CS5321 Numerical Optimization Homework 3

Due 1/6/2023

1. (20%) In the trust region method (unit 3), we need to solve the model problem m_k

$$\min_{\vec{p}} m_k(\vec{p}) = f_k + \vec{g}_k^T \vec{p} + \frac{1}{2}\vec{p}^{\,T} B_k \vec{p}.$$

$$\text{s.t.} \|\vec{p}\| \leq \Delta$$

Show that $\vec{p}*$ is the optimal solution if and only if it satisfies

$$(B_k + \lambda I)\vec{p}* = -\vec{g}$$

$$\lambda(\Delta - \|\vec{p}*\|) = 0$$

where $B_k + \lambda I$ is positive definite. (Hint: using KKT conditions.)

A：參考自講義第三章與第十章

The Lagrangian is

$$\mathcal{L} = f_k + \vec{g}_k^{\mathrm{T}} \vec{p} + \frac{1}{2}\vec{p}^{\mathrm{T}} B_k \vec{p} - \mu(\Delta - \|\vec{p}\|)$$

which μ is the Lagrange multiplier.

另因 $\|\vec{p}\| = \sqrt{\vec{p}^{\mathrm{T}}\vec{p}}$，constraint 可以改寫成 $\vec{p}^{\mathrm{T}}\vec{p} \leq \Delta^2$

Lagrangian 就變為

$$\mathcal{L} = f_k + \vec{g}_k^{\mathrm{T}} \vec{p} + \frac{1}{2}\vec{p}^{\mathrm{T}} B_k \vec{p} - \mu(\Delta^2 - \vec{p}^{\mathrm{T}}\vec{p})$$

根據 KKT condition

$$\nabla\mathcal{L} = 0$$

也就是

$$\nabla\mathcal{L} = \vec{g} + B_k\vec{p} + 2\mu\vec{p} = 0$$

移項獲得

$$(B_k + 2\mu I)\vec{p} = -\vec{g}$$

再將 2μ 替換成 λ，即可獲得題目條件

$$(B_k + \lambda I)\vec{p}^* = -\vec{g}$$

又 $\mu(\Delta - \|\vec{p}\|)$，將 μ 替換成 $\frac{1}{2}\lambda$，即可獲得另一互補條件

$$\lambda(\Delta - \|\vec{p}^*\|) = 0$$

再根據 KKT condition，$\lambda \geq 0$，

若 λ 等於 0 時，代表 $\mathcal{L}$ 的最後一項 constraint，將沒有意義，optimal solution 與 constraint 沒有關係，則 $B_k + \lambda I = B_k$ 必為半正定 (positive semi-definite).

而 λ 大於 0 時，因 $\lambda(\Delta - \|\vec{p}^*\|) = 0$，代表 $\Delta = \|\vec{p}^*\|$，所以 $\vec{p}^*$ is the minimizer，則 $m_k(\vec{p}^*) \leq m_k(\vec{p})$

$$f_k + \vec{g}_k^{\mathrm{T}}\vec{p}^* + \frac{1}{2}(\vec{p}^*)^{\mathrm{T}} B_k \vec{p}^* \leq f_k + \vec{g}_k^{\mathrm{T}}\vec{p} + \frac{1}{2}\vec{p}^{\mathrm{T}} B_k \vec{p}$$

將 $(B_k + \lambda I)\vec{p}^* = -\vec{g}$ 代入

$$f_k - (\vec{p}^*)^{\mathrm{T}}(B_k + \lambda I)\vec{p}^* + \frac{1}{2}(\vec{p}^*)^{\mathrm{T}} B_k \vec{p}^* \leq f_k - \vec{p}^{\mathrm{T}}(B_k + \lambda I)\vec{p} + \frac{1}{2}\vec{p}^{\mathrm{T}} B_k \vec{p}$$

再將不等式兩側加上 $\frac{1}{2}\lambda\Delta^2$，得

$$f_k - (\vec{p}^*)^{\mathrm{T}}(B_k+\lambda I)\vec{p}^* + \frac{1}{2}(\vec{p}^*)^{\mathrm{T}}(B_k+\lambda I)\vec{p}^* \leq f_k - \vec{p}^{\mathrm{T}}(B_k+\lambda I)\vec{p} + \frac{1}{2}\vec{p}^{\mathrm{T}}(B_k+\lambda I)\vec{p}$$

經整理得

$$0 \leq (\vec{p}^* - \vec{p})^{\mathrm{T}}(B_k + \lambda I)(\vec{p}^* - \vec{p})$$

又 $\vec{p}$ 可以為任意向量，則 $B_k + \lambda I$ 必為半正定 (positive semi-definite)

所以題目欲證明應改為：where $B_k + \lambda I$ is positive semi-definite.

2. (15%) Prove that for the matrix $\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$, if A has full row-rank and the reduced Hessian $Z^T G Z$ is positive definite, where span$\{Z\}$ is the null space of span$\{A^T\}$ then the matrix is nonsingular. (You may reference Lemma 16.1 in the textbook.)
A：參考自課本第 16.1 章節

$\begin{bmatrix} G & A^{\mathrm{T}} \\ A & 0 \end{bmatrix}$ 稱為 KKT matrix，假設有兩個向量 w 和 v 會使

$$\begin{bmatrix} G & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = 0$$

乘開得

$$Gw + A^{\mathrm{T}}v = 0, \quad Aw = 0,$$

又

$$\begin{bmatrix} G & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = 0 = \begin{bmatrix} w \\ v \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} G & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix}$$

也就是

$$w^{\mathrm{T}}Gw + w^{\mathrm{T}}A^{\mathrm{T}}v + v^{\mathrm{T}}Aw = 0,$$

又因為 $Aw = 0$，所以 w 一定在 A 的 null space 中，

且因

$$Aw = 0 = (Aw)^{\mathrm{T}} = w^{\mathrm{T}}A^{\mathrm{T}},$$

得

$$w^{\mathrm{T}}Gw + w^{\mathrm{T}}A^{\mathrm{T}}v + v^{\mathrm{T}}Aw = 0 = w^{\mathrm{T}}Gw,$$

再令 $w = Zu$，得

$$0 = w^{\mathrm{T}}Gw = u^{\mathrm{T}}Z^{\mathrm{T}}GZu$$

上式等號右側，若 $Z^{\mathrm{T}}GZ$ is positive definite，則代表 $u = 0$，
上式等號左側，因 G 一定不為零，所以 $w = 0$，再代入 $Gw + A^{\mathrm{T}}v = 0$，
得 $A^{\mathrm{T}}v = 0$，又 A 為 full row-rank，代表 $v = 0$，

至此，可以得知 $\begin{bmatrix} G & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = 0$，一定只成立於 $w = 0$ 和 $v = 0$，
所以得證：KKT matrix is nonsingular.

3. (30%) Consider the problem

$$\begin{aligned} \min_{x_1, x_2} \quad & (x_1 - 3)^2 + 10x_2^2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 1 \le 0 \end{aligned} \tag{1}$$

(a) Write down the KKT conditions for (1).

(b) Solve the KKT conditions and find the optimal solutions, including the Lagrangian parameters.

(c) Compute the reduced Hessian and check the second order conditions for the solution.

A：參考自講義第九章第 19、21 頁與第十章第 3 頁

(a) The Lagrangian is

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = (x_1 - 3)^2 + 10x_2^2 - \lambda(-x_1^2 - x_2^2 + 1)$$

which λ is the Lagrange multiplier.

The KKT conditions：

$$\nabla \mathcal{L}(\vec{x}, \vec{\lambda}) = 0$$

$$\vec{a_i}^{\mathrm{T}}\vec{x} = b_i, \quad i \in \mathcal{E}$$

$$\vec{a_i}^{\mathrm{T}}\vec{x} \ge b_i, \quad i \in \mathcal{I}$$

$$\lambda_i \ge 0, \quad i \in \mathcal{I}$$

$$\lambda_i(\vec{a_i}^{\mathrm{T}}\vec{x} - b_i) = 0, \quad i \in \mathcal{I}$$

$\Rightarrow$

$$\nabla \mathcal{L}(\vec{x}, \vec{\lambda}) = \begin{bmatrix} 2(x_1 - 3) + 2\lambda x_1 \\ 20x_2 + 2\lambda x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-x_1^2 - x_2^2 + 1 \ge 0$$

$$\lambda \ge 0$$

$$\lambda(-x_1^2 - x_2^2 + 1) = 0$$

(b) Solve the KKT conditions can get

$$\vec{x}^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \vec{\lambda}^* = \lambda^* = 2$$

(c) The critical cone $\mathcal{C}(\vec{x}^*) = \{\begin{bmatrix} 0 \\ w_2 \end{bmatrix} \mid w_2 \in \mathbb{R}\}$,

by the optimal solutions of (b), the Hessian matrix of $\mathcal{L}(\vec{x}^*, \vec{\lambda}^*)$ is

$$\nabla^2 \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = \begin{bmatrix} 2 + 2\lambda^* & 0 \\ 0 & 20 + 2\lambda^* \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 24 \end{bmatrix}$$

The reduced Hessian(projected Hessian) is

$$\vec{w}^{\mathrm{T}} \nabla^2_{xx} \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) \vec{w} = \begin{bmatrix} 0 & w_2 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 24 \end{bmatrix} \begin{bmatrix} 0 \\ w_2 \end{bmatrix} = 24 w_2^2 \geq 0$$

so, the second order condition is fit.

4. (20%) Consider the following constrained optimization problem

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1^3 + 2x_2^2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

(a) What is the optimal solution and the optimal Lagrangian multiplier?

(b) Formulate this problem to the equation of augmented Lagrangian method, and derive the gradient of Lagrangian.

(c) Let $\rho_0 = -1, \mu_0 = 1$. What is x_1 if it is solved by the augmented Lagrangian problem.

(d) To make the solution of augmented Lagrangian method exact, what is the minimum ρ should be?

A：參考自講義第九章與第十一章

(a) The Lagrangian is

$$\mathcal{L} = x_1^3 + 2x_2^2 - \lambda(x_1^2 + x_2^2 - 1)$$

which λ is the Lagrange multiplier.

根據 KKT condition

$$\nabla \mathcal{L} = 0$$

也就是

$$\nabla \mathcal{L} = \begin{bmatrix} 3x_1^2 - 2\lambda x_1 \\ 4x_2 - 2\lambda x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

So the optimal solution and the optimal Lagrangian multiplier is

$$\vec{x}^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \lambda^* = 1.5$$

or

$$\vec{x}^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \lambda^* = -1.5$$

但考慮到 $\underset{x_1,x_2}{min} \quad x_1^3 + 2x_2^2$，後者較符合。

(b) 根據講義第十一章第 5 頁，
Augmented Lagrangian method 的 Lagrangian 可以寫為

$$\mathcal{L}(\vec{x}, \vec{\rho}, \mu) = f(\vec{x}) - \sum_{i \in \mathcal{E}} \rho_i c_i(\vec{x}) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(\vec{x})$$

以 ρ 和 μ 合成 λ，

$$\mathcal{L}(\vec{x}, \vec{\rho}, \mu) = x_1^3 + 2x_2^2 - \rho(x_1^2 + x_2^2 - 1) + \frac{\mu}{2}(x_1^2 + x_2^2 - 1)^2$$

The gradient of Lagrangian is

$$\nabla \mathcal{L} = \begin{bmatrix} 3x_1^2 - 2\rho x_1 + 2\mu(x_1^2 + x_2^2 - 1)x_1 \\ 4x_2 - 2\rho x_1 + 2\mu(x_1^2 + x_2^2 - 1)x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(c) 將 $\rho_0 = -1$ 和 $\mu_0 = 1$ 代入上式，得到

$$\nabla \mathcal{L} = \begin{bmatrix} 3x_1^2 + 2x_1^3 + 2x_1 x_2^2 \\ 4x_2 + 2x_1^2 x_2 + 2x_2^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

再經過計算，扣除掉皆為零或虛數的解，可得到一個較符合的 $\vec{x}$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix}$$

所以 $x_1 = -1.5.$

(d) 將 (a) 已求得的 optimal solution $\vec{x}^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 代入 (b) 的 $\nabla \mathcal{L}$ 中，也就是 $3 + 2\rho = 0$，可得 $\rho = -\frac{3}{2}$。

5. (15%) Consider the following constrained optimization problem

$$\begin{aligned} \min_{x_1,x_2} \quad & -3x_1 + x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \leq 20 \\ & x_1 + 2x_2 \leq 16 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Formulate this problem to the equation of the interior point method, and derive the gradient and Jacobian.

A：參考自講義第十一章

根據上課講義，題目可以改寫成

$$\begin{aligned} \min_{\vec{x}} \quad & f(\vec{x}) \\ \text{s.t.} \quad & C_E(\vec{x}) = 0 \\ & C_I(\vec{x}) - \vec{s} = 0 \\ & \vec{s} \geq 0 \end{aligned}$$

其中 $f(\vec{x}) = -3x_1 + x_2, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{s}$ are slack variables，且
$C_E(\vec{x})$ is equality constraints,
$C_I(\vec{x}) - \vec{s}$ is inequality constraints.

The Lagrangian is

$$\mathcal{L}(\vec{x}, \vec{s}, \vec{y}, \vec{z}) = f(\vec{x}) - \mu \sum_{i=1}^{m} log(s_i) - \vec{y}^{\mathrm{T}} C_E(\vec{x}) - \vec{z}^{\mathrm{T}}(C_I(\vec{x}) - \vec{s})$$

其中，μ is barrier parameter，且
$\vec{y}$ is the Lagrangian multiplier of equality constraints,
$\vec{z}$ is the Lagrangian multiplier of inequality constraints

因本題沒有 equality 的 constraints，所以沒有 $\vec{y}$ 和 C_E，

$\mathcal{L}(\vec{x}, \vec{s}, \vec{y}, \vec{z}) = -3x_1 + x_2 - \mu(ln(x_1) + ln(x_2) + ln(20 - 2x_1 - x_2) + ln(16 - x_1 - 2x_2))$

又根據上課講義，gradient of Lagrangian is

$$F = \begin{bmatrix} \nabla_x \mathcal{L} \\ \nabla_s \mathcal{L} \\ \nabla_y \mathcal{L} \\ \nabla_z \mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla f - A_E \vec{y} - A_I \vec{z} \\ S\vec{z} - \mu\vec{e} \\ C_E(\vec{x}) \\ C_I(\vec{x}) - \vec{s} \end{bmatrix}$$

The Jacobian is

$$J = \nabla F = \begin{bmatrix} \nabla_{xx} \mathcal{L} & 0 & -A_E(\vec{x}) & -A_I(\vec{x}) \\ 0 & Z & 0 & S \\ A_E(\vec{x}) & 0 & 0 & 0 \\ A_I(\vec{x}) & -I & 0 & 0 \end{bmatrix}$$

Matrix A_E is the Jacobian of C_E and matrix A_I is the Jacobian of C_I , matrix $S = \text{diag}(\vec{s})$ and matrix Z $= \text{diag}(\vec{z})$.

但因本題沒有 equality 的 constraints，所以應沒有 $\nabla_y \mathcal{L}$ 的那行，

不過，根據網路上的查詢，gradient of Lagrangian 中的 $\nabla_x \mathcal{L}$ 裡

$$\nabla_x(-\vec{y}^{\mathrm{T}} C_E(\vec{x}))^{\mathrm{T}} = -\nabla_x C_E(\vec{x})^{\mathrm{T}} \vec{y} = -A_E^{\mathrm{T}} \vec{y}$$

也就是 A_E 和 A_I 上應該還要有個轉置，不知是否遺漏？

公式應為

$$F = \begin{bmatrix} \nabla_x \mathcal{L} \\ \nabla_s \mathcal{L} \\ \nabla_y \mathcal{L} \\ \nabla_z \mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla f - {A_E}^{\mathrm{T}} \vec{y} - {A_I}^{\mathrm{T}} \vec{z} \\ S\vec{z} - \mu \vec{e} \\ C_E(\vec{x}) \\ C_I(\vec{x}) - \vec{s} \end{bmatrix}$$

$$J = \nabla F = \begin{bmatrix} \nabla_{xx}\mathcal{L} & 0 & -{A_E}^{\mathrm{T}}(\vec{x}) & -{A_I}^{\mathrm{T}}(\vec{x}) \\ 0 & Z & 0 & S \\ A_E(\vec{x}) & 0 & 0 & 0 \\ A_I(\vec{x}) & -I & 0 & 0 \end{bmatrix}$$

因不知上述公式是否正確，也可採用 linear programming 算法另解，

The primal problem：

$$\begin{aligned} & \min_{\vec{x}} && \vec{c}^{\mathrm{T}} \vec{x} \\ & \text{s.t.} && A\vec{x} = \vec{b} \\ & && \vec{x} \geq 0 \\ \Rightarrow\ & \min_{x_1, x_2} && -3x_1 + x_2 \\ & \text{s.t.} && 2x_1 + x_2 + x_3 = 20 \\ & && x_1 + 2x_2 + x_4 = 16 \\ & && x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

The dual problem：

$$\begin{aligned} & \max_{\vec{\lambda}} && \vec{b}^{\mathrm{T}} \vec{\lambda} \\ & \text{s.t.} && A^{\mathrm{T}} \lambda + \vec{s} = \vec{c} \\ & && \vec{s} \geq 0 \\ \Rightarrow\ & \max_{\lambda_1, \lambda_2} && 20\lambda_1 + 16\lambda_2 \\ & \text{s.t.} && 2\lambda_1 + \lambda_2 + s_1 = -3 \\ & && \lambda_1 + 2\lambda_2 + s_2 = 1 \\ & && \lambda_1 + s_3 = 0 \\ & && \lambda_2 + s_4 = 0 \\ & && s_1, s_2, s_3, s_4 \geq 0 \end{aligned}$$

其中 $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$， $\vec{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$， $\vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}$，

且 $\vec{c} = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}$， $A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}$， $\vec{b} = \begin{bmatrix} 20 \\ 16 \end{bmatrix}$

根據上課講義，

$$F = \begin{bmatrix} A^{\mathrm{T}}\vec{\lambda} + \vec{s} - \vec{c} \\ A\vec{x} - \vec{b} \\ X\vec{s} - \mu\vec{e} \end{bmatrix} = \left[\begin{array}{c} 2\lambda_1 + \lambda_2 + s_1 + 3 \\ \lambda_1 + 2\lambda_2 + s_2 - 1 \\ \lambda_1 + s_3 \\ \lambda_2 + s_4 \\ \hline 2x_1 + x_2 + x_3 - 20 \\ x_1 + 2x_2 + x_4 - 16 \\ \hline x_1 s_1 \\ x_2 s_2 \\ x_3 s_3 \\ x_4 s_4 \end{array}\right]$$

$$\nabla F = \begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} = \left[\begin{array}{cccc|cc|cccc} 0 & 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline s_1 & 0 & 0 & 0 & 0 & 0 & x_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 & 0 & 0 & x_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 & 0 & 0 & 0 & 0 & x_3 & 0 \\ 0 & 0 & 0 & s_4 & 0 & 0 & 0 & 0 & 0 & x_4 \end{array}\right]$$