拉格朗日乘子法——带不等式约束项的函数优化

Dezeming Family

2021 年 9 月 13 日

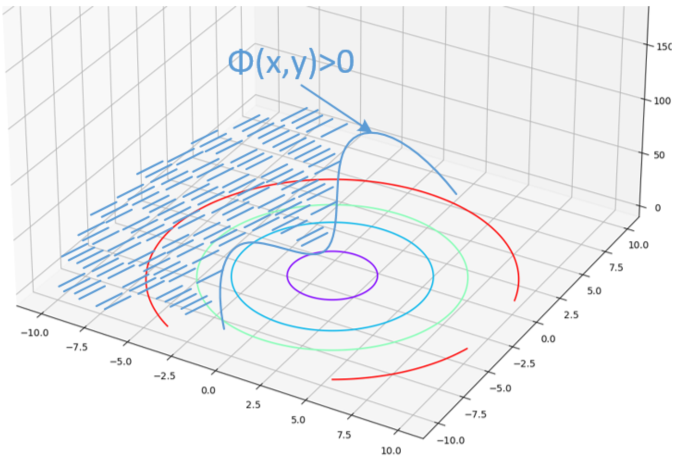
DezemingFamily 系列书和小册子因为是电子书，所以可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列书，可以从我们的网站 [https://dezeming.top/] 找到最新版。对书的内容建议和出现的错误欢迎在网站留言。

目录

一 不等式约束

在《拉格朗日乘子法——带等式约束项的函数优化》中我们可以看到，有等式约束的条件其实就是对函数的变化维度进行了限定，比如一个二元函数 $f(x_1, x_2)$，限定条件是 $g(x_1, g_2) = 0$，那么函数 g 其实就是一条曲线，而二元函数值就只能取曲线上的值。

但当这个条件变成不等式的时候，情况就变得复杂了，例如 $g(x_1, x_2) > 0$ 就不再是一条曲线，而可能是一个曲面，f 可以在这个曲面上任意取值。



现在带约束的函数极值有很多种可能，有可能函数能取到的极值还是在切线上，也有可能函数的极值点在不等式约束内。在损失函数优化中，我们希望损失函数能取到尽可能小的值，因此我们可以先求函数的所有极小值点，然后判断其是否在不等式约束内部，如果在，则说明这就是函数最小的点，如果不在，我们就可以认定函数能取到的最小值在约束项边界处，就可以使用等式约束条件下的拉格朗日乘子法。

二　KKT 条件

KKT 条件（Karush-Kuhn-Tucker conditions）是关于带不等式约束的拉格朗日乘子法得到全局最小值的必要条件，在讲解 KKT 条件时，我们先进行一些定义。

考虑最优化问题：

$$min\Big(f(\boldsymbol{x})\Big) \tag{二.1}$$

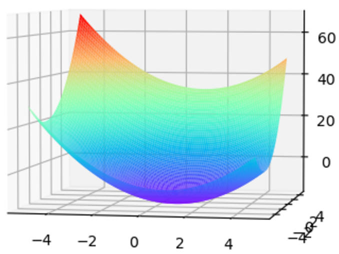
$$s.t.\quad g(\boldsymbol{x}) \leq 0 \tag{二.2}$$

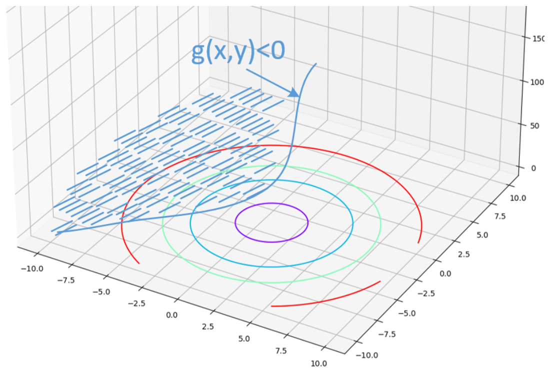
注意函数 g 的形式，如果我们的约束项是 $g(\boldsymbol{x}) \geq 0$，我们就改为 $\phi(\boldsymbol{x}) = -g(\boldsymbol{x}) \leq 0$ 就可以了。

假如 $\boldsymbol{x}^*$ 是最优解，它可能会出现两种情况，第一种是 $g(\boldsymbol{x}^*) < 0$，这时最优解在约束条件内部，约束条件不起作用；第二种是 $g(\boldsymbol{x}^*) = 0$，也就是说最优解在边界上，约束条件有效。

我们以二维函数为例，我们知道极值解一定在等高线和约束项边界线的切线上，但是我们需要保证切线处必须满足一定的条件，否则就可能出现切线处是极大值以及函数可以减小到无穷小等现象。

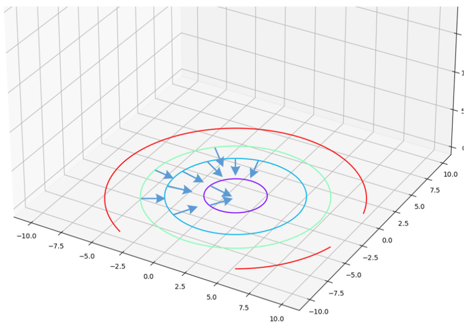
假如二维函数如下，可以看到最小值点会取在约束项边界上：

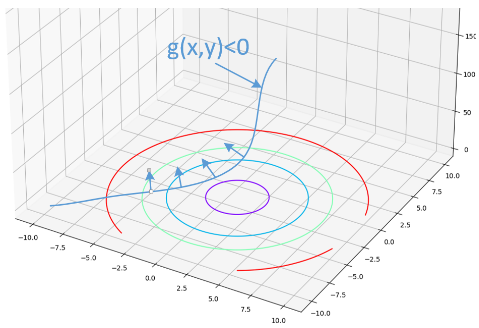




关于多元函数的梯度和切线详细讲解参见 DezemingFamily 的《函数的切线与梯度》章节。

对于函数 $a = h(x, y)$，其梯度 ∇h 指向函数值变大的一侧，即对于上图的函数 f，在等高线上绘制出多个点的函数值减小的方向（梯度的反方向），如下图左；对于约束项 $g(x)$，我们知道其内部值一定小于其边界值（因为内部值都小于 0，边界值为 0），所以如果构造一个约束项函数 $b = g(x, y)$，在 $g(x, y) = 0$ 处的梯度反方向（$g(x, y)$ 减小的方向）如下图右。





现在就很明确了，最优解点处，约束项的梯度方向需要跟函数的梯度方向恰好相反，这就是 KKT 条件中最重要的部分。我们现在把 KKT 条件列一下：

$$\nabla L(\boldsymbol{x}, \lambda) = \nabla f + \lambda \nabla g = \boldsymbol{0} \tag{二.3}$$

$$g(\boldsymbol{x}) \leq 0 \tag{二.4}$$

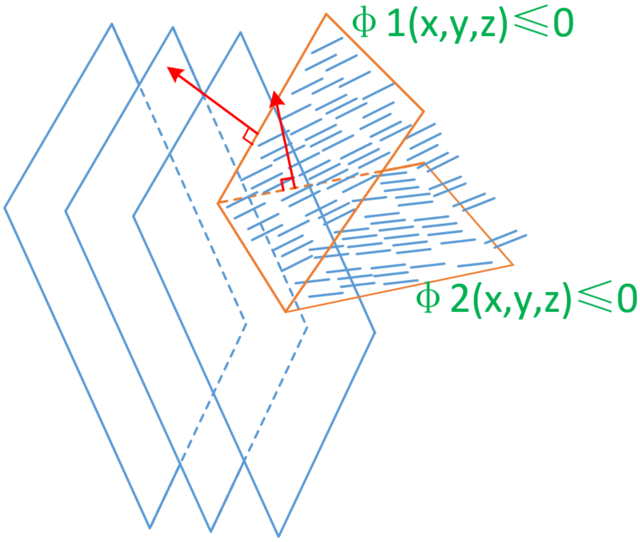
$$\lambda \geq 0 \tag{二.5}$$

$$\lambda g(\boldsymbol{x}) = 0 \tag{二.6}$$

如果我们的目标是最大化函数 f，那么根据对偶可行性原则，就可以设 $\phi(\boldsymbol{x}) = -g(\boldsymbol{x}) \geq 0$；或者令 $\lambda \leq 0$ 也是可以的。

三　多个不等式约束

其实可以类比只有等式约束项的函数优化，我们想象得到，我们还是要保证约束项的梯度组合以后为函数梯度反方向，我们观察下图，设待优化函数为 $s = f(x, y, z)$：



两个约束项的公共区域是上图中的楔形区域，假如我们的函数极小值点不在楔形区域内，则最优解就在两个区域的相交线上。在相交线上，得到的解就是等式约束的解，但还需要有一些条件。

条件一：在这个微小的局部区域上，交线就是函数 $s = f(x, y, z)$ 等值面上的一条切线（否则该点就不是极值点，沿着交线走函数值还可以变小）。

条件二：同时约束项的 $t_1 = \phi_1(x, y, z)$ 在交线处的梯度方向是朝楔形区域外部的（楔形区域内部都小于 0，边界等于 0，梯度方向朝函数增大的方向），$t_2 = \phi_2(x, y, z)$ 也是同理。约束项梯度指向的方向是函数 s 值减小的方向，而函数 $f(x, y, z)$ 在该交线处的梯度是指向函数 s 值增大的方向。

我们给出优化目标：

$$min\Big(f(\boldsymbol{x})\Big) \tag{三.1}$$

$$s.t. \quad g_j(\boldsymbol{x}) = 0, \quad j = 1, ..., m \tag{三.2}$$

$$h_k(\boldsymbol{x}) \leq 0, \quad k = 1, ..., p \tag{三.3}$$

也就是说，需要保证：

$$\nabla f = \sum_{j=1}^{m} \lambda_j \nabla g_j(\boldsymbol{x}) + (-1) \sum_{k=1}^{p} \mu_k \nabla g_k(\boldsymbol{x}) \tag{三.4}$$

$$\mu_k \geq 0, \quad k = 1, ..., p \tag{三.5}$$

构建的拉格朗日函数为：

$$L(\boldsymbol{x}, \{\lambda_j\}\{\mu_k\}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}) + \sum_{k=1}^{p} \mu_k h_k(\boldsymbol{x}) \tag{三.6}$$

多不等式约束的 KKT 条件为：

$$\nabla L = \mathbf{0} \tag{三.7}$$

$$g_j(\boldsymbol{x}) = 0, \quad j = 1, ..., m \tag{三.8}$$

$$h_k(\boldsymbol{x}) \leq 0 \tag{三.9}$$

$$\mu_k \geq 0 \tag{三.10}$$

$$\mu_k h_k(\boldsymbol{x}) = 0, \quad k = 1, ..., p \tag{三.11}$$

KKT 条件需要满足一定的约束限定条件或者正则条件，例如线性约束限定（LCQ：g_j 和 h_k 都是仿射函数，即最高次数为 1 的多项式函数），如果满足 LCQ，就不用再满足其他条件了。还有很多条件都与运筹学、函数凸优化有关，优化分析专家们不断地扩充相关的理论，大家有兴趣可以去 [3] 自行搜索。

参考文献

[1] https://www.cnblogs.com/liaohuiqiang/p/7805954.html

[2] https://zhuanlan.zhihu.com/p/38163970

[3] https://en.wikipedia.org/wiki/Karush-Kuhn-Tucker_conditions#Regularity_conditions