ESE415 OPTIMIZATION

Tutorial 9

Xiaojian Xu

04/19/2019

Guideline

HW5 RECITATION

HW5 Recitation(P1)

Solve the following convex programs by using the KKT conditions.

$$f(\boldsymbol{x}) = x_1^2 + x_2^2 - 4x_1 - 4x_2,\ g_1(\boldsymbol{x}) = x_1^2 - x_2,\ g_2(\boldsymbol{x}) = x_1 + x_2 - 2, \text{ and } \mathcal{X} = \mathbb{R}^2.$$

- *dual feasibility*: $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$;
- *primal feasibility*: $g_1(\boldsymbol{x}) \leq 0$ and $g_2(\boldsymbol{x}) \leq 0$;
- *complementary slackness*: $\lambda_1 g_1(\boldsymbol{x}) = 0$ and $\lambda_2 g_2(\boldsymbol{x}) = 0$;
- *stationarity*: $\nabla f(\boldsymbol{x}) + \lambda_1 \nabla g_1(\boldsymbol{x}) + \lambda_2 \nabla g_2(\boldsymbol{x}) = \mathbf{0}$.

Hints (p1)

b) The solution of this problem is illustrated in Fig. 1(b). First note that the problem is superconsistent and the functions f, g_1, and g_2 are all convex and continuously differentiable. Therefore, if we can find a feasible point that satisfies the KKT conditions, we have found the solution.

- *dual feasibility*: $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$;
- *primal feasibility*: $g_1(\boldsymbol{x}) \leq 0$ and $g_2(\boldsymbol{x}) \leq 0$;
- *complementary slackness*: $\lambda_1 g_1(\boldsymbol{x}) = 0$ and $\lambda_2 g_2(\boldsymbol{x}) = 0$;
- *stationarity*: $\nabla f(\boldsymbol{x}) + \lambda_1 \nabla g_1(\boldsymbol{x}) + \lambda_2 \nabla g_2(\boldsymbol{x}) = \mathbf{0}$.

Note that for our problem, we have

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 - 4 \\ 2x_2 - 4 \end{bmatrix}, \quad \nabla g_1(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix}, \quad \text{and} \quad \nabla g_1(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The second and the third KKT condition reduce respectively to

$$\begin{cases} \lambda_1(x_1^2 - x_2) = 0 \\ \lambda_2(x_1 + x_2 - 2) = 0 \end{cases} \quad \text{and} \quad \begin{cases} 2x_1 - 4 + 2\lambda_1 x_1 + \lambda_2 = 0 \\ 2x_2 - 4 - \lambda_1 + \lambda_2 = 0 \end{cases}. \tag{3}$$

Without any obvious simplification, we will consider four cases:

Case 1: $\lambda_1 = \lambda_2 = 0$. By using this in eq. (3), we obtain

$$x_1 = x_2 = 2,$$

which is not a consistent solutions since

$$x_1 + x_2 - 2 = 2 > 0.$$

Hints (p1)

Case 2: $\lambda_1 = 0$ and $x_1 + x_2 = 2$. By using this in eq. (3), we obtain

$$\begin{cases} 2x_1 - 4 + \lambda_2 = 0 \\ 2x_2 - 4 + \lambda_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 + x_2 = 2 \\ x_2 - x_1 = 0 \end{cases} \Rightarrow x_1 = x_2 = 1 \Rightarrow \lambda_2 = 4 - 2x_1 = 2.$$

Quick check reveals that the point $\boldsymbol{x}^* = (1, 1)$ with $\lambda_1 = 0$ and $\lambda_2 = 2$ is indeed a feasible point.

Case 3: $x_1^2 - x_2 = 0$ and $\lambda_2 = 0$. By using this in eq. (3), we obtain

$$\begin{cases} 2x_1 - 4 + 2\lambda_1 x_1 = 0 \\ 2x_2 - 4 - \lambda_1 = 0 \end{cases} \Rightarrow 2x_1^3 - 3x_1 - 2 = 0 \Rightarrow x_1 \approx 1.4757 \Rightarrow x_2 \approx \pm\sqrt{1.4757}.$$

Note that for $x_2 \approx \sqrt{1.4757}$, we have $x_1 + x_2 - 2 \approx 0.6895 > 0$, which is not consistent. On the other hand for $x_2 \approx -\sqrt{1.4757}$, we have

$$\lambda_1 = 2x_2 - 4 = -2 \cdot \sqrt{1.4757} - 4 < 0,$$

which does not satisfy KKT condition of $\lambda_1 \geq 0$.

Case 4: $x_1^2 - x_2 = 0$ and $x_1 + x_2 = 2$. We seek the solution of the system

$$\begin{cases} x_1^2 - x_2 = 0 \\ x_1 + x_2 = 2 \end{cases} \Rightarrow x_1^2 + x_1 - 2 = (x_1 - 1) \cdot (x_1 + 2) = 0 \Rightarrow \begin{cases} x_1 = 1 \text{ and } x_2 = 1 \\ x_1 = -2 \text{ and } x_2 = 4 \end{cases}.$$

The first solution $(1, 1)$ was already obtained in *Case 2*. For the second solution, we have that

$$\begin{cases} -4\lambda_1 + \lambda_2 = 8 \\ -\lambda_1 + \lambda_2 = -4 \end{cases} \Rightarrow \begin{cases} \lambda_1 = -4 \\ \lambda_2 = -8 \end{cases}.$$

Since those values are negative, they do not satisfy the first KKT condition.

Hence, $\boldsymbol{x}^* = (1, 1)$ is the only solution to the convex program.

HW5 Recitation(P1)

Find the closed-form solutions for the projections onto the following sets:

$\mathcal{X}_2 = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} \leq b\}$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are some known parameters.

Hints（p1）

b) We would like to minimize $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_{\ell_2}^2$ subject to $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} - b \leq 0$. We note that the problem is superconsistent and the functions f and g are convex and continuously differentiable. Therefore, if we can find a feasible point that satisfies the KKT conditions, we have found the solution. Thus, our goal is to solve the following system of equations in $n + 1$ unknowns

- *dual feasibility*: $\lambda \geq 0$;
- *primal feasibility*: $g(\boldsymbol{x}) \leq 0$;
- *complementary slackness*: $\lambda g(\boldsymbol{x}) = 0$;
- *stationarity*: $\nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = \mathbf{0}$.

hints

Note that for our problem, we have

$$\nabla f(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{y} \quad \text{and} \quad \nabla g(\boldsymbol{x}) = \boldsymbol{a}.$$

We consider two cases.

Case 1: We consider the scenario when $\lambda = 0$. From the stationarity condition, we get that

$$\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{y}.$$

Note, however, that this is a valid solution only when it is feasible, *i.e.*, $\boldsymbol{a}^\top \boldsymbol{y} \leq b$.

Case 2: We consider the scenario when $\lambda > 0$ and $\boldsymbol{a}^\top \boldsymbol{x} = b$. From the stationarity condition, we get that

$$\boldsymbol{x} - \boldsymbol{y} + \lambda \boldsymbol{a} = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{y} - \lambda \boldsymbol{a}.$$

We use this within the complementary slackness condition

$$\boldsymbol{a}^\top(\boldsymbol{y} - \lambda \boldsymbol{a}) - b = \boldsymbol{a}^\top \boldsymbol{y} - \lambda \boldsymbol{a}^\top \boldsymbol{a} - b = 0 \quad \Rightarrow \quad \lambda = \frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\boldsymbol{a}^\top \boldsymbol{a}}.$$

This leads to

$$\boldsymbol{x} = \boldsymbol{y} - \left[\frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\|\boldsymbol{a}\|^2}\right] \boldsymbol{a} = \boldsymbol{y} + \left[\frac{b - \boldsymbol{a}^\top \boldsymbol{y}}{\|\boldsymbol{a}\|}\right] \boldsymbol{a}.$$

Note that this solution is valid when it satisfies the dual feasibility

$$\lambda = \frac{\boldsymbol{a}^\top \boldsymbol{y} - b}{\boldsymbol{a}^\top \boldsymbol{a}} > 0 \quad \Leftrightarrow \quad \boldsymbol{a}^\top \boldsymbol{y} > b.$$

We combine the solutions to obtain for $\mathcal{X}_2 = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} \leq b\}$:

$$\mathrm{proj}_{\mathcal{X}_2}(\boldsymbol{y}) = \begin{cases} \boldsymbol{y} & \text{when } \boldsymbol{a}^\top \boldsymbol{y} \leq b \\ \boldsymbol{y} + \left[\dfrac{b - \boldsymbol{a}^\top \boldsymbol{y}}{\|\boldsymbol{a}\|}\right] \boldsymbol{a} & \text{when } \boldsymbol{a}^\top \boldsymbol{y} > b \end{cases}.$$

HW5 Recitation(P3)

Consider the following optimization problem

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) = x_1 + x_2 \\ &\text{subject to } g(\boldsymbol{x}) = x_1^2 - x_2 - 2 \leq 0. \end{aligned}$$

(a) Use the KKT conditions to solve this problem analytically.

(b) Plot the level sets of f by using the `contour` function for $-3/2 \leq x_1 \leq 3/2$ and $-2 \leq x_2 \leq 1$. On the same figure, use `plot` to draw the boundary of the constraint g and to show the location of the optimal point $\boldsymbol{x}^* \in \mathbb{R}^2$. Submit the plot and the printout of the code used to generate it.

(c) Find the sequence of solutions $\{\boldsymbol{x}^t\}$ obtained with the quadratic penalty method of Lecture 23. Find the convergence rate $\delta_t \triangleq (f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*))$ of the method and characterize it as sublinear, linear, superlinear, or quadratic.

(d) Show that the objective function r_t corresponding to the absolute value penalty function has no critical points outside of the set $\mathcal{G}_0 \triangleq \{\boldsymbol{x} \in \mathbb{R}^2 : g(\boldsymbol{x}) = 0\}$ for any $t > 1$. Use this to find the minimizer of r_t.

hints

(a) We form the Lagrangian

$$L(\boldsymbol{x}, \lambda) = x_1 + x_2 + \lambda(x_1^2 - x_2 - 2)$$

and seek its critical points by differentiating it with respect to $\boldsymbol{x}$

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = \begin{bmatrix} 1 + 2\lambda x_1 \\ 1 - \lambda \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad \lambda = 1,\ x_1 = -\frac{1}{2}.$$

Additionally, from complementary slackness, we have

$$\lambda(x_1^2 - x_2 - 2) = 0 \quad \Rightarrow \quad x_2 = \frac{1}{4} - 2 = -\frac{7}{4}.$$

The solution is thus given by

$$\boldsymbol{x}^* = \left(-\frac{1}{2}, -\frac{7}{4}\right), \quad \lambda^* = 1, \quad \text{and} \quad f(\boldsymbol{x}^*) = -\frac{9}{4}$$

(b) The plot is given in Fig. 1.

hints

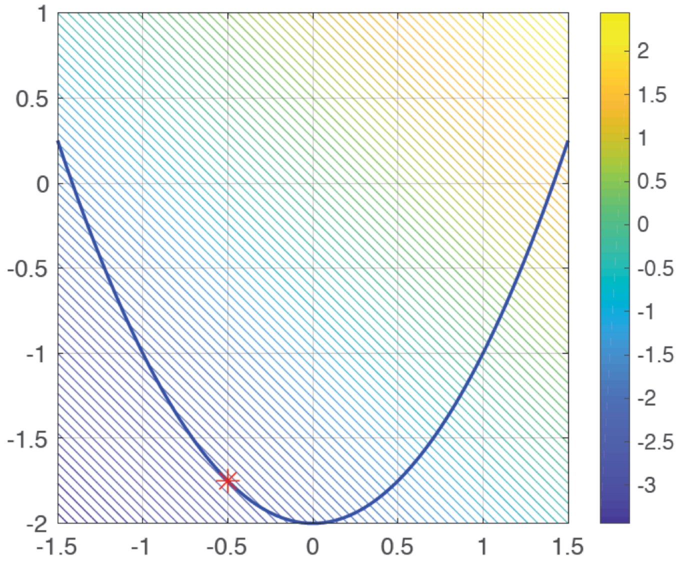

Figure 1: Visual illustration of the level sets of $f(\boldsymbol{x}) = x_1 + x_2$, of the function $x_2 = x_1^2 - 2$, and of the solution $\boldsymbol{x}^* = (-1/2, -7/4)$.

hints

(c) We form the quadratic penalty function

$$q_t(\boldsymbol{x}) = f(\boldsymbol{x}) + t \cdot p(\boldsymbol{x}) = x_1 + x_2 + \frac{t}{2}[(x_1^2 - x_2 - 2)_+]^2,$$

where $(\cdot)_+$ returns the non-negative part of its input. We seek the critical points of the function by seeking the solution to

$$\nabla q_t(\boldsymbol{x}) = \begin{bmatrix} 1 + 2tx_1(x_1^2 - x_2 - 2)_+ \\ 1 - t(x_1^2 - x_2 - 2)_+ \end{bmatrix} = \mathbf{0}.$$

Note that when $x_1^2 - x_2 \leq 2$, the function $\nabla q_t(\boldsymbol{x}) = \mathbf{0}$ does not have a solution. On the other hand, when $x_1^2 - x_2 > 2$, we find

$$x_1 = -\frac{1}{2} \quad \text{and} \quad x_2 = \frac{1}{t} - \frac{7}{4}.$$

Hence, the quadratic penalty method generates the following sequences

$$\boldsymbol{x}^t = \left(-\frac{1}{2}, -\frac{7}{4} - \frac{1}{t}\right) \quad \text{and} \quad f(\boldsymbol{x}^t) = -\frac{9}{4} + \frac{1}{t}.$$

This makes it clear that

$$\lim_{t \to \infty} f(\boldsymbol{x}^t) = f(\boldsymbol{x}^*).$$

Additionally, we have that

$$\delta_t = (f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*)) = \frac{1}{t},$$

which corresponds to sublinear convergence.

(d) We form the absolute value penalty function

$$r_t(\boldsymbol{x}) = f(\boldsymbol{x}) + t \cdot s(\boldsymbol{x}) = x_1 + x_2 + t(x_1^2 - x_2 - 2)_+.$$

hints

Note that when $x_1^2 - x_2 > 2$, the gradient of r_t is given by

$$\nabla r_t(\boldsymbol{x}) = \begin{bmatrix} 1 + 2tx_1 \\ 1 - t \end{bmatrix},$$

which indicates that $\nabla r_t(\boldsymbol{x}) = \mathbf{0}$ does not have any solutions for $t > 1$. This means that the minimizer of r_t occurs on the parabola $x_1^2 - x_2 - 2 = 0$. By substituting this into the objective, we find

$$f(x_1, x_1^2 - 2) = x_1^2 + x_1 - 2,$$

which leads to the solution

$$\boldsymbol{x}^t = \left(-\frac{1}{2}, -\frac{7}{4}\right), \quad \forall t \geq 1.$$

hints

Further explanation for (d) :

(d). $r_t(x) = f(x) + t\,|g(x)^+|$

$$= x_1 + x_2 + t\,|(x_1^2 - x_2 - 2)_+|$$

$$= \begin{cases} x_1 + x_2 & \text{for } x_1^2 - x_2 - 2 \le 0 \\ x_1 + x_2 + t(x_1^2 - x_2 - 2) & \text{for } x_1^2 - x_2 - 2 > 0 \end{cases}$$

for $g(x) > 0$

$$r_t(x)' = \begin{bmatrix} 1 + 2t x_1 \\ 1 - t \end{bmatrix}$$

for $g(x) \le 0$

$$r_t(x)' = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

for any $t > 1$, $r_t(x)' \neq 0$, $\therefore$ there is no critical point for any $x \notin \mathcal{G} = \{x \in R^2 : g(x) = 0\}$.

So we know $\begin{cases} r_t(x)' = 0 \\ g(x) = 0 \end{cases} \Rightarrow x^* = \begin{cases} x_1 = -\frac{1}{2} \\ x_2 = -\frac{7}{4} \end{cases}$

with $t = 1$

Thanks