CS532100 Numerical Optimization Homework 3

110062533 Yang, Jiun-Kai

Due Jan 12, 2022

1. (25%) Consider the problem

$$\begin{array}{ll} \min\limits_{x_1,x_2} & 0.1 \times (x_1-3)^2 + x_2^2 \\ \text{s.t.} & x_1^2 + x_2^2 - 1 \leq 0 \end{array} \tag{1}$$

   (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.

We can rewrite the problem as:

$$\begin{array}{ll} \min\limits_{\vec{x}} & f(\vec{x}) = 0.1 \times (x_1-3)^2 + x_2^2 \\ \text{s.t.} & c_1(\vec{x}) = -x_1^2 - x_2^2 + 1 \geq 0 \end{array}$$

The Lagrangian function is

$$\mathcal{L}(\vec{x}, \lambda) = f(\vec{x}) - \lambda c_1(\vec{x}) = 0.1 \times (x_1-3)^2 + x_2^2 - \lambda(-x_1^2 - x_2^2 + 1)$$

.

(a) The KKT conditions:
Suppose $\vec{x}^*$ is a solution to the problem and there exist a Lagrangian multiplier vector $\vec{\lambda}$ s.t. the following conditions are satisfied at $(\vec{x}^*, \vec{\lambda}^*) = (\hat{x}_1, \hat{x}_2, \hat{\lambda})$.

   i. $\nabla_{\vec{x}^*}\mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = \begin{bmatrix} 0.2 \times (\hat{x}_1 - 3) + 2\hat{\lambda}\hat{x}_1 \\ 2\hat{x}_2 + 2\hat{\lambda}\hat{x}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

   ii. $c_1(\vec{x}^*) = -\hat{x}_1^2 - \hat{x}_2^2 + 1 \geq 0$

   iii. $\hat{\lambda} c_1(\vec{x}^*) = \hat{\lambda}(-\hat{x}_1^2 - \hat{x}_2^2 + 1) = 0$

   iv. $\hat{\lambda} \geq 0$

(b) For the first and the third conditions, we can solve

$$\begin{cases} 0.1(\hat{x}_1 - 3) + \hat{\lambda}\hat{x}_1 & = & 0 \\ \hat{x}_2 + \hat{\lambda}\hat{x}_2 & = & 0 \end{cases}$$

and $\hat{\lambda}(-\hat{x}_1^2 - \hat{x}_2^2 + 1) = 0$,so we have these possible solutions: $(\hat{\lambda}, \hat{x}_1, \hat{x}_2) = (-1, -\frac{1}{3}, \frac{2\sqrt{2}}{3}), (0, 3, 0), (0.2, 1, 0),$ or $(-0.4, -1, 0)$. After checking the second and the forth conditions, there are only one solution left, $(\hat{\lambda}, \hat{x}_1, \hat{x}_2) = (0.2, 1, 0)$

(c) The critical cone $\mathcal{C}(\vec{x}^*, \vec{\lambda}^*)$ is a set of directions defined at the optimal solution $(\vec{x}^*, \vec{\lambda}^*)$. Since $\nabla c_1(\vec{x}^*) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$ and $\hat{\lambda} = 1 > 0$, the critical cone $\mathcal{C}(\vec{x}^*, \vec{\lambda}^*) = \{\begin{bmatrix} 0 \\ w_2 \end{bmatrix} | \forall w_2 \in \mathbb{R}\}$.

$$\nabla^2_{\vec{x}^*\vec{x}^*}\mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = \begin{bmatrix} 0.2 + 2\hat{\lambda} & 0 \\ 0 & 2 + 2\hat{\lambda} \end{bmatrix}_{\hat{\lambda}=0.2} = \begin{bmatrix} 0.6 & 0 \\ 0 & 2.4 \end{bmatrix}$$

The reduced Hessian is $\begin{bmatrix} 0 & w_2 \end{bmatrix} \begin{bmatrix} 0.6 & 0 \\ 0 & 2.4 \end{bmatrix} \begin{bmatrix} 0 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 & w_2 \end{bmatrix} \begin{bmatrix} 0 \\ 2.4w_2 \end{bmatrix} = 2.4w_2^2$.

Suppose $\vec{x}^*$ is a local minimizer at which the LICQ holds, and $\vec{\lambda}^*$ is the Lagrange multiplier. We can see that

$$\forall \vec{w} \in \mathcal{C}(\vec{x}^*, \vec{\lambda}^*), \vec{w}^T \nabla^2_{\vec{x}^*\vec{x}^*}\mathcal{L}(\vec{x}^*, \vec{\lambda}^*)\vec{w} = 2.4w_2^2 \geq 0,$$

so the second order conditions for the solution hold.

2. (20%) Consider the problem

$$\begin{array}{ll} \min\limits_{x_1,x_2,x_3} & (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 \\ \text{s.t.} & 0 \leq x_1, x_2, x_3 \leq 2. \end{array} \tag{2}$$

Find the optimization solution x^* for (2) with gradient projection method, with initial guess at $\vec{x}_0 = (x_1, x_2, x_3) = (0, 0, 2)^T$.
(Find out all segments and the minimizers of all segments, and determine whether the solution you got from this method is optimal solution. Justify your answer.)
In the quadratic model,

$$\begin{array}{ll} \min\limits_{\vec{x}} & f(\vec{x}) = \frac{1}{2}\vec{x}^T G\vec{x} + \vec{c}^T\vec{x} + d \\ \text{s.t.} & \vec{l} \leq \vec{x} \leq \vec{u} \text{ (which means } \forall i, l_i \leq x_i \leq u_i) \\ \text{where} & G = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \vec{c} = \begin{bmatrix} -2 \\ -2 \\ -2 \end{bmatrix}, d = 3, \vec{l} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{and}, \vec{u} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \end{array}$$

$$\nabla_{\vec{x}} f = G\vec{x} + \vec{c} = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 - 2 \\ 2x_3 - 2 \end{bmatrix}$$

We set $\vec{x}_0 = (0, 0, 2)^T$, so we get $\vec{g} = \nabla_{\vec{x}} f(\vec{x}_0) = (-2, -2, 2)$. Solve $\bar{t}_i = \begin{cases} (x_i - u_i)/g_i & \text{if } g_i < 0 \\ (x_i - l_i)/g_i & \text{if } g_i > 0 \\ \infty & \text{otherwise} \end{cases}$ and get $(\bar{t}_1, \bar{t}_2, \bar{t}_3) = (1, 1, 1)$. Since $\bar{t}_1 = \bar{t}_2 = \bar{t}_3 = 1$, we eliminate the duplicate values and zero values of $\bar{t}_i$ and examine the interval $[0, \bar{t}_1] = [0, 1]$. For $0 < t < 1$, the minimizer is at $\Delta t = 0.5$, $\vec{x}_1^* = (1, 1, 1)$. Since the smallest value among all segments is at $\Delta t = 0.5$, the exact solution is $\vec{x}_1^* = (1, 1, 1)$ and $f(\vec{x}_1^*) = 0$.

To justify the answer, we can see that $(x_i - 1)^2 \geq 0, \forall x_i \in \mathbb{R}$, so the smallest value we can get for each term is 0 and we should set $x_i = 1, \forall i \in 1, 2, 3$. Also, this setting do still match the constraints. The solution we get is correct!

3. (15%) Consider the problem

$$\begin{aligned} \min_{x_1, x_2} \quad & z = 8x_1 + 5x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \leq 100 \\ & 3x_1 + 4x_2 \leq 240 \\ & x_1, x_2 \geq 0 \end{aligned} \tag{3}$$

Formulate this problem to the equation of the interior point method, and derive the gradient of the Lagrangian and the Jacobian of the function F. (The "gradient" means F and the Jacobian is the derivative of F.)

Add slack variable to turn the inequality constraints to equality constraints.

$$\begin{aligned} \min_{\vec{x}} \quad & z = f(\vec{x}) \\ \text{s.t.} \quad & c_1(\vec{x}) - s_1 = 0 \\ & c_2(\vec{x}) - s_2 = 0 \\ & c_3(\vec{x}) - s_3 = 0 \\ & c_4(\vec{x}) - s_4 = 0 \\ & s_1, s_2, s_3, s_4 \geq 0 \\ \text{where} \quad & f(\vec{x}) = 8x_1 + 5x_2 \\ & c_1(\vec{x}) = -2x_1 - x_2 + 100 \\ & c_2(\vec{x}) = -3x_1 - 4x_2 + 240 \\ & c_3(\vec{x}) = x_1 \\ & c_4(\vec{x}) = x_2 \end{aligned}$$

The Lagrangian of the above problem is

$$\mathcal{L}(\vec{x}, \vec{s}, \vec{z}) = f(x) - \mu \sum_{i=1}^{4} \log(s_i) - \vec{y}^T \vec{0} - \vec{z}^T (C_I(\vec{x}) - \vec{s})$$

where $\vec{y}$ is the Lagrangian multiplier of equality constraints and $\vec{z} = [z_1, z_2, z_3, z_4]^T$ is the Lagrangian multiplier of inequality constraints.

Let A_I be the Jacobian of C_I, $A_I^T = \begin{bmatrix} -2 & -3 & 1 & 0 \\ -1 & -4 & 0 & 1 \end{bmatrix}$, $S = \begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & s_4 \end{bmatrix}$, and $\vec{e} = [1, 1, \cdots, 1]^T$. The gradient of the Lagrangian is

$$F = \begin{bmatrix} \nabla_x \mathcal{L} \\ \nabla_s \mathcal{L} \\ \nabla_y \mathcal{L} \\ \nabla_z \mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla f - A_I^T \vec{z} \\ -\mu S^{-1} \vec{e} + \vec{z} \\ \vec{0} \\ C_I(\vec{x}) - \vec{s} \end{bmatrix}$$

, but it is advantageous for Newton's method to transform the rational equation into a quadratic equation. We do so by multiplying the second equation $\nabla_s \mathcal{L}$ by S, a procedure that does not change the solution of F because the diagonal elements of S are positive. Therefore, we get the gradient of the Lagrangian is

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

where $\nabla f - A_I^T \vec{z} = \begin{bmatrix} 8 + 2z_1 + 3z_2 - z_3 \\ 5 + 1z_1 + 4z_2 - z_4 \end{bmatrix}$, $-\mu\vec{e} + S\vec{z} = \begin{bmatrix} -\mu + s_1 z_1 \\ -\mu + s_2 z_2 \\ -\mu + s_3 z_3 \\ -\mu + s_4 z_4 \end{bmatrix}$, and

$C_I(\vec{x}) - \vec{s} = \begin{bmatrix} -2x_1 - x_2 + 100 - s_1 \\ -3x_1 - 4x_2 + 240 - s_2 \\ x_1 - s_3 \\ x_2 - s_4 \end{bmatrix}$.

The Jacobian of the function F is

$$\nabla F = \begin{bmatrix} 0 & 0 & 0 & -A_I^T \\ 0 & Z & 0 & S \\ 0 & 0 & 0 & 0 \\ A_I & -I & 0 & 0 \end{bmatrix}$$

where $Z = \begin{bmatrix} z_1 & 0 & 0 & 0 \\ 0 & z_2 & 0 & 0 \\ 0 & 0 & z_3 & 0 \\ 0 & 0 & 0 & z_4 \end{bmatrix}$.

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

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 2 \end{aligned} \tag{4}$$

(a) Write the augmented Lagrangian penalty function L and Hessian of L of this problem.

(b) To make the augmented Lagrangian function L exact, what is the penalty parameter μ should be ?

Answers are put here.

(a) The augmented Lagrangian is

$$L_A(x_1, x_2) = x_1 + x_2 - \rho(x_1^2 + x_2^2 - 2) + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2$$

and the gradient of it is:

$$\nabla_x L_A = \begin{bmatrix} 1 - 2\rho x_1 + 2\mu(x_1^2 + x_2^2 - 2)x_1 \\ 1 - 2\rho x_2 + 2\mu(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix}$$

The Hessian of L_A is

$$\nabla_{xx} L_A = \begin{bmatrix} -2\rho + 2\mu(3x_1^2 + x_2^2 - 2) & 4\mu x_1 x_2 \\ 4\mu x_1 x_2 & -2\rho + 2\mu(x_1^2 + 3x_2^2 - 2) \end{bmatrix}$$

(b) It is trivial to see that the minimizer is $(x_1^*, x_2^*) = (-1, -1)$ and we get the minimum value -2. Given $(x_1^*, x_2^*) = (-1, -1)$, the gradient of the augmented Lagrangian should be $\vec{0}$, so we can solve

$$[\nabla_x L_A(x_1^*, x_2^*)]_{\rho=\rho^*} = \begin{bmatrix} 1 - 2\rho^* x_1 \\ 1 - 2\rho^* x_2 \end{bmatrix} = \begin{bmatrix} 1 + 2\rho^* \\ 1 + 2\rho^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

and get $\rho^* = -0.5$.
Put $(x_1^*, x_2^*) = (-1, -1)$ and $\rho^* = -0.5$ to $\nabla_{xx} L_A$, the Hessian of L_A becomes

$$\nabla_{xx} L_A = \begin{bmatrix} 4\mu + 1 & 4\mu \\ 4\mu & 4\mu + 1 \end{bmatrix}$$

Solve the $|\nabla_{xx} L_A| = 0$ and get the eigenvalues of it are 1 and $8\mu + 1$. Since it should still match the second-order condition with one equality constraint only, $8\mu + 1 > 0 \Rightarrow \mu > -\frac{1}{8}$.

5. (5%) Find the condition number $\kappa(A)$ of matrix A. Describe how ill- conditioned and good-conditioned matrices behave in matrix computation.

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 0 \end{bmatrix}$$

The condition number of A is $\kappa(A) = \|A\| \|A^-1\|$ and $A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 0 \end{bmatrix}$ and $A^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$, so we get

$$\kappa(A) = \|A\| \|A^{-1}\| = 4 \times \frac{1}{2} = 2$$

if we calculate the numerical value of the condition number of if by ∞-norm $\|\cdot\|_\infty$. The condition number $\kappa(A)$ means the sensitivity of the matrix when solving $Ax = b$.

For a matrix computation problem muliplying a matrix A and a vector $\vec{v}$, say $A\vec{x} = \vec{b}$ which $\vec{b}$ is a vector. Let E be a matrix that induces A into some error, and given another $\vec{x}'$ that we still get $\vec{b}$ by mulitplying $A + E$ and $\vec{x}'$:

$$
\begin{aligned}
&(A+E)\vec{x}' = A\vec{x} = b \\
\Rightarrow &\|\vec{x}' - \vec{x}\| = \|A^{-1}E\vec{x}'\| \leq \|A^{-1}\|\|E\|\|\vec{x}'\| \\
\Rightarrow &\frac{\|\vec{x}' - \vec{x}\|}{\|\vec{x}\|} \leq \|A\|\|A^{-1}\|\frac{\|E\|}{\|A\|} = \kappa(A)\frac{\|E\|}{\|A\|}
\end{aligned}
$$

Since the condition number of any matrix is greater or equal to 1, a well-conditioned matrix has a lower condition number that about 1 and an ill-conditioned matrix has a greater condition number. The difference results in different relative error $\frac{\|\vec{x}'-\vec{x}\|}{\|\vec{x}\|}$, and we can see that the relative error is lesser with well-conditioned matrix than the one with ill-conditioned matrix.

6. (15%) The problem 15.4 in textbook shows an example of Maratos effect.

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

The optimal solution is $\vec{x}^* = (1,0)$. Suppose $\vec{x}_k = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \vec{p}_k = \begin{pmatrix} \sin^2\theta \\ -\sin\theta\cos\theta \end{pmatrix}$ The Maratos effect says although $\vec{p}_k$ is a good step, but the filter method will reject it. In the textbook and slides, it says the remedy to this problem is using the second order correction. Read the textbook or the slides to understand the reason why the second order correction can help avoiding this problem. And explain it in your own words.

Denote he constraint $x_1^2 + x_2^2 - 1 = 0$ as $c(x_1,x_2) = x_1^2 + x_2^2 - 1 = 0$. We can set an iterate $\vec{x}_k$ as the form $\vec{x}_k = (\cos\theta, \sin\theta)^T$ Since the step is $\vec{p}_k = \begin{bmatrix} \sin^2\theta \\ -\sin\theta\cos\theta \end{bmatrix}$, the trial point is $\vec{x}_k + \vec{p}_k = \begin{bmatrix} \cos\theta + \sin^2\theta \\ \sin\theta(1-\cos\theta) \end{bmatrix}$.

Check the convergence and get

$$
\frac{\|\vec{x}_k + \vec{p}_k - x^*\|^2}{\|\vec{x}_k - x^*\|_2^2} = \frac{1}{2}.
$$

We can see that it is quadratic convergence, but if we put $\vec{x}_k + \vec{p}_k$ to the objective function $f(x_1,x_2)$ and the constraint $x_1 + x_2 = 1$, we get

$$
f(\vec{x}_k + \vec{p}_k) = \sin^2\theta - \cos\theta > -\cos\theta = f(\vec{x}_k)
$$

$$
c(\vec{x}_k + \vec{p}_k) = \sin^2\theta > c(\vec{x}_k) = 0
$$

when $\theta \neq 0$. Although $\vec{p}_k$ is a good step, the filter method will reject it.

Given a step p_k, the second-order correction step $\hat{p}_k$ is defined to be

$$
\hat{p}_k = -A_k^T(A_kA_k^T)^{-1}c(\vec{x}_k + \vec{p}_k),
$$

where $A_k = A(x_k)$ is the Jacobian of c at $\vec{x}_k$.
Why is that? Let take a deeper look at the constraint $c(\vec{x}_k + \vec{p}_k)$. By Taylor expansion, we get the second-order approximation of it is

$$c(\vec{x}_k + \vec{p}_k) \approx c(x_k) + \nabla c(\vec{x}_k)^T \vec{p}_k + \frac{1}{2}\vec{p}_k \nabla^2_{xx} c(\vec{x})\vec{p}_k = 0$$

Assume $\|\hat{p} - \vec{p}\| \approx 0$. The Taylor expansion above can be rearranged to get $\hat{p}$:

$$\frac{1}{2}\hat{p}_k \nabla^2_{xx} c(\vec{x})\hat{p}_k \approx \frac{1}{2}\vec{p}_k \nabla^2_{xx} c(\vec{x})\vec{p}_k \approx c(\vec{x}_k + \vec{p}_k) - c(\vec{x}_k) - \nabla c(\vec{x}_k)^T \vec{p}_k$$

Also, by Taylor expansion, we get the second-order approximation with $\hat{p}$

$$c(\vec{x}_k + \hat{p}_k) \approx c(x_k) + \nabla c(\vec{x}_k)^T \hat{p}_k + \frac{1}{2}\hat{p}_k \nabla^2_{xx} c(\vec{x})\hat{p}_k = 0$$

Therefore, since $\frac{1}{2}\hat{p}_k \nabla^2_{xx} c(\vec{x})\hat{p}_k \approx \frac{1}{2}\vec{p}_k \nabla^2_{xx} c(\vec{x})\vec{p}_k$ and $\nabla c(\vec{x}_k)^T \vec{p}_k = 0$ we have

$$\begin{aligned} c(\vec{x}_k + \hat{p}_k) &\approx c(x_k) + \nabla c(\vec{x}_k)^T \hat{p}_k + c(\vec{x}_k + \vec{p}_k) - c(\vec{x}_k) - \nabla c(\vec{x}_k)^T \vec{p}_k \\ &= \nabla c(\vec{x}_k)^T \hat{p}_k + c(\vec{x}_k + \vec{p}_k). \end{aligned}$$

Because $c(\vec{x}_k + \hat{p}_k) = 0$, rearrange the above equation and get

$$\nabla c(\vec{x}_k)^T \hat{p}_k + c(\vec{x}_k + \vec{p}_k) = 0$$

, so we finally get $\hat{p} = -A_k^T (A_k A_k^T)^{-1} c(\vec{x}_k + \vec{p}_k)$ where $A_k = \nabla c(\vec{x}_k)$.
With the second order correction, we estimate the step from $\vec{x}_k$ to $\vec{x}_k + \vec{p}_k + \hat{p}_k$ with additional evaluation of $c(\vec{x}_k + \vec{p}_k)$ to decreases the constraint violation.