对于KKT条件及其在凸优化中应用的初探

徐贤达（2016060601018）

指导老师：肖义彬

【摘要】：最优化是应用数学中非常重要的一个问题。拉格朗日乘子法被广泛地用来解决等式约束下的优化问题，而对于不等式约束下的优化问题，我们比较陌生。所以，本文主要介绍一种解决不等式约束下优化问题的常用方法，即卡罗需-库恩-塔克条件（Karush-Kuhn-Tucker conditions，简称KKT条件）。这种方法在凸优化问题中能够成为最优解的充要条件，因此得到广泛应用。本文从理论入手，结合案例，论证了KKT条件在凸优化问题中的可实现性，为KKT在机器学习等领域的应用打下了基础。

【关键词】：拉格朗日乘子法、卡罗需-库恩-塔克条件、凸优化

1.引入

毋庸置疑，最优化是应用数学中非常重要的一个内容。它在工程学、计算机科学、金融学等领域也帮助人们解决一些实际的问题。最优化问题的解，是满足所有约束条件下，能使它对应的函数达到最大值或者最小值的一个解。一般地，它是在所有可能点中，能给我们带来最小成本或者是最大收益的那个点。

最优化能被分为两类，即等式约束下的最优化和不等式约束下的最优化。前者的解决我们非常熟悉，利用拉格朗日乘子法。但是我们对于后者的解决以及卡罗需-库恩-塔克条件，我们不太了解。事实上，在我们实际生活中遇到的优化问题大多都是不等式约束下的最优化问题，所以，我们有必要对卡罗需-库恩-塔克条件有所了解。

2.拉格朗日乘子法

在介绍卡罗需-库恩-塔克条件之前，我们首先来回顾一下拉格朗日乘子法，它对我们理解卡罗需-库恩-塔克条件具有启发性。最优化问题的形式可以写成如下：

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

我们引入拉格朗日函数：（λ是一个常数）

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

最优解$\bar{x}$在拉格朗日函数中的一阶条件是：

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

$$\frac{\partial L(x,\lambda)}{\partial \lambda} = c - G(x) = 0$$

通过以上方法，我们得到最优解：$(\bar{x}, \bar{\lambda})$

3. 拉格朗日对偶

拉格朗日对偶是理解卡罗需-库恩-塔克条件的一个重要概念。

我们引入原始优化问题如下，有等式约束，也有不等式约束

$$\min_{x \in R^n} f(x)$$
$$s.t.$$
$$c_i(x) \le 0, i = 1,2, \dots, k$$
$$h_j(x) = 0, j = 1,2, \dots, l$$

同样地，我们构造拉格朗日函数：

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

我们现在再引入一个新问题：

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

我们固定x，以α和β为参数，来确定拉格朗日函数的最大值，令为新的函数$\theta_P(x)$

因为有两类等式和不等式的约束，我们可以让后面的两个求和为 0，得到这个新函数的最大值就是 f(x)：

$$\theta_P(x) = \max_{\alpha, \beta: \alpha_i \ge 0} [f(x)] = f(x)$$

然后，我们以x为参数取这个函数的最小值，发现，取这个函数最小值的问题与我们原始问题取f(x)最小值的问题等价。所以，这两个问题共解。

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

我们不妨将这个解令为 p^*

我们定义原始问题的对偶问题：

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

相比于原始问题，它固定α和β，以x为参数，来确定拉格朗日函数的最小值，令为一个新函数。同样地，我们再固定x，以α和β为参数，取这个函数的最大值，得到对偶问题解 d^*

我们将原始问题和对偶问题的等价问题放在一起，如下所示：

（原始问题） $\max_{\alpha, \beta: \alpha_i \ge 0} \theta_D(\alpha, \beta) = \max_{\alpha, \beta: \alpha_i \ge 0} \min_x L(x, \alpha, \beta)$

（对偶问题） $\min_x \theta_P(x) = \min_x \max_{\alpha, \beta: \alpha_i \ge 0} L(x, \alpha, \beta)$

它们形式上非常对称，原始问题等价问题是先求最大，再求最小；对偶问题等价问题是先求最小，再求最大。但是，它们过程却是不同的。原始问题是先固定x，优化出α和β，最后优化出最优x。对偶问题是先固定α和β，优化出x，最后确定最优α和β。

可以证明，原始问题和对偶问题的解有如下的关系：

$$d^* = \max_{\alpha, \beta: \alpha_i \ge 0} \min_x L(x, \alpha, \beta)$$
$$\le \min_x \max_{\alpha, \beta: \alpha_i \ge 0} L(x, \alpha, \beta) = p^*$$

对偶问题的最优解，是原始问题最优解下界的最大化。

我们假设x^*和α^*,β^*为原始问题和对偶问题的可行解。考虑一种特殊的情况，$p^* = d^*$，那么，此时，x^*和α^*,β^*就上升为原始问题和对偶问题的可行解。

所以，如果 $p^* = d^*$，我们就可以通过求解对偶问题的最优解来获得原始问题的最优解。

至于如何使得 $p^* = d^*$，有一个斯莱特条件（Slater conditions），它是一个充分条件：若存在x，使得不等式约束$c_i(x) < 0$恒成立，并且原问题是凸问题，则有$p^* = d^*$。

4. 一般情况下的卡罗需-库恩-塔克条件

卡罗需-库恩-塔克条件与上述拉格朗日对偶有着密切的联系，因为它的一个前提条件就是 $p^* = d^*$，即在强对偶条件下原始问题和对偶问题有着一组解。按照本人的理解，卡罗需-库恩-塔克条件的使用就是原始问题和对偶问题共同约束下求解最优解的一个过程。

我们下面先来研究一般情况下的卡罗需-库恩-塔克条件。之所以是一般情况，因为它对$f(x)$没有要求，前提条件只是 $p^* = d^*$，不一定要通过满足斯莱特条件来获得。定义一个求最小值的最优化问题并且x满足两种约束：

$$\min f(x)$$
$$s.t.$$
$$h_i(x) \ll 0, i = 1,2, \dots \dots, m$$
$$l_j(x) = 0, j = 1,2, \dots \dots, r$$

同样地，我们引入拉格朗日函数：（u_i和v_j是常数）

$$L(x, u_i, v_j) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x)$$

如下便是一般情况下最优解$\bar{x}$需满足的卡罗需-库恩-塔克条件：

(1) 静态条件（Stationarity）

$$\partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j \partial l_j(x) = 0$$

(2) 互补松弛（Complementary Slackness）

$$u_i \cdot h_i(x) = 0\ for\ all\ is$$

(3) 原始问题可行域（Primal Feasibility）

$$h_i(x) \le 0, l_j(x) = 0\ for\ all\ is\ and\ js$$

(4) 对偶问题可行域（Dual Feasibility）

$$u_i \ge 0\ for\ all\ is$$

但是，正如拉格朗日乘子法，卡罗需-库恩-塔克条件求取的结果并不一定是最优解情况下的$\bar{x}$、$\bar{\alpha}$和$\bar{\beta}$。因为一般情况下，卡罗需-库恩-塔克条件中的静态条件只是充分条件，所以卡罗需-库恩-塔克条件求取出的x、α和β是最优解情况下的$\bar{x}$、$\bar{\alpha}$和$\bar{\beta}$的充分条件。

5. 凸优化情况下的卡罗需-库恩-塔克条件

在一般情况下，卡罗需-库恩-塔克条件只是一个充分条件。但是，在凸优化情况下，这个条件就被加强成了充要条件。

凸优化问题和一般优化问题最大的不同是，它所有的函数都是凸函数。凸优化问题中的局部最优解是它的全局最优解，所以，它在解决线性回归、范数逼近、参数估计等一些问题

上有着优势。

凸优化的问题形式可以写成如下，和一般优化问题一致：

$$\min f(x)$$
$$s.t.$$
$$h_i(x) \ll 0, i = 1,2, \dots \dots, m$$
$$l_j(x) = 0, j = 1,2, \dots \dots, r$$

但是前提条件变化了：

（1）f(x) 和 $h_i(x)$必须是凸函数

（2）$l_j(x)$ 必须是仿射函数，一般是线性函数

（3）满足斯莱特条件

同样地，我们引入拉格朗日函数：（u_i和v_j是常数）

$$L(x, u_i, v_j) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x)$$

最优化情况下最优解$\bar{x}$需满足的卡罗需-库恩-塔克条件不变，但已上升为充要条件：

(1) 静态条件（Stationarity）

$$\partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j \partial l_j(x) = 0$$

(2) 互补松弛（Complementary Slackness）

$$u_i \cdot h_i(x) = 0 \ for \ all \ is$$

(3) 原始问题可行域（Primal Feasibility）

$$h_i(x) \leq 0, l_j(x) = 0 \ for \ all \ is \ and \ js$$

(4) 对偶问题可行域（Dual Feasibility）

$$u_i \geq 0 \ for \ all \ is$$

下面给出证明：

必要性：

如果$\bar{x}$、α和$\bar{\beta}$是原始问题和对偶问题的最优解且前提条件成立，

因为前提条件中对偶间距为 0，即 $p^* = d^*$，有

$$
\begin{aligned}
f(\bar{x}) &= g(\bar{u}, \bar{v}) \\
&= \min L(x, \bar{u}, \bar{v}) = \min f(x) + \textstyle\sum_{i=1}^{m} \bar{u}_i h_i(x) + \sum_{j=1}^{r} \bar{v}_j l_j(x) \\
&\leq f(\bar{x}) + \textstyle\sum_{i=1}^{m} \bar{u}_i h_i(\bar{x}) + \sum_{j=1}^{r} \bar{v}_j l_j(\bar{x}) \\
&\leq f(\bar{x})
\end{aligned}
$$

所以，不等号可以变成等号。

第一个不等号变成了等号，说明了静态条件成立。

第二个不等号变成了等号，说明了互补松弛条件成立。

原始问题和对偶问题可行域为问题中提供，显然成立。

综上，必要性条件成立

充分性：

如果如果$\bar{x}$、$\bar{\alpha}$和$\bar{\beta}$满足卡罗需-库恩-塔克条件，

根据对偶函数的定义，

$$\bar{g} = g(\bar{u}, \bar{v})$$

$$= \min f(x) + \sum_{i=1}^{m} \bar{u}_i h_i(x) + \sum_{j=1}^{r} \bar{v}_j l_j(x)$$

根据静态条件：$\partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j \partial l_j(x) = 0$

$$= f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i h_i(\bar{x}) + \sum_{j=1}^{r} \bar{v}_j l_j(\bar{x})$$

根据互补松弛条件：$u_i \cdot h_i(x) = 0 \; for\ all\ is$

$$= f(\bar{x})$$

所以，$\bar{x}$、$\bar{\alpha}$和$\bar{\beta}$是原始问题和对偶问题的最优解。

综上，充分性条件成立。

6. 卡罗需-库恩-塔克条件在凸优化条件

利用卡罗需-库恩-塔克条件，我们可以解决许多凸优化问题。下面就给出一个简单地例子，说明卡罗需-库恩-塔克条件的便捷性。我们利用它来解决经济学中拟线性偏好问题。

问题如下：

$$\max u(x_1, x_2) = x_2 + a\ln(x_1)$$
$$s.t.$$
$$p_1 x_1 + p_2 x_2 \leq I$$
$$x_1, x_2 \geq 0$$

这也是一个凸优化问题，

问题中都是凸函数且问题满足斯莱特条件。

构造拉格朗日函数：$L(x_1, x_2, \lambda) = x_2 + a\ln(x_1) + \lambda(I - p_1 x_1 - p_2 x_2)$

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

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

我们得到：$x_1 = {}^{I}/_{p_1}$ ，$x_2 = 0$，$\lambda = {}^{a}/_{I}$ and $I \leq a p_2$

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

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

$$x_1 = {}^{a p_2}/_{p_1} \,, x_2 = {}^{I - a p_2}/_{p_2} \,, \lambda = {}^{1}/_{p_2} \text{ and } I > a p_2$$

我们得到

拟线性偏好具有它的实际意义。

我们需要最大化的函数称作效用函数，x_1和x_2为必需品和奢侈品的消费量，p_1和p_2为必需品和奢侈品的单价，I为总收入。所以，我们就需要在必需品、奢侈品消费总额不超过总收入的情况下找到最大效用下的最优解。

第一种情况下，总收入不是很高，所以应该把总收入全部用来解决温饱问题。第二种情况下，总收入超过一定标准，可以用一部分收入用来购买奢侈品，但是必需品的需求应该保持不变。

7. 小结

本文回顾了拉格朗日乘子法，介绍了拉格朗日对偶性，并且利用此探究了一般函数下的卡罗需-库恩-塔克条件。但是，一般情况下的卡罗需-库恩-塔克条件只是充分条件，所以我们将其推广到凸优化情况下，这时，充分条件被强化成了充要条件。最后，本文简单介绍了卡罗需-库恩-塔克条件解决凸优化问题的一个经济学应用----拟线性问题。

优化问题是应用数学中比较重要的一类问题。本人参考总结了一些资料，将一些零散的论证整合，思考补充了一些缺失或不完善的内容。因水平有限，欢迎读者提供指导性意见！

参考文献

[1] 史蒂夫·博伊德，列文·范德伯格，凸优化，斯坦福大学出版社，2003

[2] 冯曲，微观经济学的优化分析，复旦大学出版社，2003

[3] 伯特·塞卡斯，凸分析与优化，清华大学出版社，2006

[4] 陈宝林，最优化理论与算法，清华大学出版社，2005