{Z}-\Lambda\|_F^2=[\Lambda]_+=\mathrm{diag}([\lambda_1]_+,\ldots,[\lambda_n]_+)$

// analogous with $\underset{Z\ge 0}{\operatorname{argmin}}\|Z-X\|_2^2=[X]_+$ [eq: Projection on positive orthant]

$\boxed{[X]_{\mathcal{X}}=U[\Lambda]_+U^T}$ Ans:

⊛ **Proximal map of ℓ_1 regularization:**

Lasso problem: (ℓ_1 regularized least square)

$$\inf_x \underbrace{\frac{1}{\gamma}\|Ax-b\|_2^2}_{f_0(x)}+\underbrace{\|x\|_1}_{h(x)}$$

$f_0(x)$: { $\mathcal{D}$ strongly } $\Leftrightarrow$ A {full rank}

$h(x)$: { $\mathcal{D}$, nondifferentiable }

$\nabla_{x_k}\frac{1}{\gamma}\|Ax_k-b\|_2^2=2\frac{1}{\gamma}A^T(Ax_k-b)$

Proximal algorithm:

$$\forall_k\quad x_{k+1}=\mathrm{prox}_{s_kh}\left(x_k-s_k\nabla f_0(x_k)\right)$$

$$=\mathrm{prox}_{s_kh}\left(x_k-s_k\cdot\frac{2}{\gamma}A^T(Ax_k-b)\right)$$

$$=\mathrm{prox}_{s_kh}\left(Ix_k-\frac{2s_k}{\gamma}A^TAx_k+\frac{2s_k}{\gamma}A^Tb\right)$$

$$=\mathrm{prox}_{s_kh}\left((I-\frac{2s_k}{\gamma}A^TA)x_k+\frac{2s_k}{\gamma}A^Tb\right)=\mathrm{sthr}_{s_k}\left((I-\frac{2s_k}{\gamma}A^TA)x_k+\frac{2s_k}{\gamma}A^Tb\right)\Rightarrow\therefore\forall_k\ x_{k+1}=\mathrm{sthr}_{s_k}\left((I-\frac{2s_k}{\gamma}A^TA)x_k+\frac{2s_k}{\gamma}A^Tb\right)$$

WOW! a closed form iteration :o

in terms of fixed point theory:

$x^*=\mathrm{sthr}_{s_k}\left((I-\frac{2s_k}{\gamma}A^TA)x^*+\frac{2s_k}{\gamma}A^Tb\right)$ এটা solve করে আমরা optimal solution পেতে পারি. ⓘ

{ $\mathrm{prox}_h(x)=\underset{z}{\operatorname{argmin}}\left(h(z)+\frac{1}{2}\|z-x\|_2^2\right)$ }

$\mathrm{prox}_{s_kh}(x)=\underset{z}{\operatorname{argmin}}\left(s_kh(z)+\frac{1}{2}\|z-x\|_2^2\right)$

$=\underset{z}{\operatorname{argmin}}\left(s_k\|z\|_1+\frac{1}{2}\|z-x\|_2^2\right)=\mathrm{sthr}_{s_k}(x)$

{ $\underset{z}{\operatorname{argmin}}\left(s_k\|z\|_1+\frac{1}{2}\|z-x\|_2^2\right)=[\mathrm{sthr}_{s_k}(x_i)]_{i=1}^n=\mathrm{sthr}_{s_k}(x)$

as (eq: Optimization Problem for l-1 norm projection) থেকে পাই:

$\underset{z}{\operatorname{argmin}}\left(\frac{1}{2}\|z-x\|_2^2+\lambda\|z\|_1\right)=\mathrm{sthr}_\lambda(x)=[\mathrm{sthr}_\lambda(x_i)]_{i=1}^n$ }

• ISTA (Iterative shrinkage-thresholding algorithm) for Lasso:

For lasso:

gradient Lipschitz constraint

For lasso:

$f_0(x) = \frac{1}{\gamma}\|Ax-b\|_2^2$

$\nabla f_0(x) = \frac{1}{\gamma}\, 2A^T(Ax-b) = \frac{1}{\gamma}(2A^TAx - 2A^Tb)$

$\nabla^2 f_0(x) = \frac{2}{\gamma} A^TA$ // Matrix cookbook 96-98: $f = x^TAx + b^Tx \rightarrow \begin{matrix}\nabla_x f = Ax \\ \nabla_x^2 f = A\end{matrix}$, now আমরা $f_0(x)$ এর জন্য L, m বের করেছি to come up with a stopping criterion for proximal gradient algorithm.

gradient Lipschitz constraint

strong convexity constraint

{ [eq: 12.60] থেকে পাই: $\|g_k\|_2^2 \le 2\varepsilon \frac{mL}{L-m} \Rightarrow f(x_{k+1}) - f(x^*) \le \varepsilon$ }

* Strong convexity constraint for f_0:

* $f_0(x)\{\square_{\text{strongly}}\} \leftrightarrow \forall_{x\in \text{dom} f_0}\ \nabla^2 f_0(x) \succeq mI$

$\leftrightarrow \forall_{x\in\text{dom} f_0}\ \nabla^2 f_0(x) - mI \succeq 0$

{ now the matrix A^TA symmetric positive semidefinite, so orthogonal similarity transformation is possible with all eigenvalues nonnegative:

$A^TA = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T \quad \{\lambda_1 \ge \dots \ge \lambda_n \ge 0\}$

$\therefore \nabla^2 f_0(x) - mI = \frac{2}{\gamma} Q\Lambda Q^T - m\underbrace{QQ^T}_{I} = Q\left(\frac{2}{\gamma}\Lambda - mI\right)Q^T$

eigenvalues of $\nabla^2 f_0(x) - mI$; $\nabla^2 f_0(x) - mI$ symmetric and we have just found an orthogonal similarity transformation of that, so the diagonal matrix will correspond to the eigenvalues

$\rightarrow$ all eigenvalues $\ge 0 \leftrightarrow \forall_i\ \frac{2}{\gamma}\lambda_i - m \ge 0$

$\leftrightarrow \frac{2}{\gamma}\min(\lambda_i) - m \ge 0$

$\leftrightarrow \frac{2}{\gamma}\lambda_{min}(A^TA) \ge m$

$\leftrightarrow \boxed{m_{max} = \frac{2}{\gamma}\lambda_{min}(A^TA)}$ [eq: Strong Convexity Constant Lasso]

this can be set as the strong convexity constraint of $f_0(x) = \frac{1}{\gamma}\|Ax-b\|_2^2$

*Finding a global Lipschitz constraint:

{ from Lemma 12.1.1: $f\{f: \mathbb{R}^n \to \mathbb{R}$, gradient_Lipschitz_continuous, $\delta_{\square,2}\} \leftrightarrow \forall_{\square}\ \|\nabla^2 f(\square)\|_F = \left(\sum \sigma_i^2\right)^{1/2} \le L$ }

$f_0(x) = \frac{1}{\gamma}\|Ax-b\|_2^2\ \{f: \mathbb{R}^n \to \mathbb{R}$, n ✓, n ✓$\} \leftrightarrow \forall_x\ \|\nabla^2 f(x)\|_F = \|\frac{2}{\gamma}A^TA\|_F = \frac{2}{\gamma}\|A^TA\|_F = \frac{2}{\gamma}\left(\sum_{i=1}^{n}\sigma_i^2(A^TA)\right)^{1/2} \le L$

λ_i^2 (for a symmetric PSD matrix $\lambda_i = \sigma_i$)

{ $\because \|\alpha x\| = |\alpha|\,\|x\|$ }

$\therefore L_{min} = \frac{2}{\gamma}\left(\sum_{i=1}^{n}\lambda_i^2(A^TA)\right)^{1/2}$ [eq: Gradient Lipshitz Constant Lasso]

this we set as the Gradient Lipschitz constant.

[eq: Strong Convexity Constant Lasso] [eq: Gradient Lipshitz Constant Lasso]

Now, we know both m and L, so let's give the proximal gradient algorithm for Lasso (constant stepsize):

Require: $\varepsilon > 0, x_0$, A full rank:

1. Compute $m = \frac{2}{\gamma}\lambda_{min}(A^TA)$, $L = \frac{2}{\gamma}\left(\sum_{i=1}^{n}\sigma_i^2(A^TA)\right)^{1/2}$
2. $k := 0$, $s = 1/L$
3. $\nabla f_0(x_k) = \frac{2}{\gamma}(A^T(Ax_k - b))$
4. $x_{k+1} = \text{sthr}_s(x_k - s\nabla f_0(x_k))$
5. $\|g_k\|_2 = \|x_k - x_{k+1}\|_2 / s$ { $x_{k+1} = x_k - s_k g_k \rightarrow \|g_k\|_2 = \frac{\|x_k - x_{k+1}\|}{s}$ }
6. if ($\|g_k\|_2^2 \le 2\varepsilon \frac{mL}{L-m}$)
   done!, return $x^* = x_{k+1}$
   else
   $k := k+1$,
   go to 3

*Fast Proximal gradient (constant stepsizes)

Normal proximal gradient এর convergence rate $(\frac{1}{k})$, suitable modification এর মাধ্যমে $(\frac{1}{k^2})$ convergence rate achieve করা যায়, এই type এর algorithm কে Fast Proximal Gradient algorithm বলে, it has two versions

1. when L is known, $s_k = s = 1/L$
2. L is not known. then backtracking type এর line search ব্যবহার করে S_k বের করা হয়

[Probably need to elaborate later]