# Numerical Optimization, Fall 2020
# Quiz solution

Fan Zhang and Xiangyu Yang

December 9, 2020

I. **Using KKT conditions to characterize the projections onto different convex sets**

Given a vector $\boldsymbol{y} \in \mathbb{R}^n$.

(1) (20 pts)

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}^T \boldsymbol{x} = b. \end{aligned} \tag{$\mathcal{P}_1$}$$

Solution:

We only analyse the case that the problem is feasible, that is $\boldsymbol{a} \neq \boldsymbol{0}$ if $b \neq 0$.
The associated Lagrangian of problem $(\mathcal{P}_1)$ is

$$L(\boldsymbol{x}, \lambda) = \boldsymbol{c}^T \boldsymbol{x} + \lambda(\boldsymbol{a}^T \boldsymbol{x} - b) = (\boldsymbol{c} + \lambda \boldsymbol{a})^T \boldsymbol{x} - \lambda b,$$

where $\lambda$ is the multiplier. Therefore, if there is no $\lambda$ such that $\boldsymbol{c} + \lambda \boldsymbol{a} = \boldsymbol{0}$, then the problem $(\mathcal{P}_1)$ is unbounded from below.

Furthermore, any optimal pair $(\boldsymbol{x}^*, \lambda^*)$ must satisfy following KKT conditions

$$\boldsymbol{c} + \lambda^* \boldsymbol{a} = \boldsymbol{0}, \tag{1a}$$
$$\boldsymbol{a}^T \boldsymbol{x}^* = b, \tag{1b}$$

yielding the optimal primal and dual solutions.

(2) (20 pts)

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 \\ \text{s.t.} \quad & \boldsymbol{a}^T \boldsymbol{x} = b. \end{aligned} \tag{$\mathcal{P}_2$}$$

Solution:

We only analyse the case that the problem is feasible, that is $\boldsymbol{a} \neq \boldsymbol{0}$ if $b \neq 0$. Then We consider two cases.

1). Suppose $\langle \boldsymbol{a}, \boldsymbol{z} \rangle = b$, then the solution to $(\mathcal{P}_2)$ is $\boldsymbol{x}^* = \boldsymbol{z}$.

2). Assume that $\langle \boldsymbol{a}, \boldsymbol{z} \rangle \neq b$. The associated Lagrangian of problem $(\mathcal{P}_2)$ is

$$L(\boldsymbol{x}, \lambda) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \lambda(\boldsymbol{a}^T \boldsymbol{x} - b),$$

where $\lambda$ is the multiplier. Then, any optimal pair $(\boldsymbol{x}^*, \lambda^*)$ must satisfy following KKT conditions

$$\boldsymbol{x}^* - \boldsymbol{z} + \lambda^* \boldsymbol{a} = \boldsymbol{0}, \tag{2a}$$
$$\boldsymbol{a}^T \boldsymbol{x}^* = b. \tag{2b}$$

Multiplying both sides of (2a) by $\boldsymbol{a}^T$ and combining with (2b), we have

$$\lambda^* \|\boldsymbol{a}\|_2^2 = -\boldsymbol{a}^T \boldsymbol{x}^* + \boldsymbol{a}^T \boldsymbol{z} = \boldsymbol{a}^T \boldsymbol{z} - b,$$

which yields $\lambda^* = \frac{\boldsymbol{a}^T \boldsymbol{z} - b}{\|\boldsymbol{a}\|_2^2}$. Combining with (2a)

$$\boldsymbol{x}^* = \boldsymbol{z} + \frac{(b - \boldsymbol{a}^T \boldsymbol{z})}{\|\boldsymbol{a}\|_2^2} \boldsymbol{a}.$$

(3) (20 pts)

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}. \end{array} \tag{$\mathcal{P}_3$}$$

Solution: We only analyse the case that the problem is feasible, that is $\boldsymbol{b} \in \operatorname{span}\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_n\}$, where $\boldsymbol{c}_i$ $(i = 1, \cdots, n)$ is the $i$th column of $\boldsymbol{A}$. Then We consider two cases.

1). Suppose $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}$, then the solution to $(\mathcal{P}_3)$ is $\boldsymbol{x}^* = \boldsymbol{z}$.

2). Assume that $\boldsymbol{A}\boldsymbol{z} \neq \boldsymbol{b}$. The additional assumption on matrix $\boldsymbol{A}$ is: $\boldsymbol{A}$ has full row rank. The associated Lagrangian of problem $(\mathcal{P}_3)$ is

$$L(\boldsymbol{x}, \lambda) = \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \lambda^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}),$$

where $\lambda \in \mathbb{R}^m$ is the multiplier. Then, any optimal pair $(\boldsymbol{x}^*, \lambda^*)$ must satisfy following KKT conditions

$$\boldsymbol{x}^* - \boldsymbol{z} + \boldsymbol{A}^T\lambda^* = \boldsymbol{0}, \tag{3a}$$

$$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}. \tag{3b}$$

Multiplying both sides of (3a) by $\boldsymbol{A}$ and combining with (3b), we have

$$\boldsymbol{A}\boldsymbol{A}^T\lambda^* = \boldsymbol{A}\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{z} - \boldsymbol{b}.$$

According to the assumption $\boldsymbol{A}$ has full row rank, it holds $\lambda^* = (\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{z}-\boldsymbol{b})$. Combining with (3a)

$$\boldsymbol{x}^* = \boldsymbol{z} - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{z}-\boldsymbol{b}).$$

(4) (20 pts) Projecting $\boldsymbol{y}$ onto a (closed) halfspace:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}\in\mathbb{R}^n} & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ \text{s.t.} & \boldsymbol{a}^T\boldsymbol{x} \leq b, \end{array} \tag{4}$$

Solution: We consider two cases.

1). Suppose $\langle \boldsymbol{a}, \boldsymbol{y}\rangle \leq b$, then the solution to (4) is $\boldsymbol{x}^* = \boldsymbol{y}$.

2). Assume that $\langle \boldsymbol{a}, \boldsymbol{y}\rangle > b$. The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}^*, \lambda) = \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \lambda(\boldsymbol{a}^T\boldsymbol{x}-b), \tag{5}$$

where $\lambda \in \mathbb{R}_+$ is the introduced Lagrange multiplier. Then the first-order optimality conditions are

$$\begin{aligned} \boldsymbol{x}^* - \boldsymbol{y} + \lambda\boldsymbol{a} &= \boldsymbol{0}, && \text{((3) stationay conditions)} \\ \boldsymbol{a}^T\boldsymbol{x}^* &\leq b, && \text{((4) primal feasibility)} \\ \lambda &\geq 0, && \text{((5) dual feasibility)} \\ \lambda(\boldsymbol{a}^T\boldsymbol{x}^*-b) &= 0. && \text{((6) complementary conditions)} \end{aligned}$$

By ((3) stationay conditions), we have

$$\lambda = \frac{\langle \boldsymbol{y}, \boldsymbol{a}\rangle - b}{\|\boldsymbol{a}\|_2^2}. \qquad \textit{(why?)} \tag{7}$$

By substituting (7) into ((3) stationay conditions), we have

$$\boldsymbol{x}^* = \boldsymbol{y} - \frac{\langle \boldsymbol{y}, \boldsymbol{a}\rangle - b}{\|\boldsymbol{a}\|_2^2}\boldsymbol{a}.$$

Therefore, we have

$$\boldsymbol{x}^* = \begin{cases} \boldsymbol{y} & \text{if } \langle \boldsymbol{a}, \boldsymbol{y}\rangle \leq b, \\ \boldsymbol{y} - \frac{\langle \boldsymbol{y}, \boldsymbol{a}\rangle - b}{\|\boldsymbol{a}\|_2^2}\boldsymbol{a} & \text{if } \langle \boldsymbol{a}, \boldsymbol{y}\rangle > b. \end{cases}$$

(5) (20 pts) Projecting $\boldsymbol{y}$ onto the unit $\ell_2$ ball:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \quad & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{x}\|_2^2 \leq 1. \end{aligned} \tag{8}$$

Solution: We consider two cases.

1). Suppose $\|\boldsymbol{y}\|_2^2 \leq 1$, then the solution to (8) is $\boldsymbol{x}^* = \boldsymbol{y}$.

2). Assume that $\|\boldsymbol{y}\|_2^2 > 1$. The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}^*, \lambda^*) = \frac{1}{2}\|\boldsymbol{x}^* - \boldsymbol{y}\|_2^2 + \lambda^*(\|\boldsymbol{x}^*\|_2^2 - 1), \tag{9}$$

where $\lambda \in \mathbb{R}_+$ is the introduced Lagrange multiplier. Then the first-order optimality conditions are

$$\boldsymbol{x}^* - \boldsymbol{y} + 2\lambda\boldsymbol{x}^* = \boldsymbol{0}, \tag{10a}$$
$$\|\boldsymbol{x}^*\|_2^2 \leq 1, \tag{10b}$$
$$\lambda^* \geq 0, \tag{10c}$$
$$\lambda^*(\|\boldsymbol{x}^*\|_2^2 - 1) = 0. \tag{10d}$$

By (10a), we have

$$\lambda^* = \frac{\langle \boldsymbol{y}, \boldsymbol{x}^* \rangle - 1}{2}. \qquad \textit{(why?)} \tag{11}$$

By substituting (11) into (10a), we have

$$(\boldsymbol{y}^T\boldsymbol{x}^*)\boldsymbol{x}^* - \boldsymbol{y} = 0. \tag{12}$$

Note that $\boldsymbol{y}^T\boldsymbol{x}^*$ is a scalar, it means that $\boldsymbol{x}$ and $\boldsymbol{y}$ share the same direction, and consequently, the solution is the intersection of the $\ell_2$ ball surface with the line connecting the point $\boldsymbol{y}$ to the origin. Therefore, we have

$$\boldsymbol{x}^* = \begin{cases} \boldsymbol{y} & \text{if } \|\boldsymbol{y}\|_2^2 \leq 1, \\ \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} & \text{if } \|\boldsymbol{y}\|_2^2 > 1. \end{cases} \tag{13}$$

(6) (0 points) Projection $\boldsymbol{y}$ onto the unit $\ell_1$ ball problem:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \quad & \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{x}\|_1^2 \leq \gamma. \end{aligned} \tag{14}$$

Solution: We consider two cases.

1). Suppose $\|\boldsymbol{y}\|_1^2 \leq 1$, then the solution to (14) is $\boldsymbol{x}^* = \boldsymbol{y}$.

2). Assume that $\|\boldsymbol{y}\|_1^2 > 1$. The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}^*, \lambda^*) = \frac{1}{2}\|\boldsymbol{x}^* - \boldsymbol{y}\|_2^2 + \lambda^*(\|\boldsymbol{x}^*\|_1^2 - 1), \tag{15}$$

where $\lambda \in \mathbb{R}_+$ is the introduced Lagrange multiplier. Then the first-order optimality conditions are

$$\boldsymbol{x}^* - \boldsymbol{y} + 2\lambda^*\boldsymbol{s} = \boldsymbol{0}, \tag{16a}$$
$$\|\boldsymbol{x}^*\|_1^2 \leq 1, \tag{16b}$$
$$\lambda^* \geq 0, \tag{16c}$$
$$\lambda^*(\|\boldsymbol{x}^*\|_1^2 - 1) = 0, \tag{16d}$$

where $\boldsymbol{s} = \partial\|\boldsymbol{x}^*\|_1 \in \mathbb{R}^n$, and $s_i = \text{sign}(x_i)$ if $x_i \neq 0$, $s_i \in [-1, 1]$ if $x_i = 0$. By (16a), we have

$$x_i^* = [S_{2\lambda^*}(\boldsymbol{y})] = \begin{cases} y_i - 2\lambda^* & \text{if } y_i > 2\lambda^*, \\ 0 & \text{if } -2\lambda^* \leq y_i \leq 2\lambda^*, \\ y_i + 2\lambda^* & \text{if } y_i < -2\lambda^*. \end{cases} \tag{17}$$

where $S_{2\lambda^*}(\cdot)$ is the soft-thresholding operator. Use a more compact expression, $\boldsymbol{x}^*$ can be characterized by

$$\boldsymbol{x}^* = \operatorname{sgn}(\boldsymbol{y}) \max(|\boldsymbol{y}| - 2\lambda^*, 0), \tag{18}$$

where sgn is the signum function. (Note: In (18), all operations are defined in a component-wise manner.)

Therefore, the only thing is left for obtaining $\boldsymbol{x}^*$ is to find the value of $\lambda^*$ (why?). For this, we define the function

$$f(\lambda) = \sum_i \max(|y_i| - 2\lambda, 0) - 1; \tag{19}$$

the solution to $f(\lambda) = 0$ leads to obtain $\lambda^*$. Moreover, $f(\lambda)$ is convex and piece-wise linear, and the well-known root-finding algorithm—Newton's method, can thus be applied. (Note: Regarding the projection onto the $\ell_1$ ball, please refer to [1, 2] for more details.)

# References


[1] L. Condat, "Fast projection onto the simplex and the $\ell_1$ ball," *Mathematical Programming*, vol. 158, no. 1-2, pp. 575–585, 2016.

[2] F. Zhang, H. Wang, J. Wang, and K. Yang, "Inexact primal–dual gradient projection methods for nonlinear optimization on convex set," *Optimization*, vol. 69, no. 10, pp. 2339–2365, 2020.