

目录

- [1 拉格朗日乘法基本思想](#)
 - [1.1 引入拉格朗日乘子](#)
 - [1.2 基本形态](#)
 - [1.3 示例](#)
- [2 KKT 条件](#)
 - [2.1 不等式约束](#)
 - [2.2 KKT](#)
 - [2.3 多约束条件](#)
- [3 拉格朗日乘子的来源](#)

拉格朗日乘法基本思想

【参考】

- [cnblogs - Math & Algorithm 拉格朗日乘数法](#)
- [github.io - 【直观详解】拉格朗日乘法和KKT条件](#)
- [掘金 - 存在即有理---拉格朗日乘子法\(Lagrange Multiplier Method\)](#)
- [作业部落 - 拉格朗日乘子法](#)

拉格朗日乘法 (Lagrange multiplier) 是一种在最优化的问题中寻找多元函数在其变量受到一个或多个条件的相等约束时的求局部极值的方法。这种方法可以将一个有 n 个变量和 k 个约束条件的最优化问题转换为一个有 $n + k$ 变量的无约束优化问题。

引入拉格朗日乘子

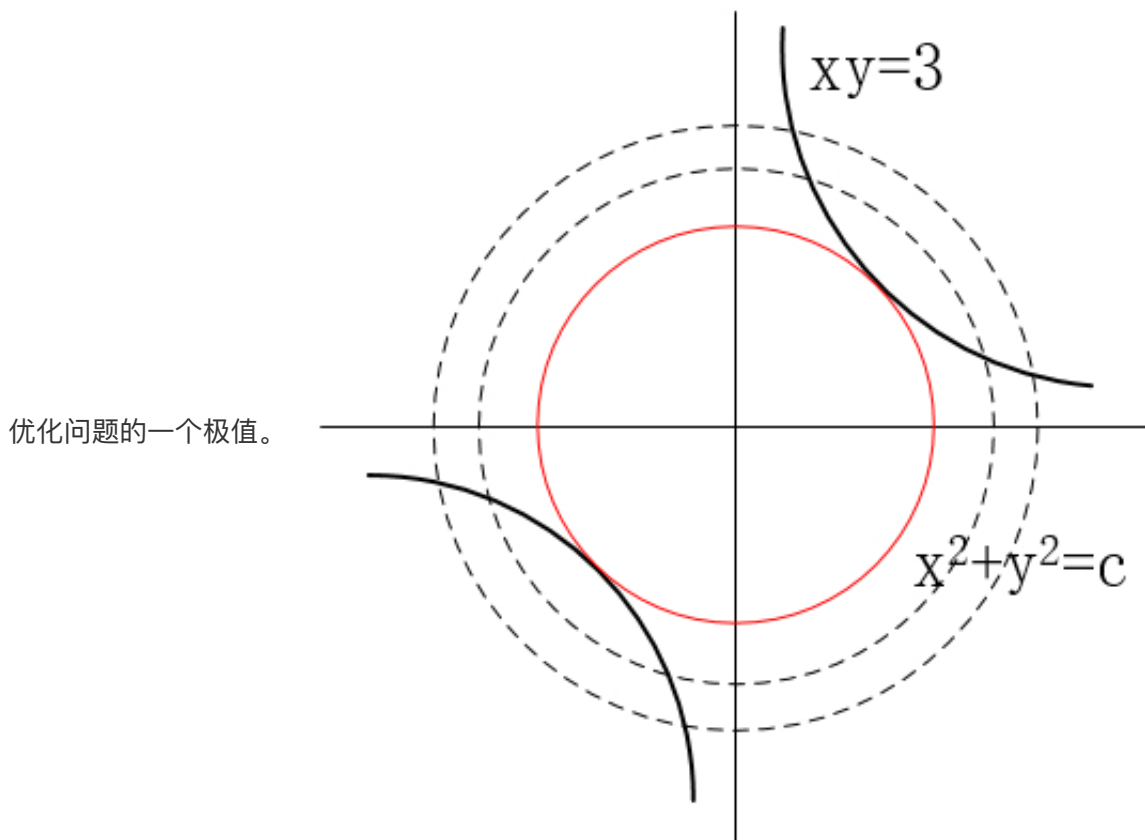
如题：求双曲线 $xy=3$ 上离原点最近的点。

求解：

首先，我们根据问题的描述来提炼出问题对应的数学模型，即： $\min f(x, y) = x^2 + y^2 \quad s.t. \quad xy = 3$ 。

根据上式我们可以知道这是一个典型的约束优化问题，其实我们在解这个问题时最简单的解法就是通过约束条件将其中的一个变量用另外一个变量进行替换，然后代入优化的函数就可以求出极值。我们在这里为了引出拉格朗日乘数法，所以我们采用拉格朗日乘数法的思想进行求解。

我们将 $x^2 + y^2 = c$ 的曲线族画出来，如下图所示，当曲线族中的圆与 $xy = 3$ 曲线进行相切时，切点到原点的距离最短。也就是说，当 $f(x, y) = c$ 的等高线和双曲线 $g(x, y)$ 相切时，我们可以得到上述



现在原问题可以转化为求当 $f(x, y)$ 和 $g(x, y)$ 相切时, x, y 的值是多少? 如果两个曲线相切, 那么它们的切线相同, 即法向量是相互平行的, $\nabla f \parallel \nabla g$. 由 $\nabla f \parallel \nabla g$ 可以得到, $\nabla f = \lambda * \nabla g$.

这时, 我们将原有的约束优化问题转化为了一种对偶的无约束的优化问题, 如下所示:

$$\begin{aligned} \text{原问题: } \min f(x, y) &= x^2 + y^2 \\ \text{s.t. } xy &= 3 \end{aligned}$$

约束优化问题

对偶问题: 由 $\nabla f = \lambda * \nabla g$ 得,

$$\nabla f_x = \lambda * \nabla g_x,$$

$$\nabla f_y = \lambda * \nabla g_y,$$

$$xy = 3$$

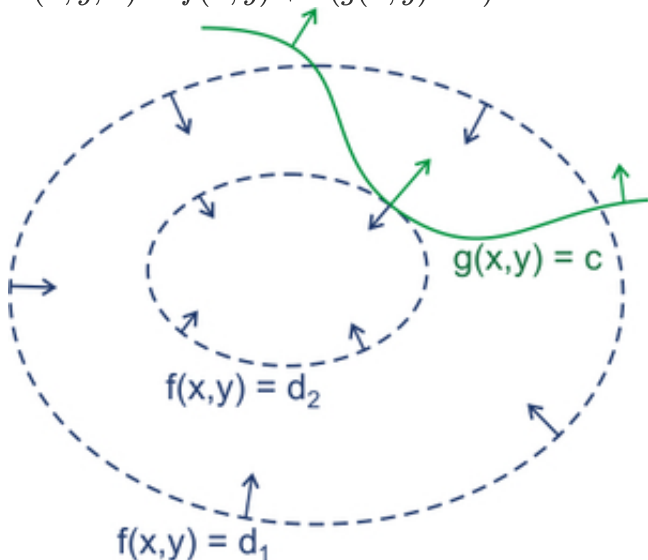
无约束方程组问题

通过求解右边的方程组我们可以获取原问题的解, 即: $2x = \lambda * y, 2y = \lambda * x, xy = 3$.

通过求解上式可得, $\lambda = 2$ 或者是 -2 ; 当 $\lambda = 2$ 时, $(x, y) = (\sqrt{3}, \sqrt{3})$ 或者 $(-\sqrt{3}, -\sqrt{3})$, 而当 $\lambda = -2$ 时, 无解。所以原问题的解为 $(x, y) = (\sqrt{3}, \sqrt{3})$ 或者 $(-\sqrt{3}, -\sqrt{3})$ 。通过举上述这个简单的例子就是为了体会拉格朗日乘数法的思想, 即通过引入拉格朗日乘子(λ)将原来的约束优化问题转化为无约束的方程组问题。

基本形态

如下：求函数 $z = f(x, y)$ 在满足 $g(x, y) = c$ 下的条件极值，可以转化为函数 $F(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - c)$ 的无条件极值问题。如下面的示意图：



绿线标出的是约束 $g(x, y) = c$ 的点的轨迹。蓝线是 $f(x, y)$ 的等高线。箭头表示斜率，和等高线的法线平行。从图上可以直观地看到在最优解处， f 和 g 的斜率平行。

$\nabla[f(x, y) + \lambda(g(x, y) - c)] = 0, \lambda \neq 0$ ，一旦求出 λ 的值，将其套入

$$F(x, y) = f(x, y) + \lambda(g(x, y) - c) \quad (1)$$

易求在无约束极值和极值所对应的点。

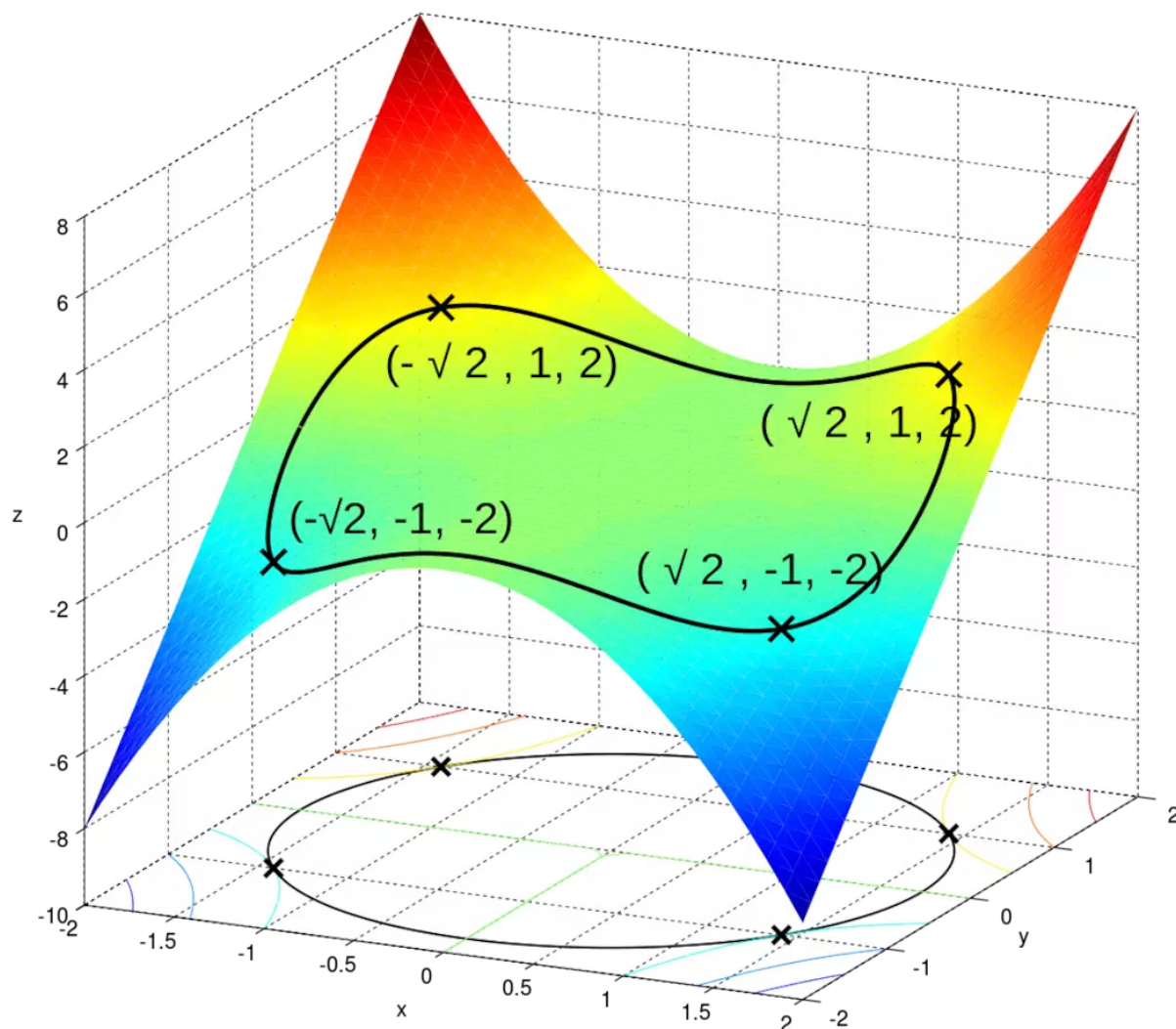
易求在无约束极值和极值所对应的点。

新方程 $F(x, y)$ 在达到极值时与 $f(x, y)$ 相等，因为 $F(x, y)$ 达到极值时 $g(x, y) - c$ 总等于零。上述式子取得极小值时其导数为0，即 $\nabla f(x) + \nabla \sum \lambda_i g_i(x) = 0$ ，也就是说 $f(x)$ 和 $g(x)$ 的梯度共线。

示例

如下示例：

$$\begin{aligned} \max_{x, y} \quad & f(x, y) = x^2 y \\ \text{s.t.} \quad & g(x, y) : x^2 + y^2 - 3 = 0 \end{aligned}$$



设拉格朗日乘子为 λ , 则有如下拉格朗日函数:

$$L(x, y, \lambda) = x^2 y - \lambda(x^2 + y^2 - 3)$$

接下来对 x, y, λ 三个变量分别求偏微分:

$$\begin{aligned} \nabla_{x,y,\lambda} L(x, y, \lambda) &= \left(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial \lambda} \right) \\ &= (2xy + 2\lambda x, x^2 + 2\lambda y, x^2 + y^2 - 3) \end{aligned}$$

令偏微分都为0, 则有:

$$\nabla_{x,y,\lambda} L(x, y, \lambda) = 0 \iff \begin{cases} 2xy + 2\lambda x = 0, & (i) \\ x^2 + 2\lambda y = 0, & (ii) \\ x^2 + y^2 - 3 = 0, & (iii) \end{cases}$$

解出L:

$$(\pm\sqrt{2}, 1, -1); (\pm\sqrt{2}, -1, 1); (0, \pm\sqrt{3}, 0)$$

代入 $f(x, y)$ 中解得最终 $f(x, y)$ 几种可能的值分别为 2, -2, 0, 则此为最大最小值的集合, 可知 $\max_{x,y} f(x, y) = 2$.

KKT 条件

我们上述讨论的问题均为等式约束优化问题, 但等式约束并不足以描述人们面临的问题, 不等式约束比等式约束更为常见, 大部分实际问题的约束都是不超过多少时间, 不超过多少人力, 不超过多少成本等等。所以有几个科学家拓展了拉格朗日乘数法, 增加了 KKT 条件之后便可以用拉格朗日乘数法来求解不等式约束的优化问题了。

KKT 条件是指在满足一些有规则的条件下, 一个非线性规划(Nonlinear Programming)问题能有最优化解法的一个必要和充分条件. 这是一个广义化拉格朗日乘数的成果.

不等式约束

我们看这样一个不等式约束的问题:

$$\min_x f(x) \quad s.t. \quad g(x) \leq 0$$

其对应的拉格朗日函数如下所示:

```

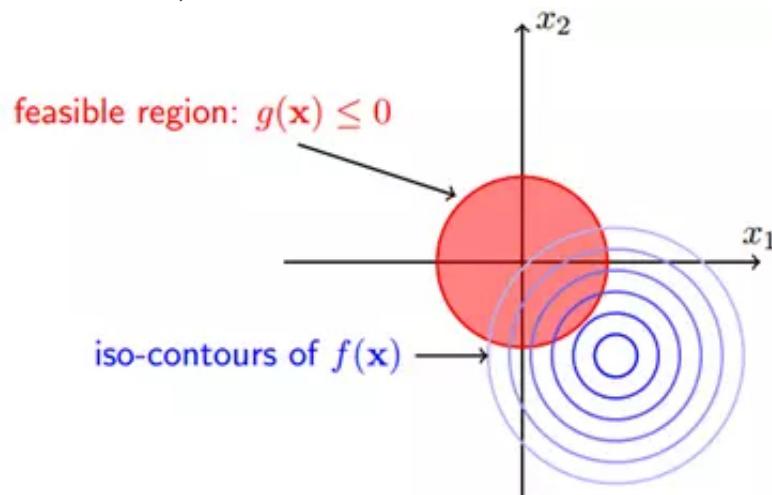

$$L(x, \lambda) = f(x) + \lambda g(x)$$


```

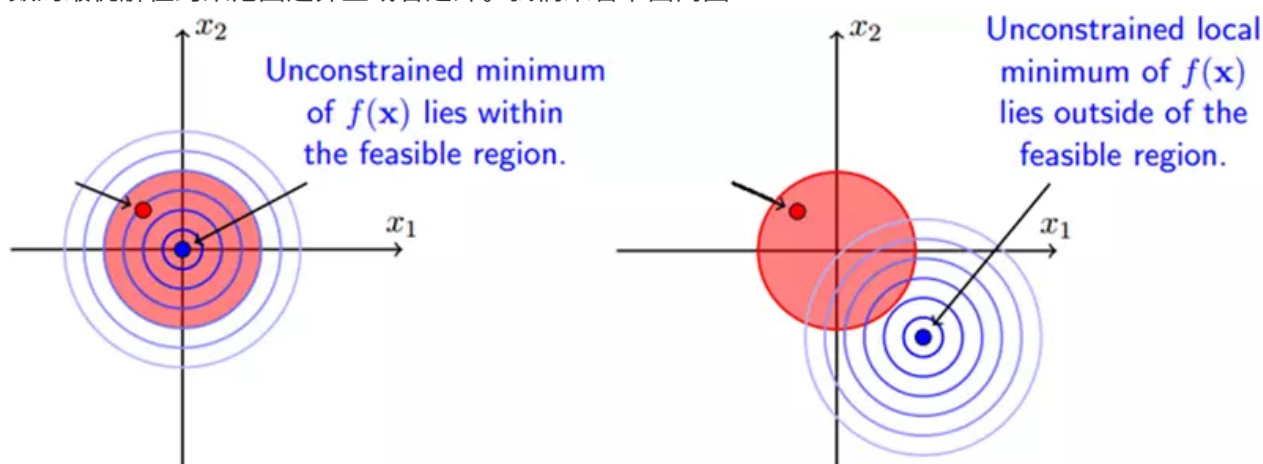
\$\$

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

这时可行解必须落在约束区域以内, 下图给了目标函数的等高线以及约束:



这里红色的区域代表的是不等式的区域，蓝色代表的原函数的等高线，这里我们可以将其想象成一个碗状的三维图形。值得一提的是，我们这里暂时只针对**凸函数问题**，所以我们可以看到这个蓝色圆心便是最优解。那么现在我们只会有两种情况了：一种是原函数的最优解就已经在约束范围内；另一种是原函数的最优解在约束范围边界上或者之外。我们来看下面两图：



我们看第一个图就代表了第一种情况，蓝色圆心在红色圆内。这样约束条件没有起到约束作用，约束之后的最优解还是原来的最优解。

对于第二个图，我们看到蓝色圆心在红色区域之外。当然这也包括了刚好在红色边界上的情况。为什么这两种情况归为一类呢？我们接着看：我们必须满足约束条件，也就是我们必须要在红色范围内找到一个最优解。即使这个最优解并不是原函数最优的。由于原函数是一个凸函数，所以是平滑单个极值。那么理论上说，我们在极值往外任意方向上，都是离极值越近，那个值就越优。所以我们便可知，这个最优值会在约束的边界上。也就是 $g(x) = 0$ 的时候。这时候就又变成了等式约束了。

KKT

以上两种情况就是说，要么可行解落在约束边界上即得 $g(x) = 0$ ，要么可行解落在约束区域内部，此时约束不起作用，令 $\lambda = 0$ 消去约束即可，所以无论哪种情况都会得到：

$$\lambda g(x) = 0$$

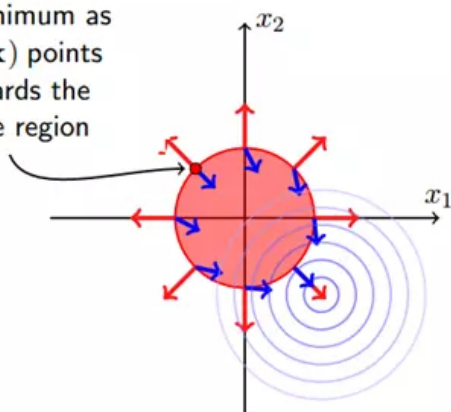
而在这种情况下，我们的不等式约束是和等式约束同解的。这种情况，就是我们所说的 KKT 条件下。所以我们总结出此问题的 KKT 条件：

$$\begin{cases} \nabla_x L(x, \lambda) = 0, & (I) \\ \lambda g(x) = 0, & (II) \\ g(x) \leq 0, & (III) \\ \lambda \geq 0, & (IV) \end{cases}$$

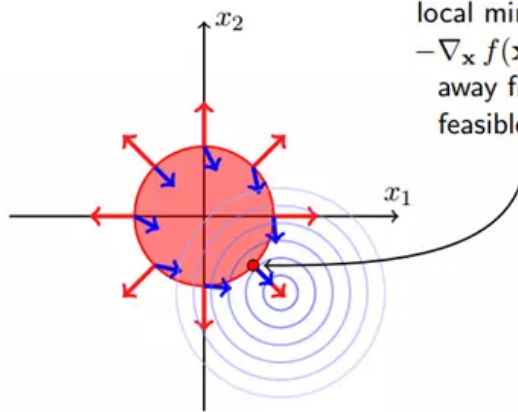
总结： KKT 是我们将不等式约束直接化成等式约束的必要条件，也是这个不等式约束的必要条件。在凸函数优化问题中，升级为充要条件。一般的， KKT 条件很容易满足，我们论证一下就可以使用乘子法解决问题。在拉格朗日对偶中， KKT 条件也是十分重要的条件，判断原问题与对偶问题是否有一致解。

上式需要满足的要求是拉格朗日乘子，这个问题可以举一个形象的例子，假设你去爬山，目标是山顶，但有一个障碍挡住了通向山顶的路，所以只能沿着障碍爬到尽可能靠近山顶的位置，然后望着山顶叹气，这里山顶便是目标函数的可行解，障碍便是约束函数的边界，此时的梯度方向一定是指向山顶的，与障碍的梯度同向，下图描述了这种情况：

✗ **Not** a constrained local minimum as $-\nabla_x f(\mathbf{x})$ points in towards the feasible region



✓ **Is** a constrained local minimum as $-\nabla_x f(\mathbf{x})$ points away from the feasible region



多约束条件

当然更一般的问题：

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

我们有更为完整的KKT条件：

$$\begin{cases} \nabla_x L(x, \alpha_i, \beta_i) = 0, & (1) \\ \alpha_i c_i(x) = 0, & i = 1, 2, \dots, k \quad (2) \\ h_j(x) = 0, & j = 1, 2, \dots, l \quad (3) \\ c_i(x) \leq 0, & i = 1, 2, \dots, k \quad (4) \\ \alpha_i \geq 0, & i = 1, 2, \dots, k \quad (5) \end{cases}$$

满足 KKT 条件后极小化 Lagrangian 即可得到在不等式约束条件下的可行解。KKT 条件看起来很多，其实很好理解：

拉格朗日乘子的来源

拉格朗日乘子法的思路是：给予违反约束适当的惩罚，使得无约束的拉格朗日函数的解与原优化函数的解一致。拉格朗日乘子法的策略是**将有约束的问题转化成无约束的优化问题**。

考虑多个约束条件的原始最优化问题：

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

我们对这种多维约束条件，有如下拉格朗日函数：

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

假设给定某个 x ，如果 x 违反原始问题的约束条件，即存在某个 i 使得 $c_i(x) > 0$ 或者存在某个 j 使得 $h_j(x) \neq 0$ ，那么就有

$$\max_{\alpha, \beta: \alpha_i \geq 0} \left[f(x) + \sum_{i=1}^k \alpha_i c_i(x) + \sum_{j=1}^l \beta_j h_j(x) \right] = \infty$$

因为若某个 i 使约束 $c_i > 0$ ，则可令 $\alpha_i \rightarrow +\infty$ ，若某个 j 使得约束 $h_j(x) \neq 0$ ，则令 β_j 使得 $\beta_j h_j(x) \rightarrow +\infty$ ，而将其余的 α_i, β_j 均取为 0。这样使得不满足约束条件的都取成 ∞ ，而相反的，满足约束条件的我们让 α, β 置 0，使得拉格朗日函数与原函数解相同。

以上陈述详细的说明了拉格朗日乘子法的思想。