CS532100 Numerical Optimization Homework 3

106033233 Sheng Yen Chou

Due Jan 12, 2022

1. (25%) Consider the problem

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

   (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.

Answers are put here.

(a)

Consider a constrained optimization problem in standard format is

$$\min_{\vec{x}} f(\vec{x})$$
$$\text{subject to} \quad \begin{aligned} & c_i(\vec{x}) = 0, i \in \mathcal{E} \\ & c_i(\vec{x}) \geqslant 0, i \in \mathcal{I} \end{aligned}$$

where $\mathcal{E}$ and $\mathcal{I}$ are index sets of the equality conditions and the inequality conditions with size $|\mathcal{E}| = p$ and $|\mathcal{I}| = q$.

We can write down the Lagrangian function $\mathcal{L}$ of the constrained optimization problem as

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \vec{\lambda}^\top c(\vec{x}) = f(\vec{x}) - \sum_{i=0}^{N} c_i(\vec{x})$$

where $N = p + q$, $\vec{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_N\}^\top \in \mathbb{R}^N$ is the Lagrangian multiplier and $c(\vec{x}) = \{c_1(\vec{x}), c_2(\vec{x}), ..., c_N(\vec{x})\}^\top \in \mathbb{R}^N$

The KKT conditions claims that for the optimal solution $\vec{x}^*$ of the constrained optimization problem, the following formulas are satisfied.

(a) Stationary condition: $\nabla_{\vec{x}^*} \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = 0$

(b) Primal feasibility condition: $\begin{aligned} & c_i(\vec{x}) = 0 \quad \forall i \in \mathcal{E} \\ & c_i(\vec{x}) \geqslant 0 \quad \forall i \in \mathcal{I} \end{aligned}$

(c) Complementary slackness: $\lambda_i^* c_i(\vec{x}) = 0$

(d) Dual feasibility: $\lambda_i^* \geqslant 0 \quad \forall i \in \mathcal{I}$

First, we need to convert the optimization problem into standard form

$$\begin{array}{ll} \min_{x_1,x_2} & 0.1 \times (x_1 - 3)^2 + x_2^2 \\ \text{subject to} & -x_1^2 - x_2^2 + 1 \geqslant 0 \end{array}$$

Let $f(x_1, x_2) = 0.1 \times (x_1 - 3)^2 + x_2^2$ and $c(x_1, x_2) = -x_1^2 - x_2^2 + 1$

We can write down the Lagrangian function for the given optimization problem

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

Thus, the KKT conditions of the equation (1) is

(a) $\nabla_x^* \mathcal{L}(x_1^*, x_2^*, \lambda^*) = 0$

(b) $-(x_1^*)^2 - (x_2^*)^2 + 1 \geqslant 0$

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

(d) $\lambda^* \geqslant 0$

(b)

First, consider the stationary condition: $\nabla_x^* \mathcal{L}(x_1^*, x_2^*, \lambda^*) = 0$

$$\frac{\partial \mathcal{L}(x_1^*, x_2^*, \lambda^*)}{\partial x_1^*} = 0.2 \times (x_1^* - 3) + 2\lambda^* x_1^* = (0.2 + 2\lambda^*)x_1^* - 0.6 = 0$$

$$\frac{\partial \mathcal{L}(x_1^*, x_2^*, \lambda^*)}{\partial x_2^*} = 2x_2^* + 2\lambda^* x_2^* = (2 + 2\lambda^*)x_2^* = 0$$

Thus, we can derive simultaneous equations

$$\begin{cases} (0.2 + 2\lambda^*)x_1^* - 0.6 & = & 0 \\ (2 + 2\lambda^*)x_2^* & = & 0 \end{cases}$$

Next, consider the simultaneous equations of the stationary condition and the complementary slackness($\lambda^* \geqslant 0$), we can derive

$$\begin{cases} x_1^* & = & \frac{0.6}{0.2+2\lambda^*} \\ x_2^* & = & \frac{0}{(2+2\lambda^*)} \end{cases}$$

Plug into the complementary slackness: $\lambda^*(-(x_1^*)^2 - (x_2^*)^2 + 1) = 0$ and we get

$$\lambda^*(-(\frac{0.6}{0.2 + 2\lambda^*})^2 - (\frac{0}{(2 + 2\lambda^*)})^2 + 1) = 0$$

$$\lambda^*(-(\frac{0.6}{0.2+2\lambda^*})^2+1)=0$$

$$-\frac{0.36\lambda^*}{0.04+4(\lambda^*)^2+0.8\lambda^*}+\lambda^*=0$$

$$0.36\lambda^*=\lambda^*(0.04+4(\lambda^*)^2+0.8\lambda^*)$$

$$0.36\lambda^*=0.04\lambda^*+4(\lambda^*)^3+0.8(\lambda^*)^2$$

$$-0.32\lambda^*+4(\lambda^*)^3+0.8(\lambda^*)^2=0$$

$$\frac{4}{25}\lambda^*(-2+5\lambda^*+25(\lambda^*)^2)=0$$

$$\frac{4}{25}\lambda^*(5\lambda^*+2)(5\lambda^*-1)=0$$

Finally, we know the solution of the variables x_1^*, x_2^*, and λ^* as the following

$$\begin{cases}\lambda^* &= & & 0,-\frac{2}{5},\frac{1}{5} \\ x_1^* &= & \frac{0.6}{0.2+2\lambda^*} = & \frac{0.6}{0.2},\frac{0.6}{0.2-0.8},\frac{0.6}{0.2+0.4}=3,-1,1 \\ x_2^* &= & & \frac{0}{(2+2\lambda^*)}=0\end{cases}$$

Thus, all possible solutions are

$$(\lambda^*,x_1^*,x_2^*)=(0,3,0),(-\frac{2}{5},-1,0),(\frac{1}{5},1,0)$$

Consider KKT conditions, we need to verify the solution with KKT conditions.

(a) $(0,3,0)$: Consider the primal feasibility, $-(x_1^*)^2-(x_2^*)^2+1=-1-9+1=-9\not\geq 0$, violate the constraint.

(b) $(-\frac{2}{5},-1,0)$: Consider the dual feasibility, $\lambda^*=-\frac{5}{2}\not\geq 0$, violate the constraint.

(c) $(\frac{1}{5},1,0)$ satisfy all constraints.

Thus, the solution is

$$(\lambda^*,x_1^*,x_2^*)=(\frac{1}{5},1,0)$$

(c)

$$\nabla_{\vec{x}*}\mathcal{L}(x_1^*, x_2^*, \lambda^*) = \begin{pmatrix} \frac{\partial \mathcal{L}(x_1^*, x_2^*, \lambda^*)}{\partial x_1^*} \\ \frac{\partial \mathcal{L}(x_1^*, x_2^*, \lambda^*)}{\partial x_2^*} \end{pmatrix}$$

$$= \begin{pmatrix} (0.2 + 2\lambda^*)x_1^* - 0.6 \\ (2 + 2\lambda^*)x_2^* \end{pmatrix}$$

The Hessian matrix

$$\nabla^2_{\vec{x}*}\mathcal{L}(x_1^*, x_2^*, \lambda^*) = \begin{pmatrix} 0.2 + 2\lambda^* & 0 \\ 0 & 2 + 2\lambda^* \end{pmatrix}$$

Compute the eigenvalues λ_{eigen} of the Hessian matrix

$$\det(\begin{pmatrix} 0.2 + 2\lambda^* - \lambda_{eigen} & 0 \\ 0 & 2 + 2\lambda^* - \lambda_{eigen} \end{pmatrix} = 0)$$

$$((0.2 + 2\lambda^*) - \lambda_{eigen})((2 + 2\lambda^*) - \lambda_{eigen}) = 0$$

$$\lambda_{eigen} = 0.2 + 2\lambda^*, 2 + 2\lambda^*$$

$$\lambda^* = \frac{1}{5} \to \lambda_{eigen} = 0.4, 2.2 \geqslant 0$$

Because the Hessian matrix is a symmetric and real matrix, which is also a Hermitian matrix. According to Rayleigh's quotient, $\vec{w}^\top \nabla^2_{\vec{x}*}\mathcal{L}(x_1^*, x_2^*, \lambda^*)\vec{w}$ can be bounded as

$$\lambda_{min} \leqslant \vec{w}^\top \nabla^2_{\vec{x}*}\mathcal{L}(x_1^*, x_2^*, \lambda^*)\vec{w} \leqslant \lambda_{max}$$

Where λ_{min} and λ_{max} are the smallest and the largest eigenvalue of the Hessian matrix $\nabla^2_{\vec{x}*}\mathcal{L}(x_1^*, x_2^*, \lambda^*)$.

According to the second-order condition, the minimal eigenvalue λ_{min} of the Hessian matrix should $\geqslant 0$. Thus, the **Hessian matrix** $\nabla^2_{\vec{x}*}\mathcal{L}(x_1^*, x_2^*, \lambda^*)$ **satisfies the second-order condition**. Also, the valid solutions are

$$\begin{cases} \lambda^* & = & \frac{1}{5} \\ x_1^* & = & 1 \\ x_2^* & = & 0 \end{cases}$$

We can get

$$(\lambda^*, x_1^*, x_2^*) = (\frac{1}{5}, 1, 0)$$

And we know the second-order condition hold.

2. (20%) Consider the problem

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

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.)
Answers are put here.

First, we denote the quadratic bound-constrained problem as

$$\min_{\vec{x}} q(\vec{x}) = \frac{1}{2}\vec{x}^\top \mathbf{G}\vec{x} + \vec{x}^\top \vec{c} \quad \text{subject to} \quad \vec{l} \leqslant \vec{x} \leqslant \vec{u}$$

Consider the given target function, the matrix $\mathbf{G}$ and $\vec{c}$ is

$$\mathbf{G} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \quad \vec{c} = \begin{pmatrix} -2 \\ -2 \\ -2 \end{pmatrix}$$

Denote the gradient of the objective function as $\vec{g}(\vec{x}) = \vec{g}(x_0, x_1, x_2)$

$$\vec{g}(\vec{x}) = \vec{g}(x_0, x_1, x_2) = \nabla_{\vec{x}}(x_0-1)^2 + (x_1-1)^2 + (x_2-1)^2 = \begin{pmatrix} 2(x_0-1) \\ 2(x_1-1) \\ 2(x_2-1) \end{pmatrix}$$

$$\bar{t}_i = \begin{cases} (x_i - u_i)/g_i & \text{if } g_i < 0 \text{ and } u_i < +\infty \\ (x_i - l_i)/g_i & \text{if } g_i > 0 \text{ and } l_i > -\infty \\ \infty & \text{otherwise} \end{cases}$$

Iteration 1

Let $\vec{x}_0 = (0, 0, 2)$

$$\vec{g}(\vec{x}_0) = \begin{pmatrix} -2 \\ -2 \\ 2 \end{pmatrix}$$

$$\begin{cases} \bar{t}_0 = \frac{(0-2)}{-2} = 1 \\ \bar{t}_1 = \frac{(0-2)}{-2} = 1 \\ \bar{t}_2 = \frac{(2-0)}{2} = 1 \end{cases}$$

Thus, we find out the segments $(0, 1)$. We need to find out an optimal step length t such that $\vec{x}(t_j) = \vec{x}(t_{j-1}) + (\Delta t)p^{j-1}, \quad \Delta t = t - t_{j-1} \in [0, t_j - t_{j-1}]$ achieves minimal value.

As we reorder the segments, we can get the ordered segments as $\{[0, 1]\}$.

$$\begin{cases} t_0 = 0 \\ t_1 = 1 \end{cases}$$

We can derive the search direction for j-1-th iteration p^{j-1}. Note that we denote i-th entry of the search direction as p_i^{j-1} and i-th entry of the gradient as $g_i(\vec{x}(t_j))$.

$$p_i^{j-1} = \begin{cases} -g_i(\vec{x}(t_{j-1})) & \text{if} \quad t_{j-1} < \bar{t}_i \\ 0 & otherwise \end{cases}$$

Thus, we can get the search direction in iteration 1

$$p^0 = \begin{pmatrix} -g_0(\vec{x}(t_0)) \\ -g_1(\vec{x}(t_0)) \\ -g_2(\vec{x}(t_0)) \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -2 \end{pmatrix}$$

Then, for each segment, we need to compute the optimal step length Δt^*, such that $q(\vec{x}(t_j))$ reaches the minimal value.

$$q(\vec{x}(t_j)) = \frac{1}{2}(\vec{x}(t_{j-1})+(\Delta t)p^{j-1})^\top \mathbf{G}(\vec{x}(t_{j-1})+(\Delta t)p^{j-1})+(\vec{x}(t_{j-1})+(\Delta t)p^{j-1})^\top \vec{c}$$

$$= \tfrac{1}{2}(\vec{x}(t_{j-1})^\top \mathbf{G}\vec{x}(t_{j-1}) + ((\Delta t)p^{j-1})^\top \mathbf{G}((\Delta t)p^{j-1}) + 2\vec{x}(t_{j-1})^\top \mathbf{G}(\Delta t)p^{j-1})) \\ +\vec{x}(t_{j-1})^\top \vec{c} + ((\Delta t)p^{j-1})^\top \vec{c}$$

$$= \tfrac{1}{2}\vec{x}(t_{j-1})^\top \mathbf{G}\vec{x}(t_{j-1}) + \vec{x}(t_{j-1})^\top \vec{c} \\ +\tfrac{1}{2}(p^{j-1})^\top \mathbf{G}(p^{j-1})(\Delta t)^2 \\ +\vec{x}(t_{j-1})^\top \mathbf{G}p^{j-1}(\Delta t) + (p^{j-1})^\top \vec{c}(\Delta t)$$

$$= (\tfrac{1}{2}(p^{j-1})^\top \mathbf{G}(p^{j-1}))(\Delta t)^2 \\ +(\vec{x}(t_{j-1})^\top \mathbf{G}p^{j-1} + (p^{j-1})^\top \vec{c})(\Delta t) \\ +\tfrac{1}{2}\vec{x}(t_{j-1})^\top \mathbf{G}\vec{x}(t_{j-1}) + \vec{x}(t_{j-1})^\top \vec{c}$$

To simplify the equation, we denote $a(t_{j-1}) = (p^{j-1})^\top \mathbf{G}(p^{j-1})$, $b(t_{j-1}) = \vec{x}(t_{j-1})^\top \mathbf{G}p^{j-1} + (p^{j-1})^\top \vec{c}$, and $c(t_{j-1}) = \frac{1}{2}\vec{x}(t_{j-1})^\top \mathbf{G}\vec{x}(t_{j-1}) + \vec{x}(t_{j-1})^\top \vec{c}$. Thus, we can get

$$= \frac{1}{2}a(t_{j-1})(\Delta t)^2 + b(t_{j-1})(\Delta t) + c(t_{j-1})$$

To achieve the minimal value of $q(\vec{x}(t_j)$, we need to find the solution Δt^* of $\nabla_{\Delta t*} q(\vec{x}(t_j)) = 0$.

$$\nabla_{\Delta t*} q(\vec{x}(t_j)) = a(t_{j-1})(\Delta t^*) + b(t_{j-1}) = 0$$

$$\Delta t^* = \frac{-b(t_{j-1})}{a(t_{j-1})} = -\frac{\vec{x}(t_{j-1})^\top \mathbf{G}p^{j-1} + (p^{j-1})^\top \vec{c}}{(p^{j-1})^\top \mathbf{G}(p^{j-1})}$$

Thus, we can get Δt^* in iteration 1

$$
\begin{aligned}
\Delta t^* &= -\frac{\vec{x}(t_0)^\top \mathbf{G} p^0 + (p^0)^\top \vec{c}}{(p^0)^\top \mathbf{G}(p^0)} \\
&= -\frac{\begin{pmatrix}0 & 0 & 2\end{pmatrix}\begin{pmatrix}\frac{1}{2} & 0 & 0\\ 0 & \frac{1}{2} & 0\\ 0 & 0 & \frac{1}{2}\end{pmatrix}\begin{pmatrix}2\\2\\-2\end{pmatrix} + \begin{pmatrix}2 & 2 & -2\end{pmatrix}\begin{pmatrix}-2\\-2\\-2\end{pmatrix}}{\begin{pmatrix}2 & 2 & -2\end{pmatrix}\begin{pmatrix}\frac{1}{2} & 0 & 0\\ 0 & \frac{1}{2} & 0\\ 0 & 0 & \frac{1}{2}\end{pmatrix}\begin{pmatrix}2\\2\\-2\end{pmatrix}} \\
&= -\frac{(-2)+(-4)}{12} = \frac{6}{12} = \frac{1}{2}
\end{aligned}
$$

Since $\Delta t^* = \frac{1}{2} \in [0, t_j - t_{j-1}] = [0,1]$, the optimal step length is $t = t_{j-1} + \Delta t^* = 0 + \frac{1}{2} = \frac{1}{2}$. The optimal solution is

$$
\vec{x}(t_1) = \vec{x}(t_0) + \Delta t^* p^0 = \begin{pmatrix}0\\0\\2\end{pmatrix} + \frac{1}{2}\begin{pmatrix}2\\2\\-2\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}
$$

Thus, we get the optimal solution $\vec{x}(t_1)$:

$$
\vec{x}(t_1) = (1,1,1)
$$

To verify the solution, we can simply plug the solution $\vec{x}(t_1)$ into the target function $q(\vec{x}(t_1))$ and we can get $q(\vec{x}(t_1)) = 0$, which is the minimal value.

3. (15%) Consider the problem

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

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.)

Answers are put here.

For interior point method, we consider the optimization problem as

$$
\begin{aligned}
\min_{\vec{x}} \quad & f(\vec{x}) \\
\text{subject to} \quad & C_E(\vec{x}) = 0 \\
& C_I(\vec{x}) - \vec{s} = 0 \\
& \vec{s} \geqslant 0
\end{aligned}
$$

Where $C_E(\vec{x}) \in \mathbb{R}^{N_E}$ are equality constrains, $C_I(\vec{x}) \in \mathbb{R}^{N_I}$ are inequality constrain, and $\{s_1, s_2, ..., s_{N_I}\}\vec{s} \in \mathbb{R}^{N_I}$ are slack variables

The interior point method(IPM) would build a barrier to prevent the slack variables to become 0. Thus, we can reformulate the optimization problem as

$$\begin{array}{rl} \min_{\vec{x},\vec{s}} & f(\vec{x}) - \mu \sum_{i=1}^{N_I} \log(s_i) \\ \text{subject to} & C_E(\vec{x}) = 0 \\ & C_I(\vec{x}) - \vec{s} = 0 \\ & \vec{s} \geqslant 0 \end{array}$$

In order to solve the optimization problem, we can write down the Lagrangian function as

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

Apply the KKT conditions

(a) Stationary condition: $\begin{array}{l} \nabla_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{s}^*, \vec{y}^*, \vec{z}^*) = 0 \\ \nabla_{\vec{s}*} \mathcal{L}(\vec{x}^*, \vec{s}^*, \vec{y}^*, \vec{z}^*) = 0 \end{array}$

(b) Primal feasibility condition: $\begin{array}{l} C_E(\vec{x}^*) = 0 \\ C_I(\vec{x}^*) \geqslant 0 \end{array}$

(c) Complementary slackness: $\vec{z}^* \odot C_I(\vec{x}^*) = 0$

(d) Dual feasibility: $\vec{z}^* \geqslant 0$

Thus, we can derive the Jacobian matrix of KKT conditions

$$J(\vec{x}^*, \vec{s}^*, \vec{y}^*, \vec{z}^*) = \begin{pmatrix} \nabla_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{s}^*, \vec{y}^*, \vec{z}^*) \\ \nabla_{\vec{s}*} \mathcal{L}(\vec{x}^*, \vec{s}^*, \vec{y}^*, \vec{z}^*) \\ C_E(\vec{x}^*) \\ C_I(\vec{x}^*) - \vec{s}^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Expand the matrix

$$J(\vec{x}, \vec{s}, \vec{y}, \vec{z}) = \begin{pmatrix} \nabla_{\vec{x}} f(\vec{x}) - \vec{y}^\top \nabla_{\vec{x}} C_E(\vec{x}) - \vec{z}^\top \nabla_{\vec{x}} C_I(\vec{x}) \\ \frac{\mu}{\vec{s}} - \vec{z} \\ C_E(\vec{x}) \\ C_I(\vec{x}) - \vec{s} \end{pmatrix}$$

$$J(\vec{x}, \vec{s}, \vec{y}, \vec{z}) = \begin{pmatrix} \nabla_{\vec{x}} f(\vec{x}) - \vec{y}^\top \nabla_{\vec{x}} C_E(\vec{x}) - \vec{z}^\top \nabla_{\vec{x}} C_I(\vec{x}) \\ \mu \vec{e} - \vec{z} \odot \vec{s} \\ C_E(\vec{x}) \\ C_I(\vec{x}) - \vec{s} \end{pmatrix}$$

Where $\vec{e} = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}$ and $\odot$ is element-wise product.

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 method can only solve the optimization problem with equality constrain. Thus, we can write down the standard form for the optimization problem as

$$\min_{\vec{x}} f(\vec{x}) \quad \text{subject to} \quad \vec{c}(\vec{x}) = 0$$

Where $f(\vec{x}) : \mathbb{R}^N \to \mathbb{R}$ is the objective function and $\vec{c}(\vec{x}) : \mathbb{R}^N \to \mathbb{R}^M$ are the equality constrains. Next, we can derive the augmented Lagrangian function as

$$\mathcal{L}(\vec{x}, \vec{\rho}, \mu) = f(\vec{x}) - \vec{\rho}^\top \vec{c}(\vec{x}) + \frac{\mu}{2} \vec{c}(\vec{x})^\top \vec{c}(\vec{x})$$

$$= f(\vec{x}) - \sum_{i=1}^{M} \rho_i c_i(\vec{x}) + \frac{\mu}{2} \sum_{i=1}^{M} c_i(\vec{x})^2$$

Derive the gradient of the augmented Lagrangian function

$$\nabla_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\rho}, \mu) = \nabla_{\vec{x}} f(\vec{x}) - \sum_{i=1}^{M} \rho_i \nabla_{\vec{x}} c_i(\vec{x}) + \mu \sum_{i=1}^{M} c_i(\vec{x}) \nabla_{\vec{x}} c_i(\vec{x})$$

$$= \nabla_{\vec{x}} f(\vec{x}) + (\mu \sum_{i=1}^{M} c_i(\vec{x}) - \sum_{i=1}^{M} \rho_i) \nabla_{\vec{x}} c_i(\vec{x})$$

$$= \nabla_{\vec{x}} f(\vec{x}) + \sum_{i=1}^{M} (\mu c_i(\vec{x}) - \rho_i) \nabla_{\vec{x}} c_i(\vec{x})$$

Compute the augmented Lagrangian function is

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

The gradient of the augmented Lagrangian function is

$$\nabla_{\vec{x}}\mathcal{L}(\vec{x},\vec{\rho},\mu) = \begin{pmatrix} \frac{\partial}{\partial x_1}\mathcal{L}(\vec{x},\vec{\rho},\mu) \\ \frac{\partial}{\partial x_2}\mathcal{L}(\vec{x},\vec{\rho},\mu) \end{pmatrix}$$

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

The Hessian matrix of the augmented Lagrangian function is

$$\nabla^2_{\vec{x}}\mathcal{L}(\vec{x},\vec{\rho},\mu) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1}\mathcal{L}(\vec{x},\vec{\rho},\mu) & \frac{\partial^2}{\partial x_1 \partial x_2}\mathcal{L}(\vec{x},\vec{\rho},\mu) \\ \frac{\partial^2}{\partial x_2 \partial x_1}\mathcal{L}(\vec{x},\vec{\rho},\mu) & \frac{\partial^2}{\partial x_2 \partial x_2}\mathcal{L}(\vec{x},\vec{\rho},\mu) \end{pmatrix}$$

$$= \begin{pmatrix} -2\rho+2\mu((2x_1)(x_1)+(x_1^2+x_2^2-2)) & 4\mu x_2 x_1 \\ 4\mu x_2 x_1 & -2\rho+2\mu((2x_2)(x_2)+(x_1^2+x_2^2-2)) \end{pmatrix}$$

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

(b)

Theorem 0.1 (Second-order condition for augmented Lagrangian method). Let $\vec{x}^*$ be the local solution of the equality constraint optimization problem at which the LICQ is satisfied and the second-order sufficient conditions are satisfied. Then, there is a threshold value $\bar{y}$ such that for all $\mu \geqslant \bar{\mu}$, x^* is a strict local minimizer of $\mathcal{L}_A(\vec{x}^*,\vec{\rho}^*,\mu)$.

First, we need to find out the relation between Lagrangian function and augmented Lagrangian function

$$\nabla_{\vec{x}*}\mathcal{L}(\vec{x}^*,\vec{\rho}^*,\mu^*) = \nabla_{\vec{x}*}f(\vec{x}^*) + \sum_{i=1}^{M}(\mu^* c_i(\vec{x}^*) - \rho_i^*)\nabla_{\vec{x}*}c_i(\vec{x}^*)$$

$$= \nabla_{\vec{x}*}f(\vec{x}^*) + (\vec{\rho}^*)^\top \nabla_{\vec{x}*}\vec{c}(\vec{x}^*) + \mu^*\vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*}\vec{c}(\vec{x}^*)$$

$$= \nabla_{\vec{x}*}\mathcal{L}(\vec{x}^*,\vec{\rho}^*) + \mu^*\vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*}\vec{c}(\vec{x}^*)$$

Thus, the Hessian matrix

$$\nabla^2_{\vec{x}*}\mathcal{L}(\vec{x}^*,\vec{\rho}^*,\mu^*) = \nabla^2_{\vec{x}*}\mathcal{L}(\vec{x}^*,\vec{\rho}^*) + \mu^*\nabla_{\vec{x}*}\vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*}\vec{c}(\vec{x}^*) + \mu^*\vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*}\vec{c}(\vec{x}^*)$$

Since the constraints are all equality constraints, the constraints are all equal to 0, $\vec{c}(\vec{x}^*) = 0$

$$= \nabla^2_{\vec{x}*}\mathcal{L}(\vec{x}^*,\vec{\rho}^*) + \mu^*\nabla_{\vec{x}*}\vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*}\vec{c}(\vec{x}^*)$$

Since second-order condition holds, we've know the Hessian of the Lagrangian function $\vec{w}^\top \nabla^2_{\vec{x}}\mathcal{L}(\vec{x}^*,\vec{\rho}^*)\vec{w} \geqslant 0$ for any search direction belongs to critical cone(feasible search direction) $\forall \vec{w} \in \mathcal{C}(\vec{x}^*, \vec{\rho}^*)$. Thus, as we can multiply an arbitrary unit search direction $\forall \vec{w}, ||\vec{w}||_2 = 1$ on the both side of the augmented Lagrangian function as

$$\vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)\vec{w}$$

$$= \vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*)\vec{w} + \mu^* \vec{w}^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*)\vec{w}$$

$$= \vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*)\vec{w} + \mu^* ||\nabla_{\vec{x}*} \vec{c}(\vec{x}^*)\vec{w}||_2^2$$

As a result, the augmented Lagrangian function also satisfies the second-order condition for all μ^* sufficiently large because $\vec{w}^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*)\vec{w} = ||\nabla_{\vec{x}*} \vec{c}(\vec{x}^*)\vec{w}||_2 \geqslant 0$.

Before continue, we define a mathmetical tool **Rayleigh's Quotient**

Definition 0.1 (Rayleigh's quotient)**.** Rayleigh's quotient for a given complex Hermitian matrix $\mathbf{M}$ and nonzero vector $\vec{x}$ is defined as

$$R(\mathbf{M}, \vec{x}) = \frac{\vec{x}^* \mathbf{M} \vec{x}}{\vec{x}^* \vec{x}}.$$

Corollary 0.1.1. The Rayleigh's quotient would be bound between the largest and the smallest eigenvalues$\lambda_{max}, \lambda_{min}$ of the Hermitian matrix $\mathbf{M}$

$$\lambda_{min} \leqslant R(\mathbf{M}, \vec{x}) \leqslant \lambda_{max}$$

We've known that usually, the Hessian matrix $\nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)$ is a real, symmetric matrix and also a Hermitian matrix. In order to derive the minimal value of μ^*, we need to use Rayleigh's quotient to get the maximal and minimal values of $\vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)\vec{w}$

$$\frac{\vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)\vec{w}}{\vec{w}^\top \vec{w}} \in [\lambda_{min}, \lambda_{max}]$$

Where λ_{min} and λ_{max} are the smallest and the largest eigenvalue respectively. Because of $||\vec{w}||_2 = 1$, the formula can be written down as

$$\vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)\vec{w} \in [\lambda_{min}, \lambda_{max}]$$

Thus, if we can ensure $\vec{w}^\top \nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)\vec{w} \geqslant \lambda_{min} \geqslant 0$, the second-order condition would be satisfied surely.

$$\nabla_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*) = \nabla_{\vec{x}*} f(\vec{x}^*) + (\vec{\rho}^*)^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*) + \mu^* \vec{c}(\vec{x}^*)^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*)$$

Since $\vec{c}(\vec{x}^*)$ are equality constraints, thus, $\vec{x}^*$ are on the boundaries of the constraints and $\vec{c}(\vec{x}^*) = 0$.

$$= \nabla_{\vec{x}*} f(\vec{x}^*) + (\vec{\rho}^*)^\top \nabla_{\vec{x}*} \vec{c}(\vec{x}^*) = 0$$

Trivially, the optimal solution of the objective function is $\vec{x}^* = (-1, -1)$ and we can plug into the gradients of the augmented Lagrangian function

$$\nabla_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*) = \begin{pmatrix} 1 - \rho^*(2x_1^*) \\ 1 - \rho^*(2x_2^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Then, we can get $\rho^* = -0.5$.

Furthermore, we now have $\vec{x}^* = (-1, -1)$ and $\rho^* = 0.5$ and we can plug them into the Hessian matrix and check the second-order condition.

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

$$= \begin{pmatrix} 1 + 2\mu^*(2) & 4\mu^*(1) \\ 4\mu^*(1) & 1 + 2\mu^*(2) \end{pmatrix}$$

$$= \begin{pmatrix} 4\mu^* + 1 & 4\mu^* \\ 4\mu^* & 4\mu^* + 1 \end{pmatrix}$$

Denote the eigenvalue of the $\nabla^2_{\vec{x}*} \mathcal{L}(\vec{x}^*, \vec{\rho}^*, \mu^*)$ are λ_{eigen} and we can formulate

$$\det(\begin{pmatrix} 4\mu^* + 1 - \lambda_{eigen} & 4\mu^* \\ 4\mu^* & 4\mu^* + 1 - \lambda_{eigen} \end{pmatrix}) = 0$$

$$(4\mu^* + 1 - \lambda_{eigen})^2 - (4\mu^*)^2 = 0$$

$$16(\mu^*)^2 + (1 - \lambda_{eigen})^2 + 8\mu^*(1 - \lambda_{eigen}) - 16(\mu^*)^2 = 0$$

$$((1 - \lambda_{eigen}) + 8\mu^*)(1 - \lambda_{eigen}) = 0$$

Thus, $\lambda_{eigen} = 1, 1 + 8\mu^*$. Recall that the second-order condition requires $\lambda_{min} \geqslant 0$,

$$1 + 8\mu^* \geqslant 0$$

$$\mu^* \geqslant -\frac{1}{8}$$

Finally, we know that the minimal value of μ^* of the augmented Lagrangian function is $-\frac{1}{8}$.

Refers to Augmented Lagrangian Method

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

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

Answers are put here.

The condition number describes the sensitivity of the solution of the linear system toward the noise of the linear system. Formally, for a linear system $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}, \mathbf{A} \in \mathbb{R}^{\mathbf{N}\times\mathbf{M}}, \tilde{\mathbf{x}} \in \mathbb{R}^{\mathbf{M}}, \tilde{\mathbf{b}} \in \mathbb{R}^{\mathbf{N}}$, assume the system is perturbed by $\vec{b} + \delta\vec{b}$ and the corresponding solution would be $\vec{x} + \delta\vec{x}$.

$$\mathbf{A}(\tilde{\mathbf{x}} + \delta\tilde{\mathbf{x}}) = \tilde{\mathbf{b}} + \delta\tilde{\mathbf{b}}$$

The condition number is $\frac{||\delta\vec{x}||}{||\vec{x}||}$

Before deriving the closed form of the condition number, we need a lemma

Definition 0.2 (Metric Space)**.** Let X be a set and a function $d : X \times X \to \mathbb{R}$ should satisfy the following 4 properties

(a) Non-negativity: $d(x_1, x_2) \geqslant 0$

(b) $d(x_1, x_2) = 0 \quad \text{if and only if} \quad x_1 = x_2$

(c) Symmetric: $d(x_1, x_2) = d(x_2, x_1)$

(d) Triangular inequality: $d(x_1, x_2) + d(x_2, x_3) \geqslant d(x_1, x_3)$

Then, the pair (X, d) is called a metric space. Note that the definition is referred to this handout.

Now, we have the formal definition of the metric and we can derive the properties that a norm should hold.

Corollary 0.1.2 (Norm). Refers to matrix norm.

Definition 0.3 (Condition Number)**.** For a matrix $\mathbf{A} \in \mathbb{R}^{N\times M}$, the condition number of the matrix is defined by $\kappa(\mathbf{A}) = ||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}||$.

Proof:

For a linear system $\mathbf{A}\vec{x} = \vec{b}$, suppose that $\vec{b}$ is disturbed as $\vec{b} + \delta\vec{b}$ and the corresponding solution is $\vec{x} + \delta\vec{x}$.

$$\mathbf{A}(\vec{x} + \delta\vec{x}) = \vec{b} + \delta\vec{b}$$

Minus the original linear system $\mathbf{A}\vec{x} = \vec{b}$

$$\mathbf{A}\delta\vec{x} = \delta\vec{b}$$

$$\delta\vec{x} = \mathbf{A}^{-1}\delta\vec{b}$$

First of all, we need to derive the upper bound of the sensitivity $\frac{||\delta\vec{x}||}{||\vec{x}||}$

$$\frac{||\delta\vec{x}||}{||\vec{x}||} = \frac{||\mathbf{A}^{-1}\delta\vec{b}||}{||\vec{x}||}$$

Consider the inequality $||\mathbf{A}^{-1}\delta\vec{b}|| \leqslant ||\mathbf{A}^{-1}|| \cdot ||\delta\vec{b}||$

$$\leqslant \frac{||\mathbf{A}^{-1}|| \cdot ||\delta\vec{b}||}{||\vec{x}||} = ||\mathbf{A}^{-1}||\frac{||\mathbf{A}|| \cdot ||\delta\vec{b}||}{||\mathbf{A}|| \cdot ||\vec{x}||} = ||\mathbf{A}^{-1}||\frac{||\mathbf{A}|| \cdot ||\delta\vec{b}||}{||\vec{b}||}$$

Consider the inequality $||\mathbf{A}|| \cdot ||\vec{x}|| \geqslant ||\mathbf{A}\vec{x}||$

$$\leqslant ||\mathbf{A}^{-1}|| \cdot ||\mathbf{A}||\frac{||\delta\vec{b}||}{||\vec{b}||}$$

Next, consider the lower bound of the sensitivity $\frac{||\delta\vec{x}||}{||\vec{x}||}$

$$\frac{||\delta\vec{x}||}{||\vec{x}||} = \frac{||\mathbf{A}|| \cdot ||\delta\vec{x}||}{||\mathbf{A}|| \cdot ||\vec{x}||}$$

Consider the inequality $||\mathbf{A}|| \cdot ||\delta\vec{x}|| \geqslant ||\mathbf{A}\delta\vec{x}||$

$$\geqslant \frac{||\delta\vec{b}||}{||\mathbf{A}|| \cdot ||\vec{x}||} = \frac{||\delta\vec{b}||}{||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}\vec{b}||}$$

Consider the inequality $||\vec{x}|| = ||\mathbf{A}^{-1}\vec{b}|| \leqslant ||\mathbf{A}^{-1}|| \cdot ||\vec{b}||$

$$\geqslant \frac{||\delta\vec{b}||}{||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}|| \cdot ||\vec{b}||} = (||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}||)^{-1}\frac{||\delta\vec{b}||}{||\vec{b}||}$$

Since the upper bound and the lower bound of the condition number $\kappa(A)$ is affected by $||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}||$. Thus, we define the condition number as

$$\kappa(\mathbf{A}) = ||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}||$$

For more detail, please refers to this blog.

If we use L2 norm as the matrix norm, we can get $\kappa(\mathbf{A}) = \frac{||\delta\vec{x}||_2}{||\vec{x}||_2}$. Since the SVD of the matrix is

$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 0 \end{pmatrix} = \mathbf{U\Sigma V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Then we know the **condition number based on the L2-norm** $\frac{||\delta\vec{x}||_2}{||\vec{x}||_2} = |\frac{\sigma_1}{\sigma_n}| = \frac{4}{3}$.

In the other hand, if we use infinite norm as matrix norm, we can get $\kappa(\mathbf{A}) = ||\mathbf{A}||_\infty||\mathbf{A}^{-1}||_\infty$. We can compute the condition number

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

Thus, the **condition number based on the infinite norm is 2.**

Ill condition refers to this blog.

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.

Answers are put here.

Intuitively to say, the second-order correction add an additional correction term $\hat{p}_k$ to the original direction p_k, which is computed at the next step $x_k + p_k$ to decrease the constraint violation.

We have 3 assumption

(a) $c(\vec{x}_k) = 0$
(b) $c(\vec{x}_k) + \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \vec{p}_k = 0$
(c) $\hat{\vec{p}}_k \approx \vec{p}_k$ s.t. $\hat{\vec{p}}_k^\top \nabla^2_{\vec{x}_k} c(\vec{x}_k) \hat{\vec{p}}_k = \vec{p}_k^\top \nabla^2_{\vec{x}_k} c(\vec{x}_k) \vec{p}_k$

Instead of first-order approximation, the second-order correction use second-order approximation to approximate $c(\vec{x}_k + \vec{p}_k)$. If we write down the formula with Taylor expansion, the first-order approximation of $c(\vec{x}_k + \vec{p}_k)$ is

$$c(\vec{x}_k + \vec{p}_k) \approx c(\vec{x}_k) + \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \vec{p}_k$$

and the second-order approximation is

$$c(\vec{x}_k + \vec{p}_k) \approx c(\vec{x}_k) + \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \vec{p}_k + \frac{1}{2} \vec{p}_k^\top \nabla^2_{\vec{x}_k} c(\vec{x}_k) \vec{p}_k$$

Rearrange the order of the second-order approximation

$$\vec{p}_k^\top \nabla^2_{\vec{x}_k} c(\vec{x}_k) \vec{p}_k \approx c(\vec{x}_k + \vec{p}_k) - c(\vec{x}_k) - \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \vec{p}_k$$

Use assumption (c)

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

The second-order approximation of the second-order correction $\hat{\vec{p}}_k$

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

Then, we can plug second-order approximation into the second-order correction to replace $\hat{\vec{p}}_k^\top \nabla^2_{\vec{x}_k} c(\vec{x}_k)\hat{\vec{p}}_k$

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

$$= c(\vec{x}_k) + \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \hat{\vec{p}}_k + c(\vec{x}_k + \vec{p}_k) - c(\vec{x}_k) - \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \vec{p}_k$$

$$= \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \hat{\vec{p}}_k + c(\vec{x}_k + \vec{p}_k) - \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \vec{p}_k$$

Use assumption (a) and (b)

$$c(\vec{x}_k + \hat{\vec{p}}_k) \approx \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \hat{\vec{p}}_k + c(\vec{x}_k + \vec{p}_k)$$

Since the constraint $c(\vec{x})$ is equality constraint, let $c(\vec{x}_k + \hat{\vec{p}}_k) = 0$

$$c(\vec{x}_k + \hat{\vec{p}}_k) \approx \nabla_{\vec{x}_k} c(\vec{x}_k)^\top \hat{\vec{p}}_k + c(\vec{x}_k + \vec{p}_k) = 0$$

Finally, we can derive the second-order correction $\hat{\vec{p}}_k$ as

$$\hat{\vec{p}}_k = -\nabla_{\vec{x}_k} c(\vec{x}_k)(\nabla_{\vec{x}_k} c(\vec{x}_k)^\top \nabla_{\vec{x}_k} c(\vec{x}_k))^{-1} c(\vec{x}_k + \vec{p}_k)$$