CS532100 Numerical Optimization Homework 3

109062710

Due Jan 12

1. (25%) Consider the problem

$$\min_{x_1, x_2} \quad 0.1 \times (x_1 - 3)^2 + x_2^2 \tag{1}$$
$$\text{s.t.} \quad x_1^2 + x_2^2 - 1 \leq 0$$

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

Let's define the Lagrangian Form first. The Langrangian form is

$$L(x_1, x_2, \lambda) = 0.1(x_1 - 3)^2 + x_2^2 - \lambda(1 - x_1^2 - x_2^2)$$

The KKT conditions are

  i. $\nabla_{x_1, x_2} L(x_1^*, x_2^*, \lambda^*) = 0$
  ii. $(x_1^*)^2, (x_2^*)^2$
  iii. $\lambda^* \geq 0$
  iv. $\lambda^*(1 - (x_1^*)^2 - (x_2^*)^2)$

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

$$\nabla L = \begin{bmatrix} 0.2(x_1 - 3) + 2\lambda x_1 \\ 2x_2 + 2\lambda x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

From one of the conditions above, it violates the constraints if $\lambda = 0$, $x_1 = 3$, and $x_2 = 0$. So when $\lambda > 0$ and $x_1^2 + x_2^2 = 1$. The second equation in $\nabla L = 0$, which is $(1 + \lambda)x_2 = 0$, indicates either $x_2 = 0$ or $\lambda = -1$. Since $\lambda > 0$ and $x_2^* = 0$, therefore $x_1^* = 1$. Calculating the λ^* the first equation in $\nabla L = 0$, thus $0.2 \cdot 2(1 - 3) + 2\lambda = 0$.

(c) Compute the reduced Hessian and check the second order conditions for the solution. The Hessian of the Langrange is

$$\nabla^2 L = \begin{bmatrix} 0.2 + 2\lambda & 0 \\ 0 & 2 + 2\lambda \end{bmatrix} = \begin{bmatrix} 0.6 & 0 \\ 0 & 2.4 \end{bmatrix}$$

Let $A = (\nabla c)^*$ is at the optimal solution.

$$A = \begin{bmatrix} \frac{\partial c}{\partial x_1} & \frac{\partial c}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \end{bmatrix}$$

The null space is $\omega = \begin{bmatrix} 0 & \omega_2 \end{bmatrix}^T$.

The reduced Hessian is

$$\omega^T \nabla^2 L \omega = \begin{bmatrix} 0 & \omega_2 \end{bmatrix} \begin{bmatrix} 0.6 & 0 \\ 0 & 2.4 \end{bmatrix} \begin{bmatrix} 0 \\ \omega_2 \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} 0 & 2.4 \cdot \omega_2 \end{bmatrix} \begin{bmatrix} 0 \\ \omega_2 \end{bmatrix} \tag{3}$$

$$= 2.4 \cdot \omega_2^2 > 0 \tag{4}$$

2. (20%) Consider the problem

$$\begin{aligned} \min_{x_1, x_2, x_3} \quad & (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 \\ \text{s.t.} \quad & 0 \le x_1, x_2, x_3 \le 2. \end{aligned} \tag{5}$$

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.) The gradient at $\vec{x}_0 = (x_1, x_2, x_3) = (002)^T$ is

$$g = \begin{bmatrix} 2(x_1 - 1) \\ 2(x_2 - 1) \\ 2(x_3 - 1) \end{bmatrix} = \begin{bmatrix} -2 & -2 & 2 \end{bmatrix}$$

$$t_1 = -2/-2 = 1 \tag{6}$$
$$t_2 = -2/-2 = 1 \tag{7}$$
$$t_3 = -2/2 = -1 \tag{8}$$

We then need to process the segments sequentially such that $(0, t_3) \rightarrow (t_3, t_2) \rightarrow (t_2, t_1)$.

$$\vec{x}(t) = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \rightarrow \begin{bmatrix} 2t \\ 2t \\ 2t \end{bmatrix} \rightarrow \begin{bmatrix} -2 \\ 2t \\ 2t \end{bmatrix}_{t=-1} \rightarrow \begin{bmatrix} -2 \\ 2 \\ 2t \end{bmatrix}_{t=1} \rightarrow \begin{bmatrix} -2 \\ 2 \\ 2 \end{bmatrix}_{t=1}$$

In the first segment, we have

$$\min_t (2t - 1)^2 + (2t - 1)^2 + (2t - 1)^2 = \min_t 3(2t - 1)^2$$

The optimal solution is at $t = \frac{1}{2}$, which is larger than $t = -1$, so we set $t = -1$. With $t = -1$, then $f(x(t)) = 27$.

In the second segment

$$\min_t (-2 - 1)^2 + (2t - 1)^2 + (2t - 1)^2 = \min_t 2(2t - 1)^2 + 9$$

Its optimal solution is at $t = \frac{1}{2}$, then $f(x(t)) = 9$.

2

Thus, the optimal solution is at $(-2, 2, 2)$. Plugging $(-2, 2, 2)$ to the original formula, we then have

$$(-2 - 1)^2 + (2 - 1)^2 + (2 - 1)^2 = 9 + 1 + 1 = 11$$

3. (15%) Consider the problem

$$\begin{aligned}
\min_{x_1, x_2} \quad & z = 8x_1 + 5x_2 \\
\text{s.t.} \quad & 2x_1 + x_2 \leq 100 \\
& 3x_1 + 4x_2 \leq 240 \\
& x_1, x_2 \geq 0
\end{aligned} \tag{9}$$

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.)

First, the primal and the dual problems should be defined.

The primal problem is

$$\begin{aligned}
\min_{x_1, x_2} \quad & 8x_1 + 5x_2 \\
\text{s.t.} \quad & 2x_1 + x_2 + x_3 = 100 \\
& 3x_1 + 4x_2 + x_4 = 240 \\
& x_1, x_2, x_3, x_4 \geq 0
\end{aligned}$$

The dual problem is

$$\begin{aligned}
\max_{y_1, y_2} \quad & 100x_1 + 240x_2 \\
\text{s.t.} \quad & 2y_1 + 3y_2 + s_1 = 8 \\
& y_1 + 4y_2 = 5 \\
& s_1, s_2, s_3, s_4 \geq 0
\end{aligned}$$

$$A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 8 \\ 5 \\ 0 \\ 0 \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} s = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix} b = \begin{bmatrix} 100 \\ 240 \end{bmatrix}$$

The gradient is

$$F = \begin{bmatrix} A^T y + s - c \\ Ax - b \\ X \cdot s \end{bmatrix} = \begin{bmatrix} 2y_1 + 3y_2 + s_1 - 8 \\ y_1 + 4y_2 + s_2 - 5 \\ y_1 + s_3 \\ y_2 + s_4 \\ 2x_1 + x_2 + x_3 - 100 \\ 3x_1 + 4x_2 + x_4 - 240 \\ x_1 s_1 \\ x_2 s_2 \\ x_3 s_3 \\ x_3 s_4 \end{bmatrix} = 0$$

3

The Jacobian matrix is

$$J = \begin{bmatrix} \nabla_x r_c & \nabla_y r_c & \nabla_s r_c \\ \nabla_x r_b & \nabla_y r_b & \nabla_s r_b \\ \nabla_x r_s & \nabla_y r_s & \nabla_s r_s \end{bmatrix} \tag{10}$$

$$= \begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \tag{11}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 2 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 4 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_1 & 0 & 0 & 0 & 0 & 0 & x_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 & 0 & 0 & x_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 & 0 & 0 & 0 & 0 & x_3 & 0 \\ 0 & 0 & 0 & s_4 & 0 & 0 & 0 & 0 & 0 & x_4 \end{bmatrix} \tag{12}$$

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

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 2 \end{aligned} \tag{13}$$

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

The augmented Lagrangian penalty function is

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

The gradient of L is

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

The Hessian of L is

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

(b) To make the augmented Lagrangian function L exact, what is the penalty parameter μ should be ? First we should find the value of x_1, x_2, and λ which will result the smallest value.

Let's define the Langrangian form first.

$$\nabla L(x_1, x_2, \lambda) = \begin{bmatrix} 1 - 2\lambda x_1 \\ 1 - 2\lambda x_2 \\ x_1^2 + x_2^2 - 2 \end{bmatrix} = 0$$

4

Solving the equations in the matrix, we then have $\lambda = 1/2$ and $\lambda = -1/2$. When $\lambda = 1/2$, the $x_1 = 1$ and $x_2 = 1$. When $\lambda = -1/2$, the $x_1 = -1$ and $x_2 = -1$. Since $\lambda = -1/2$, $x_1 = -1$ and $x_2 = -1$ configurations give us the lowest value if those values are plugged in $x_1 + x_2$, we then use these values to find the minimum μ.

At $x_1 = -1$ and $x_2 = -1$, the Hessian matrix becomes

$$\begin{bmatrix} 4\mu & 4\mu \\ 4\mu & 4\mu \end{bmatrix}$$

To make sure all the values in the Hessian matrix above are positive, $\mu > 0$.

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

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

In order to calculate the condition number of matrix A, we need to know that $cond(A) = ||A|| \cdot ||A^{-1}||$.

However, we should know that $||A||$ the n-norm of matrix A. Since the norm-2 is the most common one, we are going to use $||A||_2$ and $||A^{-1}||_2$

The inverse of matrix A is

$$A^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 0 & 1/3 \\ 0 & 1/2 & 0 \end{bmatrix}$$

Calculating both $||A||_2$ and $||A||_2^{-1}$, we get $||A||_2 = 4$ and $||A^{-1}||_2 = 0.5$. Thus, $cond_2(A) = ||A||_2 \cdot ||A^{-1}||_2 = 2$.

Other results:

(a) Norm-1 $\rightarrow cond_1(A) = 2$

(b) Norm-∞ $\rightarrow cond_\infty(A) = 2$

(c) Frobenius norm $\rightarrow cond_{fro}(A) = 3.505$

(d) Max norm $\rightarrow cond_{\max}(A) = 2$

Ill-conditioned number is a condition where a small change in the input value will result in a large change in the output value.

As oppose to ill-conditioned number, well-conditioned number is a condition where a little or large change to the input in a function will result in a little change to the output.

Therefore, a condition number of 0 would mean that the matrix removes any error, which isn't possible. However

In fact, the smallest possible condition number is 1, where an error is neither magnified nor diminished.

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

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

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.

Suppose

$$\vec{x}_k = \begin{bmatrix} cos\theta \\ sin\theta \end{bmatrix} \tag{14}$$

$$\vec{p}_k = \begin{bmatrix} sin^2\theta \\ -sin\theta cos\theta \end{bmatrix} \tag{15}$$

$$\vec{x}_{k+1} = \vec{x}_k + \vec{p}_k \tag{16}$$

$$= \begin{bmatrix} cos\theta + sin^2\theta \\ sin\theta(1 - cos\theta) \end{bmatrix} \tag{17}$$

We then calculate $||\vec{x}_k - \vec{x}^*||$ and $||\vec{x}_{k+1} - \vec{x}^*||$.

$$||\vec{x}_k - \vec{x}^*|| = \begin{bmatrix} \cos\theta - 1 \\ \sin\theta \end{bmatrix} \tag{18}$$

$$= \sqrt{\cos^2\theta - 2\cos\theta + 1 + \sin^2\theta} \tag{19}$$

$$= \sqrt{2(1 - \cos\theta)} \tag{20}$$

$$||\vec{x}_{k+1} - \vec{x}^*|| = \begin{bmatrix} \cos\theta + \sin^2\theta - 1 \\ \sin\theta - \sin\theta\cos\theta \end{bmatrix} \tag{21}$$

$$= \begin{bmatrix} \cos\theta(1 - \cos\theta) \\ \sin\theta(1 - \cos\theta) \end{bmatrix} \tag{22}$$

$$= \sqrt{\cos^2\theta(1 - \cos\theta)^2 + \sin^2\theta(1 - \cos\theta)^2} \tag{23}$$

$$= \sqrt{(1 - \cos\theta)^2} \tag{24}$$

Thus,

$$\frac{||\vec{x}_{k+1} - \vec{x}^*||}{||\vec{x}_k - \vec{x}^*||} = \frac{1}{2}$$

Unfortunately, the filter method is going to reject this step because

6

$$f(\vec{x}_k) = -\cos\theta \text{ , and } c(\vec{x}_k) = 0 \tag{25}$$

$$f(\vec{x}_{k+1}) = -\cos\theta - \sin 2\theta = \sin^2\theta - \cos\theta > f(\vec{x}_k) \tag{26}$$

$$c(\vec{x}_{k+1}) = \sin^2\theta > 0 = (\vec{x}_k) \tag{27}$$

This is what is called the Maratos Effect.

The second order correction is believed to be able to solve The Maratos Effect. Instead of stating that $\nabla c\left(\vec{x}_k\right)^T \vec{p}_k + c\left(\vec{x}_k\right) = 0$, let's express it in a quadratic approximation form.

$$c\left(\vec{x}_k\right) + \nabla c\left(\vec{x}_k\right)^T \vec{d}_k + \frac{1}{2}\vec{d}_k^T \nabla_{xx}^2 c(\vec{x})\vec{d}_k = 0$$

Suppose $||\vec{d}_k - \vec{p}_k||$ is very small. Let's use Taylor expansion in order to approximate the quadratic term.

$$c\left(\vec{x}_k + \vec{p}_k\right) \approx c\left(\vec{x}_k\right) + \nabla c\left(\vec{x}_k\right)^T \vec{p}_k + \frac{1}{2}\vec{p}_k^T \nabla_{xx}^2 c(\vec{x})\vec{p}_k \tag{28}$$

$$\frac{1}{2}\vec{d}_k^T \nabla_{xx}^2 c(\vec{x})\vec{d}_k \approx \frac{1}{2}\vec{p}_k^T \nabla_{xx}^2 c(\vec{x})\vec{p}_k \tag{29}$$

$$\approx c\left(\vec{x}_k + \vec{p}_k\right) - c\left(\vec{x}_k\right) - \nabla c\left(\vec{x}_k\right)^T \vec{p}_k \tag{30}$$

Using the (30) can be simplified to

$$\nabla c\left(\vec{x}_k\right)^T \vec{d}_k + c\left(\vec{x}_k + \vec{p}_k\right) - \nabla c\left(\vec{x}_k\right)^T \vec{p}_k = 0$$

Express the equation above with the corrected linearized constraint

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