Continuous Optimization: Assignment 10

Due on July 2, 2024

Honglu Ma **Hiroyasu Akada** **Mathivathana Ayyappan**

Exercise 1

We can reformulate the problem as follows:

$$\min_{x \in \mathbb{R}^n} \langle c, x \rangle \quad \text{s.t.} \quad \forall i \in \{1, \dots, m\} : b_i - \langle a_i, x \rangle = 0 \quad \text{and} \quad \forall j \in \{1, \dots, n\} : -x_j \leq 0$$

where a_i is the i-th row of A, b_i is the i-th element of b and x_j is the j-th element of x.
We can name the objective function as $f(x) = \langle c, x \rangle$, the equality constraints as $f_i(x) = b_i - \langle a_i, x \rangle$, $\forall i \in \{1, \dots, m\}$ and the inequality constraints as $g_j(x) = -x_j$, $\forall j \in \{1, \dots, n\}$ such that we have a problem that fits the general form provided in Corollary 15.19, namely:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad f_i(x) = 0, i \in \mathcal{E} \quad \text{and} \quad g_j(x) \leq 0, j \in \mathcal{I}$$

By Corollary 15.19, we know that at optimal $x^\star$, we have

$$\begin{aligned} \nabla f(x^\star) + \sum_{i \in \mathcal{E}} \lambda_i \nabla f_i(x^\star) + \sum_{j \in \mathcal{A}(x^\star)} \mu_j \nabla g_j(x^\star) = 0 \\ c - \sum_{i=1}^{m} \lambda_i a_i - \sum_{j \in \mathcal{A}(x^\star)} \mu_j e_j = 0 \\ c - A^\top \lambda - \mu = 0 \end{aligned}$$

where

$$\mu = \begin{cases} 0 & \text{if } x_j^\star > 0 \\ \mu_j > 0 & \text{if } x_j^\star = 0 \end{cases}$$

observe that $\mu \geq 0$ which is the fouth KKT condition and $\mu_j x_j^\star = 0$, $\forall j \in \{1, \dots, n\}$ which is the fifth KKT condition a.k.a the complementary condition. Also by reformulate the equation we derivate at optimal, we have the first KKT condition:

$$c = A^\top \lambda + \mu$$

Exercise 2

Note that

$$\text{tr}(B^\top X) = \sum_{i=1}^{n} \langle B^\top_{i,} X_{,i} \rangle$$

where $B^\top_{i,}$ denotes the i-th row of $B^\top$ and $X_{,i}$ denotes the i-th column of X which can also just be seen as a sum of dot products between each column of B and X i.e. $\sum_{i=1}^{n} \langle b_i, x_i \rangle$ where b_i is the i-th column of B and x_i is the i-th column of X.
To find the minimum of a sum is the same as finding the minimum of each term in the sum i.e. instead of 1 objective function $f(x) = \sum_{i=1}^{n} \langle b_i, x_i \rangle$, we have n objective functions $f_i(x) = \langle b_i, x_i \rangle$ with same constraints on x_i as such

$$\min_{x_i \in \mathbb{R}^n} \langle b_i, x_i \rangle \quad \text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = 1 \quad \text{and} \quad \forall j \in \{1, \dots, n\} : x_{ij} \geq 0$$

for all $i \in \{1, \dots, n\}$

Exercise 2 continued on next page...

The Lagrangian for the i-th objective function is (to simplify notation, we will drop the i subscript from now on)

$$\begin{aligned} L(x, \lambda, \mu) &= \langle b, x\rangle - \lambda \left(\sum_{j=1}^{n} x_j - 1 \right) - \sum_{j=1}^{n} \mu_j x_j \\ &= \langle b, x\rangle - \lambda \left(\sum_{j=1}^{n} x_j - 1 \right) - \langle \mu, x\rangle \end{aligned}$$

where

$$\mu = \begin{cases} 0 & \text{if } x_j > 0 \\ \mu_j > 0 & \text{if } x_j = 0 \end{cases}$$

We take the derivative of the Lagrangian with respect to x, λ and μ respectively:

$$\begin{aligned} &\nabla_x L(x, \lambda, \mu) = b - \lambda \mathbb{1} - \mu = 0 \Leftrightarrow \mu_j = b_j - \lambda, \ \forall j \in \{1, \dots, n\} \\ &\nabla_\lambda L(x, \lambda, \mu) = -\left(\sum_{j=1}^{n} x_j - 1 \right) = 0 \Leftrightarrow \sum_{j=1}^{n} x_j = 1 \\ &\nabla_\mu L(x, \lambda, \mu) = -x \le 0 \Leftrightarrow x \ge 0 \\ &\mu \ge 0 \\ &\mu_j \cdot x_j = 0, \ \forall j \in \{1, \dots, n\} \end{aligned}$$

We have such system of equations ($n + 1$ unknowns and $n + 1$ equations)

$$\begin{aligned} \mu_j \cdot x_j &= 0 \\ \sum_{j=1}^{n} x_j &= 1 \end{aligned}$$

Solve for each column in X and we have the solution to the original problem.

Exercise 3

Observe that the shape of the constraint set C is like a box (as in $\mathbb{R}^3$)

(a)

The Conditional Gradient Method first finds

$$\tilde{x}^{(k)} \in \operatorname{argmin}_{x \in C} \langle \nabla f(x^{(k)}), x - x^{(k)} \rangle$$

and the descend direction is defined as $d^{(k)} = \tilde{x}^{(k)} - x^{(k)}$. The time step τ_k is determined by the backtracking line search that satisfies the Armijo condition with parameter γ. At last, the next point is updated by $x^{(k+1)} = x^{(k)} + \tau_k d^{(k)}$.

Exercise 3 continued on next page...

(b)

The Projected Gradient Method first find a point $\bar{x}^{(k)} := x^{(k)} - \alpha \nabla f(x^{(k)})$ and calculate the its projection onto the feasible set C by $\tilde{x}^{(k)} := \text{proj}_C(\bar{x}^{(k)})$. Then the desend direction is the difference between the projected point and the current point i.e. $d^{(k)} = \tilde{x}^{(k)} - x^{(k)}$ and the time step τ_k is determined by the backtracking line search that satisfies the Armijo condition with paramter γ. At last, the next point is updated by $x^{(k+1)} = x^{(k)} + \tau_k d^{(k)}$.
The only step that we need to derive is the projection onto the feasible set C which is defined by the following minimization problem:

$$\tilde{x}^{(k)} = \text{argmin}_{x \in C} ||x - \bar{x}^{(k)}||^2$$

by the optimality condition, we have

$$\tilde{x}^{(k)} - \bar{x}^{(k)} \in N_C(\tilde{x}^{(k)}),\, \tilde{x}^{(k)} \in C$$

and the projection is given by

$$\tilde{x}_i^{(k)} = \begin{cases} q_i & \text{if } \bar{x}_i^{(k)} \geq q_i \\ p_i & \text{if } \bar{x}_i^{(k)} \leq p_i \\ \bar{x}_i^{(k)} & \text{otherwise} \end{cases}$$