Assignment 4 A0236307J WANG JIANGYI

Qusetion 1 : $H = \{ x \in \mathbb{R}^n : b^T x \le c \}$ is a convex set

Proof: $\forall x_1, x_2 \in H$, we have:

$$\begin{cases} b^T x_1 \le c \\ b^T x_2 \le c \end{cases}$$

this implies, for $\forall \lambda \in [0,1]$, we have,

$$\begin{aligned} & b^T (\lambda x_1 + (1-\lambda) x_2) \\ = \; & \lambda b^T x_1 + (1-\lambda) b^T x_2 \\ \le \; & \lambda \cdot c + (1-\lambda) c \\ = \; & c \end{aligned}$$

Therefore, $\underline{\lambda x_1 + (1-\lambda) x_2 \in H}$

That is, we show that for $\forall x_1, x_2 \in H$ and $\lambda \in [0,1]$,

it holds that $\underline{\lambda x_1 + (1-\lambda) x_2 \in H}$

Therefore, H is a convex set.

Question 2 : C_1 and C_2 are convex sets in $\mathbb{R}^n$

(a) $C_1 \cap C_2 = C$ is also a convex set

Proof: $\forall x_1, x_2 \in C$, we have:

$$\begin{cases} x_1 \in C_1 \text{ and } x_1 \in C_2 \\ x_2 \in C_1 \text{ and } x_2 \in C_2 \end{cases}$$

Since C_1 is a convex set, then for $\forall \lambda \in [0,1]$, we have:

$$\underline{\lambda x_1 + (1-\lambda) x_2 \in C_1} \quad ①$$

Similarly, since C_2 is also a convex set, we have

$$\underline{\lambda x_1 + (1-\lambda) x_2 \in C_2} \quad ②$$

Based on ① and ②, we have:

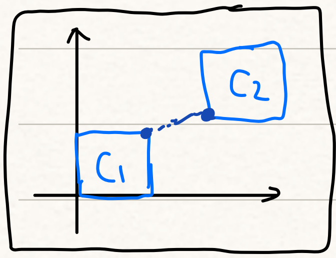
$$\lambda x_1 + (1-\lambda) x_2 \in C_1 \cap C_2 = C$$

this holds for $\forall x_1, x_2 \in C$ and $\forall \lambda \in [0,1]$.

Therefore, $C = C_1 \cap C_2$ is a convex set.

(b) Counterexample for C_1, C_2 convex but $C_1 \cup C_2$ non-convex.

Example: consider $C_1, C_2 \subseteq \mathbb{R}^2$ case $(n=2)$



$$\begin{cases} C_1 = [0,1] \times [0,1] \\ C_2 = [2,3] \times [2,3] \end{cases}$$

Firstly, we briefly show the convexity of C_1 & C_2

As for C_1, it can be expressed as:

$$C_1 = \bigcap_{i=1}^{4} H_i \quad \text{where } H_i = \{ x \in \mathbb{R}^2 : b_i^T x \le c_i \}$$

$\boxed{\text{question 1}}$ $\boxed{\text{question 2 (a)}}$

Since H_i is convex & convexity is preserved by intersection

it implies that C_1 is a convex set

For the same reasoning, we also have C_2 is a convex set

Secondly, we want to show $C_1 \cup C_2$ is non-convex

Here, $x_1 = (1,1) \in C_1 \cup C_2$ and $x_2 = (2,2) \in C_1 \cup C_2$

However, consider $x_\lambda = \lambda x_1 + (1-\lambda) x_2 \quad (\lambda \in (0,1))$

$$= (2-\lambda, 2-\lambda)$$

since $2-\lambda \in (1,2)$

Therefore, we have $x_\lambda \notin C_1 \cup C_2$, which implies that $C_1 \cup C_2$ is not a convex set.

Question 3 : $C = \{ x \in \mathbb{R}^n : \|x\|_2 \leq r \}$

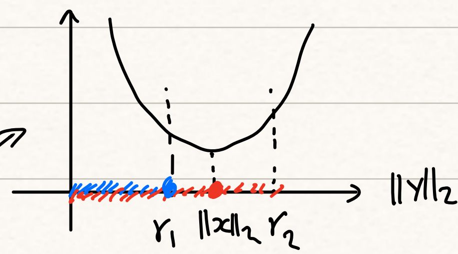
(a) determine $\Pi_C(x)$

Calculate : $$\Pi_C(x) := \underset{y \in C}{\operatorname{argmin}} \; \|y - x\|_2^2$$

$$= \underset{y}{\operatorname{argmin}} \; \|y\|_2^2 - 2x^T y \quad \text{s.t.} \quad \|y\|_2 \leq r$$

(From Cauchy-Schwaz Inequality) $$= \underset{y}{\operatorname{argmin}} \; \|y\|_2^2 - 2\|x\|_2 \|y\|_2 \quad \text{s.t} \quad \|y\|_2 \leq r$$

$y = kx$ for some $k \in \mathbb{R}$



(the minizer of quadratic function) $$= \begin{cases} r \cdot \dfrac{x}{\|x\|_2} & , \|x\|_2 > r \\ x & , \|x\|_2 \leq r \end{cases}$$

(b) Calculate $\Pi_C(x)$ for $r = 2$ and $x = (3, 1)^T$

Calculate : $\|x\|_2 = \sqrt{3^2 + 1^2} = \sqrt{10} > 2$

Therefore $$\Pi_C(x) = \frac{2}{\sqrt{10}} x$$

$$= \frac{\sqrt{10}}{5} (3, 1)^T$$

Question 4: $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \quad x = (x_1, x_2)^T \in \mathbb{R}^2$

(a) Gradient & Hessian

$$\frac{\partial}{\partial u}(x_2-u)^2 = 2(x_2-u)\cdot(-1)$$

Calculate: $\frac{\partial f}{\partial x_1} = 100 \times 2(x_2 - x_1^2)\cdot(-1)\cdot 2x_1 + 2(1-x_1)\cdot(-1)$

$$= 200(x_1^2 - x_2)\cdot 2x_1 + 2(x_1 - 1)$$

$$\frac{\partial f}{\partial x_2} = 100 \times 2(x_2 - x_1^2)$$

$$= 200(x_2 - x_1^2)$$

$$\Rightarrow \nabla f(x) = \begin{pmatrix} 200(x_1^2 - x_2)\cdot 2x_1 + 2(x_1 - 1) \\ 200(x_2 - x_1^2) \end{pmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = 200\left(2x_1 \cdot 2x_1 + 2(x_1^2 - x_2)\right) + 2$$

$$= 200(6x_1^2 - 2x_2) + 2$$

$$\frac{\partial^2 f}{\partial x_2^2} = 200$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -400x_1$$

$$\Rightarrow \nabla^2 f(x) = \begin{pmatrix} 200(6x_1^2 - 2x_2) + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

(b) Prove $x^* = (1,1)^T$ is local minimizer.

Proof: ① $\nabla f(x^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

② $\nabla^2 f(x^*) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} := H$

consider the eigenvalue of matrix H:

λ is one eigenvalue of $H \Leftrightarrow |H - \lambda I| = 0$

$$|H-\lambda I| = \begin{vmatrix} 802-\lambda & -400 \\ -400 & 200-\lambda \end{vmatrix}$$

$$= (802-\lambda)(200-\lambda) - 160000$$

$$= \lambda^2 - 1002\lambda + 400$$

Therefore, $|H-\lambda I| = 0 \Rightarrow \begin{cases} \lambda_1 = \dfrac{1002+1001.2}{2} > 0 \\ \lambda_2 = \dfrac{1002-1001.2}{2} > 0 \end{cases}$

This implies that, $H \succ 0$ (Positive Definite)

(Sufficent Optimality Condition)

Therefore, since $\begin{cases} \nabla f(x^*) = \underline{0} \\ \nabla^2 f(x^*) \succ 0 \end{cases} \Rightarrow$ x^* is one local minimizer

(c) Prove $x^* = (1,1)^T$ is the only local minimizer.

Proof: From the Necessary Optimality Condition,

if x is the local minimizer of $f(\cdot)$,

then we will have $\nabla f(x) = 0$

That is, $\begin{cases} 200(x_1^2 - x_2) \cdot 2x_1 + 2(x_1 - 1) = 0 \\ 200(x_2 - x_1^2) = 0 \end{cases}$

$$\Rightarrow \begin{cases} x_2 = 1 \\ x_1 = 1 \end{cases}$$

This shows that, $x^* = (1,1)^T$ is the only possible local minimizer

Moreover, from Q4 (b), we have already proved that,

$x^* = (1,1)^T$ is the local minimizer

Therefore, $x^* = (1,1)^T$ is the only local minimizer!

#

Question 5 : $f(x) = (x_1 + x_2^2)^2 \qquad x = (x_1, x_2)^T \in \mathbb{R}^2$

(a) $x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, compute Steepest Descent Direction

Calculation : $\nabla f(x) = \begin{pmatrix} 2(x_1 + x_2^2) \\ 2(x_2^2 + x_1) \cdot 2x_2 \end{pmatrix}$

Therefore, $\nabla f(x^{(0)}) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

Therefore, the Steepest Descent Direction $d_{SD}^{(0)}$ for $x^{(0)}$ is :

$$d_{SD}^{(0)} = -\nabla f(x^{(0)}) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

(b) $x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, compute Another Descent Direction

Calculation : $d^{(0)}$ is a descent direction at $x^{(0)}$

$$\Leftrightarrow \quad \nabla f(x^{(0)})^T d^{(0)} < 0$$

Here, $\nabla f(x^{(0)}) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

Therefore, $d^{(0)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ is one valid descent direction.

(c) $x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, exact line search w.r.t Steepest Descent

Calculation : $d_{SD}^{(0)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$

Therefore, $x^{(1)} = x^{(0)} + \alpha_0 \cdot d_{SD}^{(0)}$

$$= \begin{pmatrix} 1 - 2\alpha_0 \\ 0 \end{pmatrix}$$

Through Exact Line Search Framework, α_0 is determined by :

$$\alpha_0 = \underset{\alpha}{\operatorname{argmin}}\ f(x^{(0)} + \alpha_0 d_{SD}^{(0)})$$

$$= \underset{\alpha}{\operatorname{argmin}}\ (1 - 2\alpha_0)^2$$

$$= \frac{1}{2}$$

Therefore, $x^{(1)} = \begin{pmatrix} 1-1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

(d) $x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, Backtracking Line Search w.r.t Steepest Descent

Calculation : $d_{SD}^{(0)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$ $\quad \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

① $\bar{\alpha} = 1 \Rightarrow \begin{cases} f(x^{(0)} + \bar{\alpha} d_{SD}^{(0)}) = 1 \\ f(x^{(0)}) + c_1 \bar{\alpha} \nabla f(x^{(0)})^T d_{SD}^{(0)} = 1 - 10^{-4} \cdot 4 \end{cases}$

since $1 > 1 - 4 \cdot 10^{-4}$, we shrink $\bar{\alpha}$ by ρ

② $\bar{\alpha} = 0.9 \Rightarrow \begin{cases} f(x^{(0)} + \bar{\alpha} d_{SD}^{(0)}) = 0.64 \\ f(x^{(0)}) + c_1 \bar{\alpha} \nabla f(x^{(0)})^T d_{SD}^{(0)} = 1 - 10^{-4} \cdot 3.6 \end{cases}$

since $0.64 < 1 - 3.6 \times 10^{-4}$

we set $x^{(1)} = x^{(0)} + \bar{\alpha} d_{SD}^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0.9 \begin{pmatrix} -2 \\ 0 \end{pmatrix}$

$$= \begin{pmatrix} -0.8 \\ 0 \end{pmatrix}$$

(e) $x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, calculate Newton Direction at $x^{(0)}$

Calculation: $\nabla^2 f(x) = \begin{pmatrix} 2 & 4x_2 \\ 4x_2 & 12x_2^2 + 4x_1 \end{pmatrix}$

Therefore, $\nabla^2 f(x^{(0)}) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$

Therefore. New Direction $d_{NT}^{(0)}$ at $x^{(0)}$ is.

$$d_{NT}^{(0)} = -\nabla^2 f(x^{(0)})^{-1} \nabla f(x^{(0)})$$

$$= -\begin{pmatrix} \frac{1}{2} & \\ & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

(f) $x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, pure Newton's method ($\alpha = 1$)

Calculation : $x^{(1)} = x^{(0)} + 1 \cdot d_{NT}^{(0)}$

$$= \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Question 6 : [effective domain & conjugate function]

(a) $f(x) = \frac{1}{2}x^2 + 4x$, $x \in \mathbb{R}$

effective domain & conjugate function f^*

Calculation : By definition, $\operatorname{dom} f := \{ x \in \mathbb{R} : f(x) < +\infty \}$

Therefore, ① $\operatorname{dom} f = (-\infty, +\infty)$

② $f^*(x) = \sup_y \{ \langle y, x \rangle - f(y) : y \in \mathbb{R} \}$

$$= \sup_{y \in \mathbb{R}} \{ yx - \frac{1}{2}y^2 - 4y \}$$

$$= \frac{1}{2}(4 - x)^2$$

(b) $f(x) = -\sum_{i=1}^{n} \log(x_i)$, $x \in \mathbb{R}^n$

effective domain & conjugate function f^*

Calculation : ① $\text{dom } f = (0, +\infty]^n$

② $f^*(x) = \sup_y \{ \langle y, x \rangle - f(y) : y \in (0, +\infty]^n \}$

$$= \sup_y \{ x^T y + \sum_{i=1}^{n} \log(y_i) : y \in (0, +\infty]^n \}$$

$$= \sum_{i=1}^{n} \sup_{y_i} \{ x_i y_i + \log(y_i) : y_i > 0 \}$$

$$= \begin{cases} -n - \sum_{i=1}^{n} \log(-x_i) , & \boxed{x_i < 0 \quad \forall i \in [n]} \\ +\infty , & o/w \end{cases}$$

$$\boxed{x \in [-\infty, 0)^n}$$

Question 7 : $f(x) = \lambda \|x\|_2$ $x \in \mathbb{R}^n$

(a) conjugate function $f^*(\cdot)$

Calculation : $f^*(x) = \sup_y \{ \langle x, y \rangle - \lambda \|y\|_2 : y \in \mathbb{R}^n \}$

$$= \sup_y \{ x^T y - \lambda \sqrt{y^T y} : y \in \mathbb{R}^n \} \quad (1)$$

consider optimization problem (1) :

$g(y) = x^T y - \lambda (y^T y)^{\frac{1}{2}}$ is a positive homogeneous function

i.e., $g(\alpha y) = \alpha g(y)$

Therefore, we must have $\sup_y g(y) = +\infty$ or 0

Moreover, when $\sup_y \{ g(y) : y^T y \le 1 \} > 0$

we must have $\sup_y \{g(y)\} = +\infty$ (easily)

when $\sup_y \{g(y) : y^T y \le 1\} \le 0$

we must have : $\sup_y \{g(y)\} = 0$ (easily)

Therefore, when $\sup_y \{x^T y - \lambda : \|y\|_2 \le 1\} \le 0$

$$\Leftrightarrow \sup_y \{x^T y : \|y\|_2 \le 1\} \le \lambda$$

☆

$$\Leftrightarrow \boxed{\|x\|_2 \le \lambda}$$

since $x^T y \le \|x\|_2 \|y\|_2$ and the equality is achieved at $y = kx$ for some $k \in \mathbb{R}$

we have $x^T y \le \|x\|_2$ and achieves equality at $y = \frac{x}{\|x\|_2}$, which implies $\sup_y \{x^T y : \|y\|_2 \le 1\} = \|x\|_2$

$$\Leftrightarrow \sup_y \{g(y)\} = 0$$

otherwise, $\sup_y \{g(y)\} = +\infty$

Therefore, $f^*(x) = \sup_y \{g_x(y) : y \in \mathbb{R}^n\}$

$$= \delta_{B^2_\lambda}(x)$$

where $B^2_\lambda := \{x \in \mathbb{R}^n : \|x\|_2 \le \lambda\}$

(b) calculate $P_{f^*}(\cdot)$

<u>Calculation</u> : $P_{f^*}(x) = P_{\delta_{B^2_\lambda}}(x)$

$= \Pi_{B^2_\lambda}(x)$

(result in Q3(a)) $= \begin{cases} x, & \|x\|_2 \le \lambda \\ \lambda \cdot \dfrac{x}{\|x\|_2}, & \|x\|_2 > \lambda \end{cases}$

(c) calculate $P_f(\cdot)$

<u>Calculation</u> : from <u>Moreau-Yosida Decomposition</u> thm,

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

$$= \begin{cases} 0, & \|x\|_2 \le \lambda \\ \left(1 - \dfrac{\lambda}{\|x\|_2}\right) x, & \|x\|_2 > \lambda \end{cases}$$

(d) $\lambda = 1$, $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, calculate $P_{f^*}(b)$, $P_f(b)$

<u>Calculation</u> : since $\|b\|_2 = \sqrt{2} > 1$.

then we have $P_{f^*}(b) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$

$$P_f(b) = b - P_{f^*}(b) = \begin{pmatrix} 1 - \frac{\sqrt{2}}{2} \\ 1 - \frac{\sqrt{2}}{2} \end{pmatrix}$$

#

Question 8 : $\begin{cases} \min\limits_{x} & \frac{1}{2} \|Ax-b\|_2^2 \\ \text{s.t.} & \|x\|_\infty \le r \end{cases}$

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

(a). re-formulate the optimization problem

Solution : $\min\limits_{x\in\mathbb{R}^n} \frac{1}{2} \|Ax-b\|_2^2 + \delta_{B_r^\infty}(x)$

Here, $B_r^\infty := \{ x : \|x\|_\infty \le r \}$

$$\begin{cases} f(x) = \frac{1}{2} \|Ax-b\|_2^2 \longrightarrow \underline{\text{smooth}} \\ g(x) = \delta_{B_r^\infty}(x) \longrightarrow \underline{\text{non-smooth}} \end{cases}$$

(b) PG iteration

Solution : In PG Framework, at step k we have $\underline{\alpha, \beta^{(k)}}$ then $\beta^{(k+1)}$ is constructed via :

$$\boxed{\begin{aligned} \nabla f(x) &= \nabla_x u \, \nabla_u f \\ &= A^T \cdot (Ax-b) \end{aligned}}$$

① $\bar{x}^{(k)} = x^{(k)} - \alpha \nabla f(x^{(k)})$
$= x^{(k)} - \alpha A^T (Ax^{(k)} - b)$

② $x^{(k+1)} = P_{\alpha g}(\bar{x}^{(k)})$

$= P_{B_r^\infty}(\bar{x}^{(k)})$

$= \prod_{B_r^\infty}(\bar{x}^{(k)})$

$$\Rightarrow x_i^{(k)} = \begin{cases} \bar{x}_i^{(k)}, & |\bar{x}_i^{(k)}| \le r \\ r, & \bar{x}_i^{(k)} > r \\ -r, & \bar{x}_i^{(k)} < -r \end{cases} \quad \text{for } \forall i \in [n]$$

(c) $r = 1,\ \alpha = 1$, $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ $b = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

$x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ calculate first iteration of PG

Solution: ① $\bar{x}^{(0)} = x^{(0)} - \alpha A^T (A x^{(0)} - b)$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

$$= \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

② $x^{(1)} = \Pi_{B_r^\infty} (\bar{x}^{(0)})$

$$= \Pi_{B_1^\infty} (\bar{x}^{(0)})$$

$$= \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Question 9: $\begin{cases} \min\limits_{x} & \|x\|_1 \\ \text{s.t.} & \|x - b\|_2 \leq r \end{cases}$

(a) Transform to 2-stage separable problem

Solution: $\begin{cases} \min\limits_{x} & \|x\|_1 \\ \text{s.t.} & \|x - b\|_2 \leq r \end{cases}$

$\Leftrightarrow \min\limits_{x} \|x\|_1 + \delta_{B_r^2}(x - b)$ Here, $B_r^2 := \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$

$\Leftrightarrow \begin{cases} \min\limits_{x, z} & \|x\|_1 + \delta_{B_r^2}(z) \\ \text{s.t.} & x - b = z \end{cases}$

(b) $L_\sigma(x, z; \xi) = \|x\|_1 + \delta_{B_r^2}(z) + \langle \xi, x-b-z \rangle + \frac{\sigma}{2}\|x-b-z\|_2^2$

$= \|x\|_1 + \delta_{B_r^2}(z) + \frac{\sigma}{2}\|x-b-z+\sigma^{-1}\xi\|_2^2 - \frac{1}{2\sigma}\|\xi\|_2^2$

① subproblem − x : $\tilde{x} = \underset{x}{\operatorname{argmin}}\ L_\sigma(x, z; \xi)$

$= \underset{x}{\operatorname{argmin}}\ \|x\|_1 + \frac{\sigma}{2}\|x-b-z+\sigma^{-1}\xi\|_2^2$

$= \underset{x}{\operatorname{argmin}}\ \sigma^{-1}\|x\|_1 + \frac{1}{2}\|x-b-z+\sigma^{-1}\xi\|_2^2$

$= P_{\sigma^{-1}\|\cdot\|_1}(b+z-\sigma^{-1}\xi)$

consider what is $P_f(u)$ for $f(x) = \sigma^{-1}\|x\|_1$

since $f(x)$ is a positive homogeneous function,
then $f^*(x) = \delta_{\partial f(0)}(x)$

Then, figure out $\partial f(0)$:

$\xi \in \partial f(0)$

$\Leftrightarrow f(y) - f(0) \geq \xi^T(y-0) \quad \forall y \in \mathbb{R}^n$

$\Leftrightarrow \sigma^{-1}\|y\|_1 \geq \xi^T y \quad \forall y \in \mathbb{R}^n$

(Positive homogeneous) $\Leftrightarrow \xi^T y \leq \sigma^{-1} \quad \forall \|y\|_1 \leq 1,\ y \in \mathbb{R}^n$

$\Leftrightarrow \sup_y \{\langle \xi, y\rangle : \|y\|_1 \leq 1\} \leq \sigma^{-1}$

$\Leftrightarrow \|\xi\|_* \leq \sigma^{-1} \quad \left((\|\cdot\|_1)_* = \|\cdot\|_\infty\right)$

$\Leftrightarrow \|\xi\|_\infty \leq \sigma^{-1}$

Therefore, $\underbrace{\partial f(0) = B^{\infty}_{6^{-1}}}$

$$\Rightarrow f^*(x) = \delta_{B^{\infty}_{6^{-1}}}(x)$$

$$\Rightarrow P_{f^*}(x) = \Pi_{B^{\infty}_{6^{-1}}}(x)$$

$$\Rightarrow P_f(x) = x - P_{f^*}(x)$$
$$= x - \Pi_{B^{\infty}_{6^{-1}}}(x)$$

$$\Leftrightarrow P_f(x)_i = \begin{cases} x_i - x_i = 0, & |x_i| \le 6^{-1} \\ x_i - 6^{-1}, & x_i > 6^{-1} \\ x_i + 6^{-1}, & x_i < -6^{-1} \end{cases}$$

Therefore, since $\tilde{x} = P_f(b + z - 6^{-1}\xi)$

we have (1) $\bar{x} = b + z - 6^{-1}\xi$

(2) $\tilde{x}_i = \begin{cases} 0, & |\bar{x}_i| \le 6^{-1} \\ \bar{x}_i - 6^{-1}, & \bar{x}_i > 6^{-1} \\ \bar{x}_i + 6^{-1}, & \bar{x}_i < -6^{-1} \end{cases}$

② subproblem – z : $\tilde{z} = \underset{z}{\operatorname{argmin}}\ L_6(x, z; \xi)$

$$= \underset{z}{\operatorname{argmin}}\ \delta_{B^2_r}(z) + \frac{6}{2}\| z - x + b - 6^{-1}\xi \|_2^2$$

$$= P_{\delta_{B^2_r}}(x - b + 6^{-1}\xi)$$
$$= \Pi_{B^2_r}(x - b + 6^{-1}\xi)$$

Therefore, we have (1) $\bar{z} = x - b + 6^{-1}\xi$

(2) $\tilde{z} = \begin{cases} \bar{z}, & \|z\|_2 \le r \\ r \cdot \frac{\bar{z}}{\|\bar{z}\|_2}, & \|z\|_2 > r \end{cases}$

Therefore, to conclude. ADMM iteration is:

$$\begin{cases}
\to \text{ at step } k, \text{ we have } \underline{(x^{(k)}, z^{(k)}; \xi^{(k)})}, (\sigma, \tau) \\[1em]
\to \bar{x}^{(k+1)} = b + z^{(k)} - \sigma^{-1} \xi^{(k)} \\
\quad (x^{(k+1)})_i = \begin{cases} 0, & |(\bar{x}^{(k+1)})_i| \le \sigma^{-1} \\ (\bar{x}^{(k+1)})_i - \sigma^{-1}, & (\bar{x}^{(k+1)})_i > \sigma^{-1} \\ (\bar{x}^{(k+1)})_i + \sigma^{-1}, & (\bar{x}^{(k+1)})_i < -\sigma^{-1} \end{cases} \\[1em]
\to \bar{z}^{(k+1)} = x^{(k+1)} - b + \sigma^{-1} \xi^{(k)} \\
\quad z^{(k+1)} = \begin{cases} \bar{z}^{(k+1)}, & \|\bar{z}^{(k+1)}\|_2 \le r \\ r \cdot \dfrac{\bar{z}^{(k+1)}}{\|\bar{z}^{(k+1)}\|_2}, & \|\bar{z}^{(k+1)}\|_2 > r \end{cases} \\[1em]
\to \xi^{(k+1)} = \xi^{(k)} + \tau\sigma (x^{(k+1)} - b - z^{(k+1)})
\end{cases}$$

(c) $r = 1$, $\tau = \sigma = 1$, $b = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \to$ first iteration of ADMM

$x^{(0)} = z^{(0)} = \xi^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Calculation:

① $\bar{x}^{(1)} = b + z^{(0)} - \xi^{(0)}$

$= \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$x^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

② $\bar{z}^{(1)} = x^{(1)} - b + \xi^{(0)}$

$= \begin{pmatrix} -1 \\ -1 \end{pmatrix} \Rightarrow \underline{\|\bar{z}^{(1)}\|_2 = \sqrt{2} > 1}$

thus, $z^{(1)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$

③ $\xi^{(1)} = \xi^{(0)} + (x^{(1)} - b - z^{(1)}) = \begin{pmatrix} -1 + \frac{\sqrt{2}}{2} \\ -1 + \frac{\sqrt{2}}{2} \end{pmatrix}$

Question 10 : [Coordinate Descent] $\rightarrow$ $(x^{(1)}, x^{(2)})$ given $x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\min_{x_1, x_2} \quad x_1^4 - 4x_1 x_2 + 5x_2^2 - 10x_2 := f(x_1, x_2)$$

Solution : first, consider the iteration formula.

(1) $$\tilde{x}_1 = \arg\min_{x_1} f(x_1, x_2)$$

$$= \arg\min_{x_1} x_1^4 - 4x_1 x_2$$

$$= x_2^{1/3}$$

(2) $$\tilde{x}_2 = \arg\min_{x_2} f(x_1, x_2)$$

$$= \arg\min_{x_2} -4x_1 x_2 + 5x_2^2 - 10x_2$$

$$= \frac{2}{5} x_1 + 1$$

① Therefore, to construct $x^{(1)}$: $\left(x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right)$

$$x_1^{(1)} = (x_2^{(0)})^{1/3} = 0$$

$$x_2^{(1)} = \frac{2}{5} x_1^{(1)} + 1 = 1$$

② to construct $x^{(2)}$:

$$x_1^{(2)} = (x_2^{(1)})^{1/3} = 1.$$

$$x_2^{(2)} = \frac{2}{5} x_1^{(2)} + 1 = \frac{7}{5}$$

Therefore, we have : ① $x^{(1)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ② $x^{(2)} = \begin{pmatrix} 1 \\ 7/5 \end{pmatrix}$

Question 11 : [Transportation Plan]

(a) Formulation :

$$\min_{x_{11}, x_{12}, x_{21}, x_{22}} \quad 2x_{11} + 3x_{12} + x_{21} + 4x_{22}$$

$$\text{s.t.} \quad \left.\begin{aligned} x_{11} + x_{12} &= 0.4 \\ x_{21} + x_{22} &= 0.6 \end{aligned}\right\} \to \text{total supplies requirement}$$

$$\left.\begin{aligned} x_{11} + x_{21} &= 0.2 \\ x_{12} + x_{22} &= 0.8 \end{aligned}\right\} \to \text{meet demands}$$

$$x_{11}, x_{12}, x_{21}, x_{22} \geq 0 \longrightarrow \text{non-negative constraint}$$

(b) Langrangian $L(x; \mu, \lambda) = \sum_{i=1}^{2} \sum_{j=1}^{2} C_{ij} x_{ij} + \mu_1 (x_{11} + x_{12} - 0.4)$
$+ \mu_2 (x_{21} + x_{22} - 0.6) + \mu_3 (x_{11} + x_{21} - 0.2) + \mu_4 (x_{12} + x_{22} - 0.8)$
$- \lambda_1 x_{11} - \lambda_2 x_{12} - \lambda_3 x_{21} - \lambda_4 x_{22}$

Then, KKT Condition for $(x^*; \mu^*, \lambda^*)$ can be written as:

① Stationarity Condition :

$$\nabla_x L(x^*; \mu^*, \lambda^*) = 0$$

$$\Leftrightarrow \begin{cases} 2 + \mu_1^* + \mu_3^* - \lambda_1^* = 0 \\ 3 + \mu_1^* + \mu_4^* - \lambda_2^* = 0 \\ 1 + \mu_2^* + \mu_3^* - \lambda_3^* = 0 \\ 4 + \mu_2^* + \mu_4^* - \lambda_4^* = 0 \end{cases}$$

② Primal Feasibility :

$$\begin{cases} x_{11}^* + x_{12}^* = 0.4 \\ x_{21}^* + x_{22}^* = 0.6 \\ x_{11}^* + x_{21}^* = 0.2 \\ x_{12}^* + x_{22}^* = 0.8 \end{cases}$$

③ Dual Feasibility :

a) $\mu^* \in \mathbb{R}^4 \Leftrightarrow \mu_i^* \in \mathbb{R} \quad i \in [4]$

b) $\lambda^* \in \mathbb{R}^4_+ \Leftrightarrow \lambda_i^* \geq 0 \quad i \in [4]$

④ Complementary Slackness Condition:

$$\lambda_1^* x_{11}^* = 0 \qquad \lambda_3^* x_{21}^* = 0$$

$$\lambda_2^* x_{12}^* = 0 \qquad \lambda_4^* x_{22}^* = 0$$

(c) Feasible point 1 : $x^{(1)}$

here, $\begin{cases} x_{11}^{(1)} = 0.2 \\ x_{12}^{(1)} = 0.2 \\ x_{21}^{(1)} = 0 \\ x_{22}^{(1)} = 0.6 \end{cases}$

Feasible point 2 : $x^{(2)}$

here, $\begin{cases} x_{11}^{(2)} = 0 \\ x_{12}^{(2)} = 0.4 \\ x_{21}^{(2)} = 0.2 \\ x_{22}^{(2)} = 0.4 \end{cases}$

(d) Compare between costs for $x^{(1)}$ & $x^{(2)}$

Calculation : $f(x^{(1)}) = 2 \times 0.2 + 3 \times 0.2 + 4 \times 0.6$

$$= 0.4 + 0.6 + 2.4$$

$$= 3.4$$

$$f(x^{(2)}) = 3 \times 0.4 + 1 \times 0.2 + 4 \times 0.4$$

$$= 1.2 + 0.2 + 1.6$$

$$= 3$$

since $f(x^{(2)}) = 3 < 3.4 = f(x^{(1)})$,

it implies that transportation plan $x^{(2)}$ is better!

Question 12 : $\begin{cases} \max\limits_{x_1, x_2} & 6x_1 + 4x_2 - 13 - x_1^2 - x_2^2 \\ \text{s.t} & x_1 + x_2 \le 3 \end{cases}$ (1)

(a) ① transform to equivalent minimization problem

② Solve for $(x_1^*; x_2^*)$

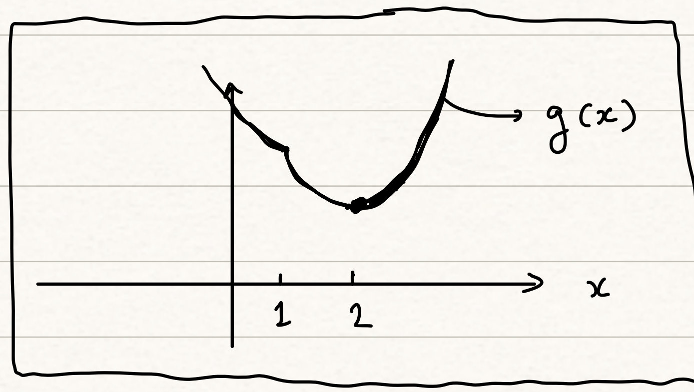
Solution : (1) $\Leftrightarrow$ $\begin{cases} -\min\limits_{x_1, x_2} & (x_1-3)^2 + (x_2-2)^2 \\ \text{s.t} & x_1 + x_2 \le 3 \end{cases}$

then we focus on $\begin{cases} \min\limits_{x_1, x_2} & (x_1-3)^2 + (x_2-2)^2 \\ \text{s.t} & (x_1 + x_2) \le 3 \end{cases}$

$$\Leftrightarrow \min_{x} \left\{ (x-3)^2 + \underbrace{\left[\min_{x_2 \le 3-x} (x_2-2)^2 \right]}_{g(x)} \right\}$$

Here, $g(x) = \begin{cases} (x-3)^2 + 0, & \text{if } x \le 1 \\ (x-3)^2 + (1-x)^2, & \text{if } x > 1 \end{cases}$

its graph is shown as follows:



Therefore, we achieve the minimizer at $\underline{x = 2}$

since $\underline{x = 2 > 1}$, it implies the optimal $x_2^* = 3 - x = 1$

Therefore, $\begin{cases} x_1^* = x = 2 \\ x_2^* = 1 \end{cases}$

(b) KKT Condition :

Solution : (1) $\Leftrightarrow$ $\begin{cases} -\min_x & (x_1-3)^2+(x_2-2)^2 \\ \text{s.t} & x_1+x_2 \le 3 \end{cases}$

Therefore $(x_1^*, x_2^*) \in \arg\max_x \; -(x_1-3)^2-(x_2-2)^2$

s.t $x_1+x_2 \le 3$

$\Leftrightarrow (x_1^*, x_2^*) \in \arg\min_x \; (x_1-3)^2+(x_2-2)^2$

s.t $x_1+x_2 \le 3$

$\Rightarrow$ denote $L(x;\mu) = (x_1-3)^2+(x_2-2)^2+\mu(x_1+x_2-3)$

KKT Condition is , there exists μ^* such that

KKT Condition

① Sationarity Condition:

$\nabla_x L(x^*;\mu^*)=0 \Rightarrow \begin{cases} 2(x_1^*-3)+\mu^*=0 \\ 2(x_2^*-2)+\mu^*=0 \end{cases}$

② Primal Feasibility:

$x_1^*+x_2^*-3 \le 0$

③ Dual Feasibility:

$\mu^* \ge 0$

④ Complementary Slackness Condition:

$\mu^*(x_1^*+x_2^*-3)=0$

(c) Lagrangian Dual Problem :

Solution : denote $f(x) = x_1^2+x_2^2-6x_1-4x_2+13$

then our problem is exactly $\begin{cases} -\min_x & f(x) \\ \text{s.t} & x_1+x_2 \le 3 \end{cases}$

Therefore, the dual Problem is derived as follows,

$$L(x;\mu) = f(x)+\mu(x_1+x_2-3)$$

$$= x_1^2+x_2^2-6x_1-4x_2+13+\mu(x_1+x_2-3)$$

$$\theta(\mu) = \min_{x} L(x; \mu)$$

$$= \left(\min_{x_1} x_1^2 - 6x_1 + \mu x_1\right) + \left(\min_{x_2} x_2^2 - 4x_2 + \mu x_2\right)$$

$$+ 13 - 3\mu$$

$$= -\frac{1}{4}(6-\mu)^2 - \frac{1}{4}(4-\mu)^2 + 13 - 3\mu$$

$$= -\frac{1}{2}\mu^2 + 2\mu$$

Therefore, the dual problem for $\begin{cases} \min & f(x) \\ \text{s.t.} & x_1 + x_2 \leq 3 \end{cases}$ is:

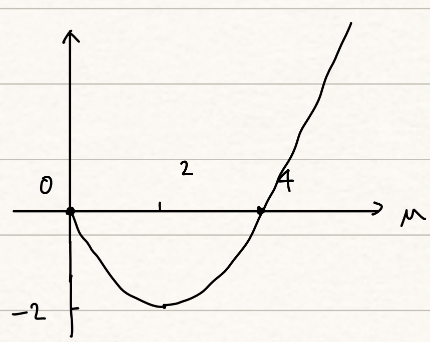
$$\begin{cases} \max\limits_{\mu} & \theta(\mu) \\ \text{s.t} & \mu \geq 0 \end{cases} \Rightarrow \begin{cases} \max\limits_{\mu} & -\frac{1}{2}\mu^2 + 2\mu \\ \text{s.t} & \mu \geq 0 \end{cases}$$

Therefore, the dual problem for $\begin{cases} \max & -f(x) \\ \text{s.t} & x_1 + x_2 \leq 3 \end{cases}$ is:

$$\begin{cases} -\max\limits_{\mu} & \theta(\mu) \\ \text{s.t} & \mu \geq 0 \end{cases} \Rightarrow \star \begin{cases} \min\limits_{\mu} & \frac{1}{2}\mu^2 - 2\mu \\ \text{s.t} & \mu \geq 0 \end{cases}$$

(d) Solve Lagrangian Dual Problem

Solution: We can draw the graph of $g(\mu) = \frac{1}{2}\mu^2 - 2\mu$:



Therefore, we achieve the minima at $\mu^* = 2$

And the corresponding objective value $= -2$

That is, the dual solution $y^*(\mu^*) = 2$

$\boxed{\text{Extra Part}} \rightarrow$ consider

$$\begin{cases} \min\limits_{x} & f(x) \\ s.t. & x_1 + x_2 \le 3 \end{cases} \qquad (*)$$

$$\begin{cases} f(x) = (x_1 - 3)^2 + (x_2 - 2)^2 \rightarrow \text{convex} \\ x_1 + x_2 \le 3 \rightarrow \text{affine inequality constraint} \end{cases}$$

From weak Slater's condition, it implies Strong Duality holds

Moreover, $(*)$ is a convex programming

Actually we have the conclusion that:

If strong duality holds, $(*)$ is convex programming, then

$(x^*; \mu^*)$ is the (primal optimal solution; dual optimal solution)

$\Leftrightarrow$ $(x^*; \mu^*)$ satisfies KKT Condition

This means that Q12(e) must hold generally if the optimization when Strong Duality holds

(e) verify that $(x_1^*, x_2^*; y^*)$ satisfy KKT Condition

Verification: Here $\begin{cases} x_1^* = 2 \\ x_2^* = 1 \end{cases}$ & $y^* = 2$

Recap: KKT Condition is:

① Sationarity Condition:

$$\nabla_x L(x^*; \mu^*) = 0 \Rightarrow \begin{cases} 2(x_1^* - 3) + \mu^* = 0 \\ 2(x_2^* - 2) + \mu^* = 0 \end{cases}$$

② Primal Feasibility:

$$x_1^* + x_2^* - 3 \leq 0$$

③ Dual Feasibility:

$$\mu^* \geq 0$$

④ Complementary Slackness Condition:

$$\mu^* (x_1^* + x_2^* - 3) = 0$$

$\rightarrow$ ① $\begin{cases} 2(2-3) + 2 = 0 \\ 2(1-2) + 2 = 0 \end{cases}$

$\rightarrow$ ② $2 + 1 - 3 = 0 \leq 0$

$\rightarrow$ ③ $\mu^* = 2 > 0$

$\rightarrow$ ④ $2(2+1-3) = 0$

$\Rightarrow$ $(x_1^*, x_2^*; y^*)$ satisfies KKT Condition!

#