ECE 490: Introduction to Optimization **Fall 2018**

Lecture 21

Karush-Kuhn-Tucker (KKT) Conditions

Lecturer: Bin Hu, Date:11/13/2018

In this lecture, we discuss the optimality conditions for optimization problems with inequality constraints. Specifically, we consider the following problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0,\ i = 1, \ldots, m \\ & g_j(x) \le 0,\ j = 1, \ldots, l \end{aligned} \tag{21.1}$$

Suppose x^* is a local min for (21.1). Suppose $\mathcal{A}(x^*)$ is the set of inequality constraints that are active, i.e. $g_j(x^*) = 0$ if $j \in \mathcal{A}(x^*)$, and $g_j(x^*) < 0$ if $j \notin \mathcal{A}(x^*)$. Then the point x^* is also a local min for the following optimization problem with equality constraints:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0,\ j = 1, \ldots, m \\ & g_j(x) = 0,\ j \in \mathcal{A}(x^*) \end{aligned} \tag{21.2}$$

Explanations. When looking at a local min, we only need to consider a sufficiently small neighborhood around the point. For a sufficiently small neighborhood around x^*, the condition $g_j(x) < 0$ for $j \notin \mathcal{A}(x^*)$ will be automatically guaranteed for all the points given the continuity of g_j and the fact $g(x^*) < 0 \forall j \notin \mathcal{A}(x^*)$. Therefore, these constraints can be dropped, and the original problem (21.1) behaves like the reformulated problem (21.2).

Now we can rewrite the optimality conditions for optimization problems with equality constraints as optimality conditions for the general problem (21.1). This leads to the famous Karush-Khun-Tucker (KKT) conditions.

21.1 KKT Conditions

Suppose x^* is a regular local min for (21.2). Then based on the Lagrange theorem, we know that there exists Lagrange multipliers λ_1^*, λ_2^*, …, λ_m^* and μ_j^* for all $j \in \mathcal{A}(x^*)$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_j^* \nabla h_i(x^*) + \sum_{j \in \mathcal{A}(x^*)} \mu_j^* \nabla g_j(x^*) = 0.$$

Here λ_i^* and μ_j^* are just scalers. Assigning zero Lagrange multipliers to inactive inequality constraints, the above condition becomes

$$\begin{aligned} &\nabla f(x^*)+\sum_{i=1}^{m}\lambda_j^*\nabla h_i(x^*)+\sum_{j=1}^{l}\mu_j^*\nabla g_j(x^*)=0, \\ &\mu_j^*=0,\ \forall\, j\notin\mathcal{A}(x^*) \end{aligned} \tag{21.3}$$

Complementary slackness. The condition $\mu_j^*=0$ for any $j\notin\mathcal{A}(x^*)$ can be compactly written as $\mu_j^* g_j(x^*)=0$ for all j. This is the so-called complementary slackness condition. It just states that either μ_j^* or $g_j(x^*)$ has to be 0 if x^* is a local min.

Non-negativity of μ_j^*. Using some sensitivity analysis, we can show that $\mu_j^*\geq 0$. We skip the proof here. Putting this with (21.3), we obtain the famous KKT conditions.

Theorem 21.1 (KKT conditions). *Suppose x^* is a local min and a regular point for* (21.1)*. Then there exist unique scalers λ_i^* ($i=1,2,\ldots,m$) and μ_j^* ($j=1,2,\ldots,l$) such that*

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_j^*\nabla h_i(x^*)+\sum_{j=1}^{l}\mu_j^*\nabla g_j(x^*)=0, \tag{21.4}$$

$$\mu_j^*\geq 0,\ \forall\, j=1,\ldots,l \tag{21.5}$$

$$\mu_j g_j(x^*)=0,\ \forall\, j=1,\ldots,l. \tag{21.6}$$

21.2 More Discussions

We have argued that the inequality constrained optimization problem (21.1) and the equality constrained optimization problem (21.2) behave similarly when x is in a sufficiently small neighborhood around x^*. Here we introduce a more global approach to convert (21.1) into an equality constrained optimization problem. Specifically, (21.1) is equivalent to

$$\begin{aligned} \text{minimize}\quad & f(x) \\ \text{subject to}\quad & h_i(x)=0,\ i=1,\ldots,m \\ & g_j(x)+s_j^2=0,\ j=1,\ldots,l \end{aligned} \tag{21.7}$$

Here s_j and x are both decision variables. We introduced the auxiliary variable s_j to convert an inequality constraint $g_j(x)\leq 0$ to an equality constraint $g_j(x)+s_j^2=0$. Therefore, any optimality conditions for (21.7) are also optimality conditions for (21.1). This provides one proof for the KKT conditions.

We can write everything in vector form as follows

$$h(x)=\begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x)\end{bmatrix},\ g(x)=\begin{bmatrix} g_1(x) \\ \vdots \\ g_l(x)\end{bmatrix},\ \lambda=\begin{bmatrix}\lambda_1 \\ \vdots \\ \lambda_m\end{bmatrix},\ \mu=\begin{bmatrix}\mu_1 \\ \vdots \\ \mu_l\end{bmatrix}$$

Then we can define the Lagrangian $L(x, \lambda, \mu) = f(x) + \lambda^\mathsf{T} h(x) + \mu^\mathsf{T} g(x)$, and the condition (21.4) is just $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$.

When f, g, and h are twice differentiable, there are also second-order necessary conditions for x^* being a local min of (21.1). One such necessary condition is that we have $d^\mathsf{T} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d \geq 0$ for all d satisfying $d^\mathsf{T} \nabla h_i(x^*) = 0$ or $d^\mathsf{T} g_j(x^*) = 0$ ($j \in \mathcal{A}(x^*)$).

A sufficient condition for optimality. Similar to the case where only equality constraints are involved, there are also sufficient conditions for x^* being a local min of (21.1). Suppose there exist x^*, λ^*, and μ^* satisfying (21.4), (21.5), and (21.6). Then x^* is a local min for (21.1) if the following two extra conditions are met

- $d^\mathsf{T} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d > 0$ for all d satisfying $d^\mathsf{T} \nabla h_i(x^*) = 0$ or $d^\mathsf{T} g_j(x^*) = 0$ where $j \in \mathcal{A}(x^*)$.
- $\mu_j^* > 0$ for all $j \in \mathcal{A}(x^*)$.

21.3 Examples

Now we apply KKT conditions to two examples.

1. Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & x_1^2 + x_2^2 + x_3^2 \\ \text{subject to} \quad & x_1 + x_2 + x_3 \leq -3 \end{aligned}$$

The first order condition (21.4) yields

$$\begin{aligned} 2x_1^* + \mu^* &= 0 \\ 2x_2^* + \mu^* &= 0 \\ 2x_3^* + \mu^* &= 0 \end{aligned}$$

In addition, we require $\mu^* \geq 0$ and also have the complementary slackness condition $\mu^*(x_1^* + x_2^* + x_3^* + 3) = 0$. We have two possibilities here: either $\mu^* = 0$ or $x_1^* + x_2^* + x_3^* = -3$. If $\mu^* = 0$, we have $x_1^* = x_2^* = x_3^* = 0$ and this violates the constraint $x_1^* + x_2^* + x_3^* \leq -3$. Hence we have $x_1^* + x_2^* + x_3^* = -3$. This leads to $x_1^* = x_2^* = x_3^* = -1$ and $\mu^* = 2$. Clearly $\mu^* > 0$. In addition, we have $\nabla^2_{xx} L(x^*, \mu^*) = 2I > 0$. Consequently the sufficient conditions are met and this is a local min for the given problem.

2. Given an arbitrary non-zero vector y and a positive definite matrix Q, we consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & -y^\mathsf{T} x \\ \text{subject to} \quad & x^\mathsf{T} Q x \leq 1 \end{aligned}$$

The first order condition (21.4) yields

$$-y + \mu^* Q x^* = 0$$

Based on the complementary slackness, we have $\mu^* = 0$ or $(x^*)^\mathsf{T} Q x^* = 1$. If $\mu^* = 0$, then there exists no solution for $y - \mu^* Q x^* = 0$ if $y \neq 0$. Hence we have $x^* = \frac{1}{\mu^*} Q^{-1} y$ and $(x^*)^\mathsf{T} Q x^* = 1$. This leads to $\mu^* = \sqrt{y^\mathsf{T} Q^{-1} y}$ and $x^* = \frac{1}{\sqrt{y^\mathsf{T} Q^{-1} y}} Q^{-1} y$. Notice $\mu^* > 0$ and $\nabla^2_{xx} L(x^*, \mu^*) = 2\mu^* Q > 0$. Hence x^* is a local min for the given problem.