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.)

Answer:

The Lagrangian function for the trust region method is:

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

The KKT condition shows that

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

Let $\lambda = 2\mu$

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

The complementarity condition shows that

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

$\Rightarrow \lambda = 0 \quad or \quad \Delta = \|\vec{p}*\|$

(a) If $\lambda = 0$, the constrain is inactive, $\|\vec{p}*\| < \Delta$. It is an unconstrained optimization problem. So by the second order condition of an unconstrained optimization problem, $B + \lambda I = B$is positive semi-definite.

(b) If $\lambda > 0$, by the complementarity condition, $\|\vec{p}*\| = \Delta$ Thus, we only need to consider the position$\vec{p}|$ such that $\|\vec{p}\| = \Delta$

Since $\vec{p}*$ is the minimizer, $m(\vec{p}*) \leq m(\vec{p})$, which implies that

$$f_k + \vec{g}_k^T\vec{p}* + \frac{1}{2}\vec{p}*^T B_k\vec{p}* \leq f_k + \vec{g}_k^T\vec{p} + \frac{1}{2}\vec{p}^{\,T} B_k\vec{p} \tag{1}$$

使用$\vec{g}_k = -(B_k + \lambda I)\vec{p}*$，代入公式(1)，得到

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

兩邊都加$\frac{1}{2}\lambda\Delta^2$，並且$\|\vec{p}\|^2 = \Delta^2$，因此

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

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

Since the only constraint of $\vec{p}$ is $\|\vec{p}\| = \Delta$，$(\vec{p}* - \vec{p})$ may be any vector $\Rightarrow (B_k + \lambda I)$ is positve 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^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.)

Answer:

Because A has full row-rank and the reduced Hessian Z^TGZ is positive definite, the KKT matrix is nonsingular, so there is a vector (x^*, λ^*) satisfying.

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

假設(p, q)為向量

$$\Rightarrow \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = 0$$

因為$Ap = 0$，所以

$$0 = \begin{bmatrix} p \\ q \end{bmatrix}^T \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = p^TGp$$

因為p lies in the null space of A, 因此

$$p = Zu, u \in R^{n-m}$$

$$\Rightarrow 0 = p^TGp = u^TZ^TGZu$$

由於Z^TGZ是正定義，所以推出$u = 0, p = 0$。
因為$A^Tq = 0$，Full row rank A implies that $q = 0$
因此當$(p, q) = (0, 0)$時

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = 0 \quad is \quad satisfied$$

所以矩陣是nonsingular matrix。

3. (30%) Consider the problem

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

(a) Write down the KKT conditions for (2).

(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.

Answer:
(a)
$\mathcal{L}(\vec{x},\vec{\lambda}) = (x_1-3)^2 + 10x_2^2 - \lambda(-x_1^2 - x_2^2 + 1)$
KKT condition:
令$\vec{x^*} = (x_1^*, x_2^*) \quad \vec{\lambda^*} = \lambda^*$
$\Rightarrow$ 1. $\nabla\mathcal{L}(\vec{x^*},\vec{\lambda^*}) = \begin{bmatrix} 2\times(x_1^*-3) + 2\times\lambda^* x_1^* \\ 20x_2^* + \lambda^* x_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
$\Rightarrow$ 2. $x_1^{*2} - x_2^{*2} + 1 \geq 0$
$\Rightarrow$ 3. $\lambda^*(-x_1^{*2} - x_2^{*2} + 1) = 0$
$\Rightarrow$ 4. $\lambda^* \geq 0$
(b)
從(a)的第一行condition$\nabla\mathcal{L}(\vec{x^*},\vec{\lambda^*}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
解出
$\vec{x^*} = (x_1^*, x_2^*) = (1, 0)$
$\lambda^* = 2$
(c)
The critical cone $\mathcal{C}(\vec{x^*},\vec{\lambda^*}) = \{\vec{w} = \begin{bmatrix} 0 \\ w_2 \end{bmatrix} \mid \forall w_2 \in \mathbb{R}\}$
The Hessian of $\mathcal{L}(\vec{x^*},\vec{\lambda^*})$ is
$\Rightarrow \nabla^2\mathcal{L}(\vec{x^*},\vec{\lambda^*}) = \begin{bmatrix} 2+2\lambda^* & 0 \\ 0 & 20+\lambda^* \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 22 \end{bmatrix}$
The reduced Hessian is
$\Rightarrow \vec{w^T}\nabla^2\mathcal{L}(\vec{x^*},\vec{\lambda^*})\vec{w} = \begin{bmatrix} 0 & w_2 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 22 \end{bmatrix} \begin{bmatrix} 0 \\ w_2 \end{bmatrix} = 22w_2^2 \geq 0$
$\vec{w^T}\nabla^2\mathcal{L}(\vec{x^*},\vec{\lambda^*})\vec{w}$ is positive define
$\Rightarrow$ Therefore the second order conditions for the solution of this problem hold.

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?

Answer:
(a)
The optimal solution is $x^* = (-1, 0)^T$
the optimal Lagrangian multiplier $\lambda^* = -1.5$
(b)
By the formula of augmented Lagrangian penalty function $\mathcal{L}$ in slides:

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

derive the gradient of Lagrangian:
Hessian of $\mathcal{L}(\vec{x}, \rho, \mu)$:

$$\nabla^2 \mathcal{L}(\vec{x}, \rho, \mu) = \begin{bmatrix} 6x_1 - 2\rho + 2\mu(3x_1^2 + x_2^1 - 1) & 4\mu x_1 x_2 \\ 4\mu x_1 x_2 & 4 - 2\rho + 2\mu(x_1^2 + 3x_2^2 - 1) \end{bmatrix}$$

(c)
Let $\rho_0 = -1, \mu_0 = 1$,

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

The solution $\vec{x_1} = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix}$
(d)
$\rho_1 = \rho_0 - \mu_0 c(\vec{x_1})$,which is closer to λ^*

$$\rho_1 = -1 - 1 * \left[(-1.5)^2 + (0)^2 - 1\right] = -2.25$$

repeat the procedure until convergence.

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.

Answer:
The primal problem:

$$\begin{aligned} \min_{x_1, x_2} \quad & -3x_1 + x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 + x_3 = 20 \\ & x_1 + 2x_2 + x_4 = 16 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

The dual problem:

$$\begin{aligned} \max_{y_1,y_2} \quad & 20y_1 + 16y_2 \\ \text{s.t.} \quad & 2y_1 + y_2 + s_1 = -3 \\ & y_1 + 2y_2 + s_2 = 1 \\ & y_1 + s_3 = 0 \\ & y_2 + s_4 = 0 \\ & s_1, s_2, s_3, s_4 \geq 0 \end{aligned}$$

令

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}$$

令

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

derive the gradient of the Lagrangian :

$$F = \begin{bmatrix} r_c \\ r_b \\ r_s \end{bmatrix} = \begin{bmatrix} A^T y + s - c \\ Ax - b \\ Xs \end{bmatrix} = \left[\begin{array}{c} 2y_1 + y_2 + s_1 + 3 \\ y_1 + 2y_2 + s_2 - 1 \\ \hline y_1 + s_3 \\ y_2 + s_4 \\ 2x_1 + x_2 + x_3 - 20 \\ x_1 + 2x_2 + x_4 - 16 \\ \hline x_1 s_1 \\ x_2 s_2 \\ x_3 s_3 \\ x_4 s_4 \end{array}\right]$$

而the Jacobian:

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