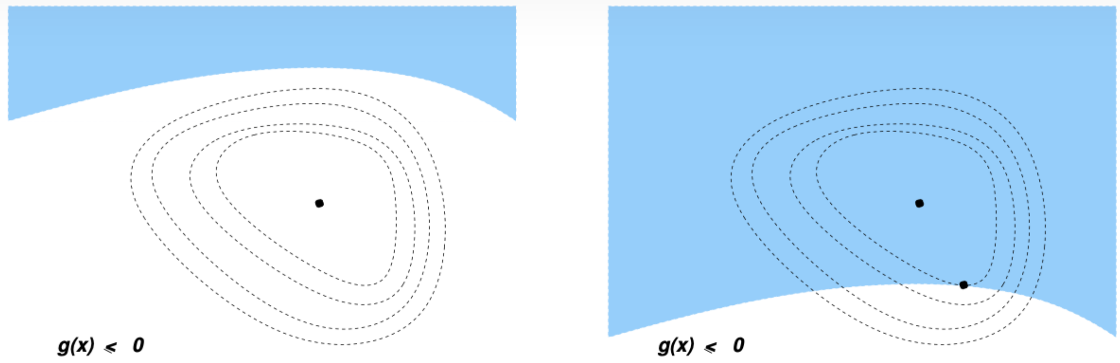



# 拉格朗日乘子法与KKT条件

Mars



**前言：本文主要讨论约束优化问题。首先介绍等式约束，然后介绍拉格朗日乘子法如何解决等式约束，再介绍不等式约束和KKT条件。**

## 目录：

(一)拉格朗日乘子法

1.1 等式约束

1.2 拉格朗日乘子法（Lagrange multipliers）

1.3.等式约束示例

(二)KKT条件( Karush-Kuhn-Tucker condition)

2.1 导出KKT条件

2.2.不等式约束示例

## (一) . 拉格朗日乘子法(Lagrange multipliers)

### 1.1 等式约束

假设 $\boldsymbol{x}$ 为 $d$ 维向量，欲想找到 $\boldsymbol{x}$ 的某个取值 $\boldsymbol{x}^*$ ，使目标函数 $f(\boldsymbol{x})$ 最小并且同时满足等式 $g(\boldsymbol{x})=0$ 的约束条件，即：

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$s.t\ \ g(\boldsymbol{x})=0$$

> 问题中 $\boldsymbol{x}$ 为 $d$ 维向量，可能会有些抽象。如果从几何的角度去思考，那么就是在由 $g(\boldsymbol{x})=0$ 所确定的 $d-1$ 维平面（超平面）找到一个使得 $f(\boldsymbol{x})$ 取得最小值的点。

### 1.2 拉格朗日乘子法（Lagrange multipliers）

如何找到这样的点？这就用到拉格朗日乘子法。

拉格朗日乘数法用于求解多元函数在一组约束条件(注意这里的约束条件时等式约束条件，如果约束条件出现了不等式，那么就会用到下面的KKT条件)下的极值问题。**如果我们的目标函数存在 $d$ 个变量并且我们有 $k$ 个等式约束条件，那么我们可以借助拉格朗日乘数法将原问题转化为 $d+k$ 个变量的无约束优化问题。**

具体做法:

1.引入拉格朗日乘子 $\lambda$ ，构造拉格朗日函数:

$$L(\boldsymbol{x},\lambda)=f(\boldsymbol{x})\ +\ \lambda g(\boldsymbol{x})$$

2.对 $\boldsymbol{x}$ 分别求偏导，并 $f(\boldsymbol{x})$ 令其为零得到我们要求的极值点:

$$\frac{\partial L(\boldsymbol{x},\lambda)}{\partial \boldsymbol{x}}=0$$

> 这里令其偏数为零的原因一是因为我们可以看做一般函数求极值的方法;二是令其偏数为零得到的结果与原问题相吻合,如对 $\lambda$ 求偏导数:
>
> $$\frac{\partial L(\boldsymbol{x},\lambda)}{\partial\lambda}=g(\boldsymbol{x})=0$$

我们通过拉格朗日乘子法，构造了拉格朗日函数，将其转化为了关于拉格朗日函数的无约束优化问题。虽然通过上面的做法的确可以正确地求得极值和极值点，但是如果第一次看到这样的求极值的方法，难免会问为什么要这样构造拉格朗日函数或者说我们为什么要把约束项加入到我们的目标函数中?

> 首先，拉格朗日函数和我们的在约束条件 $g(\boldsymbol{x})$ 下的目标函数 $f(\boldsymbol{x})$ 二者是等价的。这一点我们可以将符合约束条件的点 $\boldsymbol{x}$ 代入到拉格朗日函数中，第二项为零，得到的结果和原目标函数一致;
> 第二点，解释起来有点突兀 。这样的做法其实有些类似于我们求解常系数线性微分方程时，如果我们知道了通解 $Ce^{\lambda t}$ 代入到微分方程中的过程，也就是拉格朗日函数类似于“通解”这个概念，我们想要求得其实是一个“特解”（极值）。

### 1.3.等式约束示例

$$\begin{aligned}&\underset{\boldsymbol{x}}{min}\ J=x_1^2+x_2^2+x_3^2+x_4^2\\&s.t\ \ x_1+x_2+x_3+x_4=1\end{aligned}$$

$Solution:$

$$\overline{J}=x_1^2+x_2^2+x_3^2+x_4^2-\lambda(1-x_1-x_2-x_3-x_4)$$

求偏导数

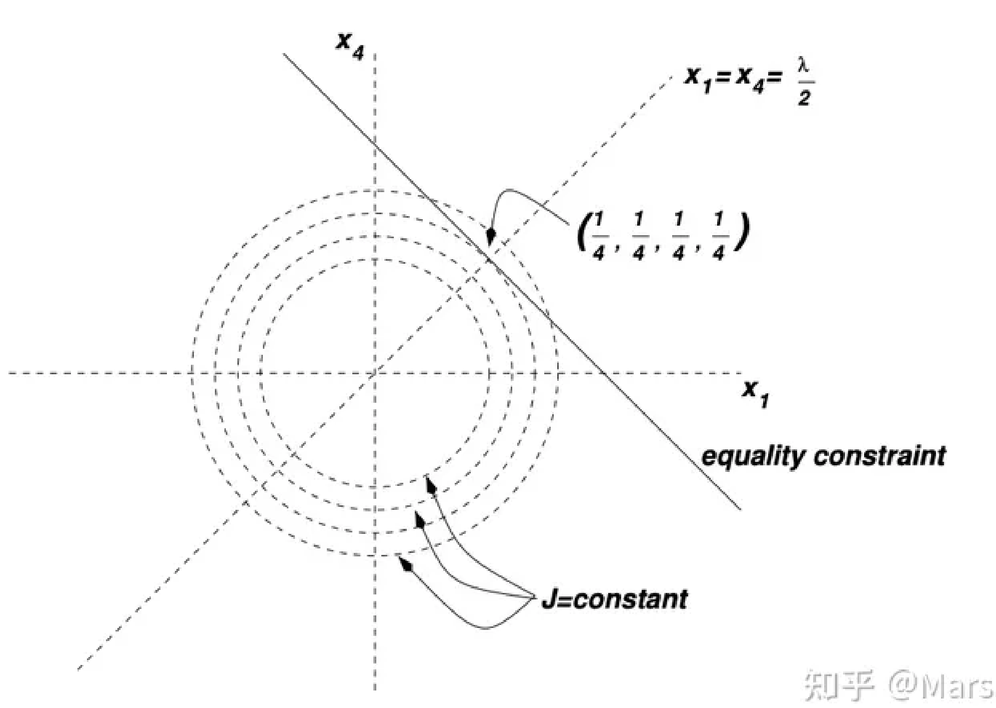
$$\frac{\partial\overline{J}}{\partial\boldsymbol{x}}=\begin{pmatrix}2x_1-\lambda\\2x_2-\lambda\\2x_3-\lambda\\2x_4-\lambda\end{pmatrix}=0$$

联立

$$x_1+x_2+x_3+x_4=1$$

得:

$$x_1=x_2=x_3=x_4=\frac{1}{4},J=\frac{1}{4}$$




## （二）.KKT条件( Karush-Kuhn-Tucker condition)

### 2.1 导出KKT条件

以上是关于等式约束的问题，拉格朗日乘子法可以直接的解决。那么如果遇到不等式约束的寻找最优解的问题，该如何解决？考虑如下问题：

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$s.t \ \ g(\boldsymbol{x}) \le 0$$

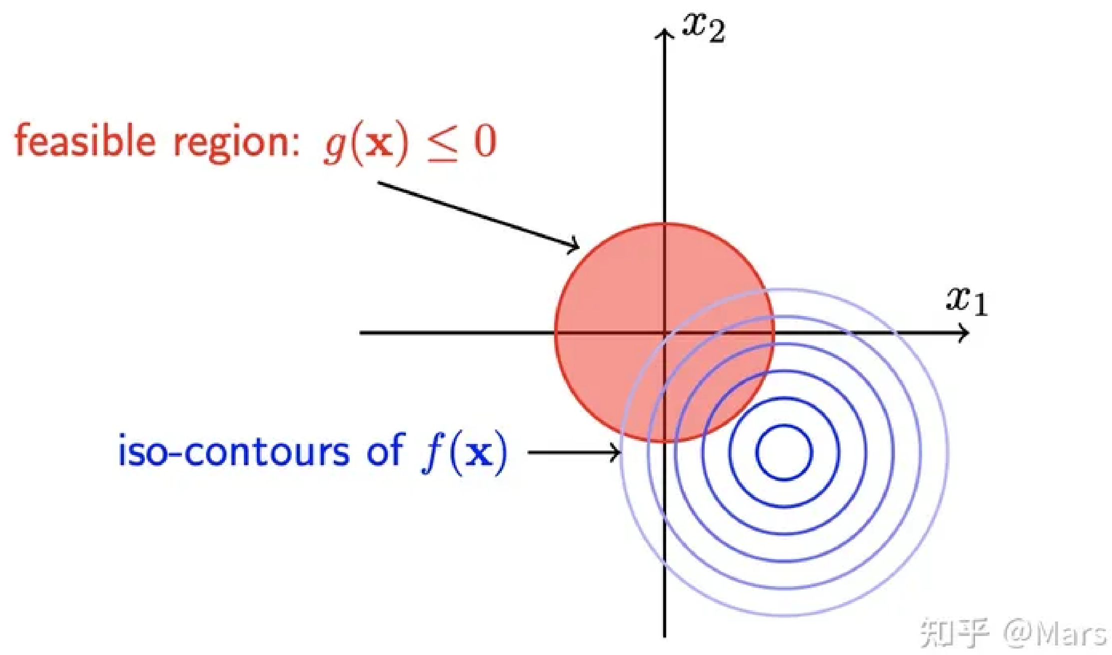
如果按照之前求解**等式约束**的做法，我们构造拉格朗日函数：

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) \ + \ \lambda g(\boldsymbol{x})$$

设最优点为 $\boldsymbol{x}^*$ 。最优点 $\boldsymbol{x}^*$ 要么处在约束区域内 $(g(\boldsymbol{x}) < 0)$ ，要么处在约束区域的边界 $(g(\boldsymbol{x}) = 0)$ 。我们分别来讨论它们。

当最优点处于 $g(\boldsymbol{x}) < 0$ 所决定的约束区域内，此时约束条件 $g(\boldsymbol{x}) \le 0$ 不起作用，目标函数 $f(\boldsymbol{x})$ 的最小值完全取决其本身，我们直接取 $\nabla f(\boldsymbol{x}) = 0$ 得到极值点。**等价于将拉格朗日函数的 $\lambda$ 取为零，然后求 $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = 0$；**

> 如何理解最优点在约束区域内部，约束不起作用？可以参考下图，蓝色的线是目标函数 $f(\boldsymbol{x})$ 的等值线，红色区域代表约束区域，如果约束点 $\boldsymbol{x}^*$ 在红色区域内部，那么目标函数的极值就和约束条件无关了，其极值完全取决于 $f(\boldsymbol{x})$ 本身。




当最优点处于$g(\boldsymbol{x}) = 0$ 所决定的约束边界上，此时的情况就和等式约束类似，约束条件起作用。最优点 $\boldsymbol{x}^*$ 在边界上，**约束条件函数在 $\boldsymbol{x}^*$ 的梯度正交于约束曲面**，**目标函数在 $\boldsymbol{x}^*$ 的梯度正交于约束曲面**，**并且二者方向必相反**，所以必有： $\nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0 (\lambda > 0)$；

> 关于约束曲面边界上约束函数,目标函数的梯度问题：

- 约束函数的边界表达式为： $g(\boldsymbol{x}) = 0$ ，**其上任意一点 $\boldsymbol{x}$ 的梯度必正交于该约束曲面。**这里可以类比高等数学（多元函数微分学在几何上的应用）中函数曲面**对其变量分别求偏导**，得到的向量就是**法向量**；
- 函数等值线与约束曲面相切。所以对于最优点而言，目标函数在该点的梯度正交与约束曲面；
- 综合上述两点，又约束曲面的梯度指向外，而函数的梯度指向增长的方向，所以必有：$\nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0 (\lambda > 0)$ ；

综合以上最优点在约束区域内部和约束区域边界的分析，我们可以得到**KKT条件**：

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = \nabla f(\boldsymbol{x}) \ + \ \lambda \nabla g(\boldsymbol{x}) = \mathbf{0}$$
$$g(\boldsymbol{x}) \le 0$$
$$\lambda \ge 0$$
$$\lambda g(\boldsymbol{x}) = 0$$

> KKT条件的一些说明：

- **KKT条件是不等式约束的最优解的必要条件；**
- $g(\boldsymbol{x}) \le 0$ 是约束条件所限制的，即**原始可行性(primal feasibility)**；
- $\lambda \ge 0$ ：最优点在约束区域内部时： $\lambda = 0$ ; 最优点在约束区域边界时： $\lambda > 0$ ，即**对偶可行性(dual feasibility)**;
- $\lambda g(\boldsymbol{x}) = 0$ :最优点在约束区域内部时： $\lambda = 0$；最优点在约束区域边界时： $g(\boldsymbol{x}) = 0$，即**互补松弛性(complementary slackness)**；

### 2.2.不等式约束示例

$$\min_{x} J = x_1^2 + x_2^2 + x_3^2 + x_4^2$$
$$s.t \;\; x_1 + x_2 + x_3 + x_4 = 1$$
$$s.t \quad x_4 \leq A$$

$Solution:$

$$\overline{J} = x_1^2 + x_2^2 + x_3^2 + x_4^2 - \lambda(1 - x_1 - x_2 - x_3 - x_4) + \mu(x_4 - A)$$

求偏导数：

$$\frac{\partial \overline{J}}{\partial \boldsymbol{x}} = \begin{pmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \\ 2x_3 - \lambda \\ 2x_4 - \lambda + \mu \end{pmatrix} = 0$$

得：

$$x_1 = x_2 = x_3 = \frac{\lambda}{2}, x_4 = \frac{\lambda - \mu}{2}$$

代入等式约束，得： $\lambda = \frac{2+\mu}{4}$ ，得到的结果再代入 $x$ ：

$$x_1 = x_2 = x_3 = \frac{1}{4} + \frac{\mu}{8} \, , x_4 = \frac{1}{4} - \frac{3\mu}{8}$$

要想得到该问题的最优解，则必须满足kkt条件：

$x_4 - A \leq 0 \qquad (1)$
$\mu \geq 0 \qquad (2)$
$\mu(x_4 - A) = 0 \qquad (3)$

该问题的未知量是 $A$ ,我们对 $A$ 做以下讨论：

当 $A \geq \frac{1}{4}$时：

此时 $\frac{1}{4} - A \leq 0$ ,也就意味着 $\mu \geq 0$ , 从而 $x_1 = x_2 = x_3 \geq \frac{1}{4}, x_4 \leq \frac{1}{4}$ ，又 $x_4 - A \neq 0$

所有此时 $\mu = 0$ ，最优解为： $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$

当 $A < \frac{1}{4}$ 时：

如果 $x_4 < A$ ，那么$\mu = 0$，那么条件（1）不成立，所以 $x_4 = A$ ， $x_1 = x_2 = x_3 = \frac{1-A}{3}$

将上面讨论的两种最优解的极值点代入目标函数,可得到最小值，即：

$$J = \begin{cases} \frac{1}{4} & \text{if } A \geq \frac{1}{4} \\ \frac{1}{3}(4A^2 - 2A + 1) & otherwise \end{cases}$$

## 总结：

> 对于等式约束，我们可以直接使用拉格朗日乘子法。对于存在不等式的约束，我们在使用拉格朗日乘子法的同时要必须满足kkt条件。另外，文中那里有错误，请读者留言指教 。

参考文献：

周志华：《机器学习》，清华大学出版社；

Stanley B. Gershwin ,KKT Examples, Massachusetts Institute of Technology；

编辑于 2021-07-27 14:03

机器学习 SVM

2 条评论

默认 最新

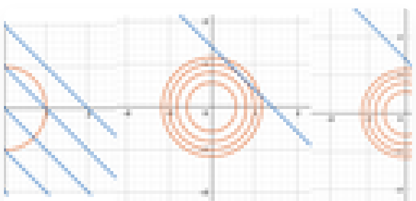
**元稹2009**
约束曲面的梯度指向外，为什么呀?
2022-12-18
回复 赞

**元稹2009**
约束曲面的梯度是怎么得到的呀?
2022-12-18
回复 赞

## 文章被以下专栏收录

**机器学习**
记录我在学习机器学习时的一些算法的数学推导和原理。

## 推荐阅读



**从拉格朗日乘子法到KKT条件**

二圈妹 发表于数理大乱炖

**拉格朗日乘子法与KKT条件详解**

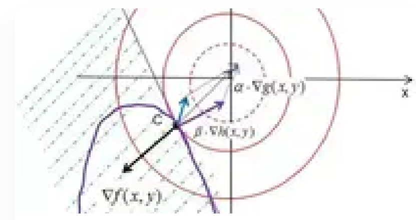
文章来自 深入理解拉格朗日乘子法（Lagrange Multiplier) 和KKT条件。在求取有约束条件的优化问题时，拉格朗日乘子法（Lagrange Multiplier) 和KKT条件是非常重要的两个求取方法。 对于等式...

漫漫成长

**拉格朗日乘数法理解2——KKT条件**

回顾 前面一篇文章 拉格朗日乘数法理解 中讲了求 \displaystyle f(x,y) 在 等式限制(equality constraint) \displaystyle g(x,y)=c 条件下取得极值的问题，此时 abla f 和 abla g 是平行的...

清雅白鹿记



**图解KKT条件和拉格朗日乘子法**

胡卫雄 发表于机器学习入...