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.)
lagrangian function

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

j為langrangian multiplier
根據KKT condition 的定義

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

$$\|\vec{p}\| = \sqrt{\vec{p}^T \vec{p}}$$

所以

$$\nabla_{\vec{p}} \mathcal{L} = \vec{g} + B_k \vec{p} + 2j\vec{p} = 0$$

經過整理

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

將$2j = \lambda$，則會變成，又只有達到極值，才會出現$\nabla \mathcal{L}$

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

又根據朗葛朗日乘數的定義constrain function 要等於零，因為達到極值，會符合$\nabla \mathcal{L} = \nabla m(\vec{p})$，所以

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

也可以寫成

$$\frac{1}{2}\lambda(\Delta - \|\vec{p}^*\|) = 0 \Rightarrow \lambda(\Delta - \|\vec{p}^*\|) = 0$$

這時要探討一個問題λ的大小，λ必定大於等於零(朗葛朗日乘數的條件)
但如果今天極值情況發生在$\lambda = 0$的情況，表示此問題的最佳解與其constrain function沒有關係，則B_k必定為半正定，如果不為半正定，等於m_k屬於一個有最大值，而沒有最小值的問題(二次曲線定義)，則至

信譽半徑的方向會隨著Δ變動，變成沒有意義

現在探討$\lambda > 0$的問題如果今天能找到最小值p^*
可以表示成下列

$$f_k + \vec{g}_k^T \vec{p} + 0.5\vec{p}^T B_k \vec{p} \geq f_k + \vec{g}_k^T \vec{p}^* + 0.5(\vec{p}^*)^T B_k \vec{p}^*$$

將上面的方程式帶入，因為$\vec{g}$為常數

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

$f_k - \vec{p}^T(B_k + \lambda I)\vec{p} + 0.5\vec{p}^T B_k \vec{p} \geq f_k - (\vec{p}^*)^T(B_k + \lambda I)\vec{p}^* + 0.5(\vec{p}^*)^T B_k \vec{p}^*$
將兩邊都加上$\frac{1}{2}\lambda\Delta^2$ 整理後為下列

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

又$\vec{p}$為任意實數，所以上述的式子也證明了$(B_k + \lambda I)$為半正定

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^TGZ 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.)
先假設有兩個vaectors w and v, KKT matrix 乘上這兩個Vectors 等於0

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

$$\to Gw + A^T v = 0,\ Aw = 0$$

設

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

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

$$\to w^T G w + v^T A w + w^T A^T v = 0$$

又

$$Aw = 0 = (Aw)^T = w^T A^T$$

$$w^T G w + v^T A w + w^T A^T v = w^T G w$$

因為$Aw = 0$，所以w一定落在A的null space 設

$$w = zy$$

因為span z is the null space of span A

$$w^T G w = 0 = y^T z^T G z y$$

如果z^TGz為正定，則表示z^TGz一定大於零，所以$y=0$才會符合上述的式子又G一定不為零，上述式子的左式$w=0$ 再看一開始的function

$$Gw + A^Tv = 0 \xrightarrow{w=0} A^Tv = 0$$

又A為Full rank表$v=0$ 當$w=v=0$ 才可以確認KKT matrix是nonsingular matrix 因為nonsingular matrix 表示$\det[KKTmatrix] \neq 0$
這邊證明為何成立
設

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

當M是nonsingular matrix，表示其$\det M \neq 0$，所以m的反矩陣存在

$$\Rightarrow M^{-1}M \begin{bmatrix} w \\ v \end{bmatrix} = M^{-1} * 0 \Rightarrow \begin{bmatrix} w \\ v \end{bmatrix} = 0$$

所以w and v都是零

3. (30%) Consider the problem

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

(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.

參考自講義第九章!!
(a)

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

The KKT conditions 為下列五個(from 講義)

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

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

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

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

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

將其算出來為下列

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

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

$$\lambda \geq 0$$

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

(b)
藉由上面的公式，有三個變數，λ, x_1, x_2

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

$$2(x_1 - 3) + 2\lambda x_1 = 0$$

$$20x_2 + 2\lambda x_2 = 0$$

可以看出，第一條式子可以得出x_1, x_2其中之一為零，又看向第三條，又$\lambda \neq 0$ ，所以可以得出$x_2 = 0$，最後將其放入到第二條，得到$\lambda = 2$ 最佳解為

$$x_1^* = 1, x_2^* = 0, \lambda^* = 2$$

(c)
The critical cone

$$\mathcal{C}(\vec{x}^*) = \{\begin{bmatrix} 0 \\ w_2 \end{bmatrix}, w_2 \in R\}$$

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 projected hessian is

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

因為w_2^2必定大於零或等於零
這樣$f(x^*)$才會比周邊的$f(\vec{z})$來的小，表示x^*為local minima，符合The second order necessary condition.

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 the augmented Lagrangian method exact, what is the minimum ρ should be?

(a)
lagrangian function

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

gradient of lagrangian function為下列

$$\nabla\mathcal{L} = \begin{bmatrix} \nabla_{x_1}\mathcal{L} \\ \nabla_{x_2}\mathcal{L} \\ \nabla_{\lambda}\mathcal{L} \end{bmatrix} = \begin{bmatrix} 3x_1^2 - \lambda 2x_1 \\ 4x_2 - x\lambda x_2 \\ -x_1^2 - x_2^2 + 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

得出

$$x_1^* = 1, x_2^* = 0, \lambda = 1.5$$

或是

$$x_1^* = -1, x_2^* = 0, \lambda = -1.5$$

而將其帶回目標函數，$x_1^* = -1, x_2^* = 0, \lambda = -1.5$有最小值，所以optimal solution and optimal lagrangian multiplier 為

$$x_1^* = -1, x_2^* = 0, \lambda = -1.5$$

(b)
根據講義的example
the augmented lagrangian function,ρ和μ合起來為langrangian multiplier

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

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

the gradient of lagrangian

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

(c)
將$\rho_0 = -1$ and $\mu_0 = 1$帶入the augmented lagrangian function
得出

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

也有可能為都是零的情況，但這是trivial soultion，就不會納入探討，如果x_1 and x_2都不等於零，會出現虛數，也不符合題目，所以不納入討論
(d)
已經透過(a)選項知道最小值的$\vec{x}$，將其帶入找最小值ρ

$$x_1 = -1, x_2 = 0,\ \lambda = \rho - \mu$$

帶入到gradient of lagrangian裡

$$3 + 2\rho + 2\mu(0) = 0 \Rightarrow \rho = \frac{-3}{2}$$

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

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

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

根據上課講義，題目可以改寫成有equality and inequality equations

$$f(\vec{x}) = -3x_1 + x_2$$

$\vec{s}$是slack variables

$$C_E(\vec{x}) = 0, C_I(\vec{x}) - \vec{s} \ge 0, \vec{s} \ge 0$$

下標E為限制等式函數，I為限制不等式函數
Lagrangian function 為下列

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

μ is barrier parameter，$\vec{y}$為lagrangian multiplier for 限制等式函數, $\vec{z}$則為不等式函數的lagrangian multiplier the

gradient of Lagrangian 為下列

$$F = \begin{bmatrix} \nabla_x \mathcal{L} \\ \nabla_s \mathcal{L} \\ \nabla_y \mathcal{L} \\ \nabla_z \mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla f - A_E \vec{y} - A_I \vec{z} \\ S\vec{z} - \mu \vec{e} \\ C_E(\vec{x}) \\ C_I(\vec{x}) - \vec{s} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

其jacobian 是

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

A_I, A_E是C_I, C_E的jacobian，S, Z 為分別取對角線矩陣$\vec{s}, \vec{z}$ 上面為IPM基本架構
下面會使用課本上教學的IPM 的linear programming來解題將原本的問題寫成求maximum

$$\max_{\lambda_1,\lambda_2} \; 20\lambda_1 + 16\lambda_2$$

$$2\lambda_1 + \lambda_2 + s_1 = -3$$

$$\lambda_1 + 2\lambda_2 + s_2 = 1$$

$$\lambda_1 + s_3 = 0$$
$$\lambda_2 + s_4 = 0$$
$$s_1, s_2, s_3, s_4 \geq 0$$

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

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

根據上課講義，分別寫出其gradient and jacobian，

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

$$\nabla F = \begin{bmatrix} O & A^{\mathrm{T}} & I \\ A & O & O \\ S & O & X \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_1 & 0 & 0 & 0 & 0 & 0 & x_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 & 0 & 0 & x_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 & 0 & 0 & 0 & 0 & x_3 & 0 \\ 0 & 0 & 0 & s_4 & 0 & 0 & 0 & 0 & 0 & x_4 \end{bmatrix}$$