A Preliminary Study of Krush-Kuhn-Tucker Conditions

Xianda Xu
University of Electronic Science and Technology of China
xiandaxu@std.uestc.edu.cn

Yibing Xiao
University of Electronic Science and Technology of China
xiaoyb9999@uestc.edu.cn

Abstract

Optimization is an important part of Applied Mathematics. For optimization of equality constraints, the Lagrange multiplier is commonly used to solve the problem. However, we are not familiar with how to deal with inequality constrained optimization. So, in this essay, we introduce Krush-Kuhn-Tucker conditions (KKT conditions in short) which are widely used to reach optimization with inequality constraints. This method can be sufficient and necessary for the optimal solution of convex optimization problems. We will do theory and case studies on the realizability of Krush-Kuhn-Tucker conditions.

1 Introduction

Optimization plays an important role in Applied Mathematics. It is also useful in fields like engineering and other domains using mathematics to solve practical problems. The solution of an optimization problem is one whose functional value is maximum or minimum meeting all requirements of constraints. To be general, it is the most satisfying one among all potential points, which can bring us the minimum cost or maximum productivity.

Optimization can be classified into two categories, namely optimization of equality constraints and optimization of inequality constraints. How to solve the former problem is widely known, using the Lagrange multiplier [5]. But perhaps we are not familiar with how to solve the latter problem. Actually, optimization of inequality constraints is more often seen in our daily lives than optimization of equality constraints. In the rest of this essay, we introduce a method called "Krush-Kuhn-Tucker conditions" which is commonly used to optimize with inequality constraints.

1.1 Lagrange Multiplier

Firstly, let's have a review of the Lagrange multiplier.

One optimization problem with equality conditions can be provided as:

$$\max f(x)$$
$$s.t.$$
$$G(x) = c$$

We import a function called the Lagrange function: (λ is a constant)

$$L(x, \lambda) = f(x) + \lambda(c - G(x))$$

The first-order conditions of the Lagrange function at the optimal $\overline{x}$ are:

$$\frac{\partial L(x,\lambda)}{\partial x_i} = f_i - \lambda G_i = 0, i = 1, 2, ..., n$$
$$\frac{\partial L(x,\lambda)}{\partial \lambda} = c - G(x) = 0$$

In this case, we can get the optimal solution of the optimization problem with equality conditions: $(\overline{x}, \lambda)$

1.2 Lagrange Duality

Next, let's have a look at the Lagrange duality, which is a precondition of understanding how Krush-Kuhn-Tucker conditions work.

One optimization problem with both equality conditions and inequality conditions can be provided as:

$$\begin{aligned} \min_{x \in R^n} f(x) \\ s.t. \\ c_i(x) \leq & 0, i = 1, 2, ..., k \\ h_j(x) = & 0, j = 1, 2, ..., l \end{aligned}$$

Similarly, we import a Lagrange function:

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{k} a_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)$$

Now, we provide with a new problem which is going to find the maximum of the Lagrange function defined as a new function called $\theta_P(x)$ with x fixed and both α and β acting as parameters:

$$\theta_P(x) = \max_{\alpha, \beta : \alpha_i \geq 0} L(x, \alpha, \beta)$$

Since there are two kinds of constraints in the optimization problem, we can set the latter two sums to zero, getting the maximum of this new function which is $\theta_P(x) = \max_{\alpha, \beta : \alpha_i \geq 0} [f(x)] = f(x)$. After figuring out the minimum of this function with x as a parameter, which is

$$\min_x \theta_P(x) = \min_x \max_{\alpha, \beta : \alpha_i \geq 0} L(x, \alpha, \beta) = \min_x f(x)$$

We find that the problem of getting the minimum of this function is equivalent to the original problem of getting the minimum of $f(x)$. Therefore, these two problems share the same solution and we define it as $p*$.

Afterwards, we provide the dual problem of the original problem:

$$\theta_D(\alpha, \beta) = \min_x L(x, \alpha, \beta)$$

Compared with the original problem, it takes x as a parameter and fixes α and β to determine the minimum of the Lagrange function. In the same way, we fixes x and takes α and β as parameters to seize the maximum of this function, which is the solution to the dual problem. We name it $d*$.

Now, we can put the equivalent problems of the original problem and its dual problem together:

$$\begin{aligned} \max_{\alpha,\beta:\alpha_i \geq 0} \theta_D(\alpha, \beta) &= \max_{\alpha,\beta:\alpha_i \geq 0} \min_x L(x, \alpha, \beta) \quad \text{original} \\ \min_x \theta_P(x) &= \min_x \max_{\alpha,\beta:\alpha_i \geq 0} L(x, \alpha, \beta) \quad \text{dual} \end{aligned}$$

They look symmetrical. The equivalent problem of the original problem is figuring out the maximum first and the minimum then; the equivalent problem of the dual problem is figuring out the minimum first and the maximum then. However, they have different procedures in the sequence of fixing x and fixing α and β.

[1] proves that the optimal solution of the dual problem is the maximization of the lower bound of the optimal solution of the original problem:

$$\begin{aligned} d* &= \max_{\alpha,\beta:\alpha_i \geq 0} \min_x L(x, \alpha, \beta) \\ &\leq \min_x \max_{\alpha,\beta:\alpha_i \geq 0} = p* \end{aligned}$$

[2] provides a *Slater's condition*, which is a sufficient condition to make $p* = d*$ in a convex problem so that in this case, we can obtain the optimal

solution of the original problem by solving the optimal solution of the dual problem:

$$\exists\ x_0 \in D : f_i(x_0) < 0, i = 1, ..., k\ ; h_j(x_0) = 0, j = 1, ..., l$$

2 General Krush-Kuhn-Tucker Conditions

There is a close connection between Krush-Kuhn-Tucker conditions and the Lagrange duality since one precondition of the former is $p* = d*$, which means that under the strong duality condition, the original problem shares one common solution with its dual problem. Therefore, the use of Krush-Kuhn-Tucker conditions is to obtain the optimal solution with constraints of both the original and the dual problem.

2.1 Formula

In the general case, though it requires $p* = d*$, the precondition does not necessarily meet the *Slater's condition*. The optimization problem and two constraints are:

$$\begin{aligned} &\min f(x) \\ &s.t. \\ c_i(x) &\leq 0, i = 1, 2, ..., k \\ h_j(x) &= 0, j = 1, 2, ..., l \end{aligned}$$

In the same way, we import the Lagrange function (both u_i and v_j are constants):

$$L(x, u_i, v_j) = f(x) + \sum_{i=1}^{k} u_i c_i(x) + \sum_{j=1}^{l} v_j l_j(x)$$

The followings are Krush-Kuhn-Tucker conditions that the optimal solution $\overline{x}$ needs to meet:

$$\partial f(x) + \sum_{i=1}^{k} u_i \partial c_i(x) + \sum_{j=1}^{l} v_j \partial l_j(x) = 0 \qquad \text{(1) Stationarity}$$

$$u_i \cdot c_i(x) = 0 \; for \; all \; is \qquad \text{(2) Complementary Slackness}$$

$$c_i(x) \leq 0; \; h_j(x) = 0 \; for \; all \; is \; and \; js \qquad \text{(3) Primal Feasibility}$$

$$u_i \geq 0 \; for \; all \; is \qquad \text{(4)Dual Feasibility}$$

2.2 Discussion

Suppose that x, α and β are the solution to the optimization problem using Krush-Kuhn-Tucker conditions. In the general case, however, they are not necessarily equal to the optimal solution, which is $\overline{x}$, $\overline{\alpha}$ and $\overline{\beta}$. This is because the *Stationarity* condition here is only a sufficient condition.

3 Krush-Kuhn-Tucker Conditions in Convex Optimization

In the general case, Krush-Kuhn-Tucker conditions are just sufficient. However, in the convex case, with the $Slater'scondition$, they are strengthened to be sufficient and necessary.

The biggest difference between a general optimization problem and a convex optimization problem is that all functions in the latter are convex. In a convex optimization problem, the local optimal solution is its global optimal solution. Therefore, convex optimization is useful in solving problems like linear regression, norm approximation and parameter estimation.

3.1 Formula

The optimization problem and two constraints are the same as before:

$$
\begin{aligned}
&\min f(x)\\
&s.t.\\
c_i(x) &\leq 0, i = 1, 2, ..., k\\
h_j(x) &= 0, j = 1, 2, ..., l
\end{aligned}
$$

However, preconditions have changed to:

$$
\begin{aligned}
f(x) \text{ and } c_i(x) \text{ are convex} \quad &(1)\\
h_j \text{ is a affine function,usually a linear function} \quad &(2)\\
\text{The } \textit{Slater's condition} \text{ is met} \quad &(3)
\end{aligned}
$$

In the same way, we import the Lagrange function (both u_i and v_j are constants):

$$L(x, u_i, v_j) = f(x) + \sum_{i=1}^{k} u_i c_i(x) + \sum_{j=1}^{l} v_j l_j(x)$$

Krush-Kuhn-Tucker conditions are the same but have strengthened to be sufficient and necessary:

$$
\begin{aligned}
\partial f(x) + \sum_{i=1}^{k} u_i \partial c_i(x) + \sum_{j=1}^{l} v_j \partial l_j(x) = 0 \quad &\text{(1) Stationarity}\\
u_i \cdot c_i(x) = 0 \ for\ all\ is \quad &\text{(2) Complementary Slackness}\\
c_i(x) \leq 0;\ h_j(x) = 0 \ for\ all\ is\ and\ js \quad &\text{(3) Primal Feasibility}\\
u_i \geq 0 \ for\ all\ is \quad &\text{(4)Dual Feasibility}
\end{aligned}
$$

3.2 Proof

3.2.1 Necessity

Assume that $\overline{x}$, $\overline{\alpha}$ and $\overline{\beta}$ are the common optimal solution of both the original and its dual problem. At the same time, all preconditions are met.

The preconditions make $p* = d*$, so:

$$f(\overline{x}) = g(\overline{u}, \overline{v}) \tag{4}$$

$$= \min L(x, \overline{u}, \overline{v} = \min f(x) + \sum_{i=1}^{k} \overline{u_i} c_i(x) + \sum_{j=1}^{l} \overline{v_j} l_j(x) \tag{5}$$

$$\leq f(\overline{x}) + \sum_{i=1}^{k} \overline{u_i} c_i(\overline{x}) + \sum_{j=1}^{l} \overline{v_j} l_j(\overline{x}) \tag{6}$$

$$\leq f(\overline{x}) \tag{7}$$

$$= f(\overline{x}) \tag{8}$$

(8) indicates that all inequalities can be set to equalities.

The transformation in (6) proves the *Stationarity* condition.

The transformation in (7) proves the *Complementary Slackness* condition.

Both the *Primal Feasibility* condition and the *Dual Feasibility* condition are provided as constraints.

Therefore, the necessity is proved.

3.2.2 Sufficiency

Assume that $\overline{x}$, $\overline{\alpha}$ and $\overline{\beta}$ meet Krush-Kuhn-Tucker conditions.

$$
\begin{aligned}
\overline{g} &= g(\overline{u}, \overline{v}) && (9)\\
&= \min f(x) + \sum_{i=1}^{k} \overline{u_i} c_i(\overline{x}) + \sum_{j=1}^{l} \overline{v_j} l_j(\overline{x}) && (10)\\
&= f(\overline{x}) + \sum_{i=1}^{k} \overline{u_i} c_i(\overline{x}) + \sum_{j=1}^{l} \overline{v_j} l_j(\overline{x}) && (11)\\
&= f(\overline{x}) && (12)
\end{aligned}
$$

From (9) to (10), the property of convex functions is used.

From (10) to (11), the *Stationarity* condition is used.

From (11) to (12), the *Complementary Slackness* condition is used.

So, $\overline{x}$, $\overline{\alpha}$ and $\overline{\beta}$ are the optimal solution of both the original problem and its dual problem.

Therefore, sufficiency is proved.

3.3 Example

Using Krush-Kuhn-Tucker conditions, we can solve many real-life convex optimization problems. For example, [3] builds a *SVM classifier* in machine learning based on the use of Krush-Kuhn-Tucker conditions. Another example is dealing with *Quasilinear preferences* in economics, as [4] provides.

The problem can be extracted as:

$$\max u(x_1, x_2) = x_2 + a\ln(x_1)$$
$$s.t.$$
$$p_1x_1 + p_2x_2 \leq I$$
$$x_1, x_2 \geq 0$$

It is a convex optimization problem with all functions convex and meeting the *Slater's condition.*

The Lagrange function is: $L(x_1, x_2, \lambda) = x_2 + a\ln(x_1) + \lambda(I - p_1x_1 - p_2x_2)$

Case 1: $x_1 > 0, x_2 = 0, \lambda > 0$

$$\frac{\partial L}{\partial x_1} = \frac{a}{x_1} - \lambda p_1 = 0$$
$$\frac{\partial L}{\partial x_2} = 1 - \lambda p_2 = 0$$
$$\frac{\partial L}{\lambda} = I - p_1x_1 = 0$$

We get: $x_1 = \frac{I}{p_1},\ x_2 = 0,\ \lambda = \frac{\alpha}{I},\ I \leq \alpha p_2$

Case 2: $x_1 > 0, x_2 > 0, \lambda > 0$

$$\frac{\partial L}{\partial x_1} = \frac{a}{x_1} - \lambda p_1 = 0$$
$$\frac{\partial L}{\partial x_2} = 1 - \lambda p_2 = 0$$

We get: $x_1 = \frac{ap_2}{p_1},\ x_2 = \frac{I - ap_2}{p_2},\ \lambda = \frac{1}{p_2},\ I > \alpha p_2$

Quasilinear preferences has its practical meaning. $u(x_1, x_2)$ is called utility function. x_1 and x_2 are the amounts of necessities and extravagances. p_1 and p_2 are the unit price of necessities and extravagances. I is the gross income. Therefore, we have to find the optimal solution of the utility function with the constraint that the total consumption does not exceed the income.

In case 1, the gross income is not very high. It means that we have to invest all the money into necessities to solve the problem of fooding and clothing. In case 2, the gross income is comparatively high. It means that we can use part of it to buy some extravagances but the total consumption of necessities stays the same.

4 Summary

This essay takes a review of the Lagrange multiplier and introduces the Lagrange duality, based on which general Krush-Kuhn-Tucker conditions are explored. However, in the general case, Krush-Kuhn-Tucker conditions are only sufficient, so we extend them to the convex case where they become both sufficient and necessary. At last, we introduce a classical application of Krush-Kuhn-Tucker conditions in solving *Quasilinear preferences* in economics.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] Stephan Dempe. Directional differentiability of optimal solutions under slater's condition. *Mathematical Programming*, 59(1-3):49–69, 1993.

[3] Thorsten Joachims. Making large-scale svm learning practical. Technical report, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten , 1998.

[4] Q.Feng. A brief lecture notes on optimization for microeconomic analysis. *Fudan University*, 2003.

[5] R Tyrrell Rockafellar. Lagrange multipliers and optimality. *SIAM review*, 35(2):183–238, 1993.