



拉格朗日乘数

在数学中的最优化问题中，拉格朗日乘数法（英語：Method of Lagrange multiplier，以数学家约瑟夫·拉格朗日命名）是一种寻找多元函数在其变量受到一个或多个条件的约束时的局部极值的方法。

方法說明

對一個有 n 个变量与 k 个约束条件的最优化问题，拉格朗日乘数法會將其转换成一个 $n + k$ 个变量的方程组，稱作拉格朗日方程（英語：Lagrange function），這個方程組的解將包括所有最優化問題的解。拉格朗日乘数法將會引入一个或一组新的未知数，即拉格朗日乘数（英語：Lagrange multiplier），又称拉格朗日乘子，或拉氏乘子，它们是在转换后的方程，即约束方程中作为梯度的线性组合中各个向量的系数。

比如，求 $f(x, y)$ 在 $g(x, y) = 0$ 时的局部極值时，我们可以引入新变量拉格朗日乘数 λ ，这时我们只要求下列拉格朗日函数的局部极值：

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot g(x, y)$$

更一般地，对含 n 个变量和 k 个约束的情况，有：

$$\mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k) = f(x_1, \dots, x_n) - \sum_{i=1}^k \lambda_i g_i(x_1, \dots, x_n),$$

拉格朗日乘数法所得的臨界點会包含原问题的所有臨界點，但并不保证每个拉格朗日乘数法所得的臨界點都是原问题的臨界點。拉格朗日乘数法的正确性的证明牵涉到偏微分，全微分或連鎖律。

介绍

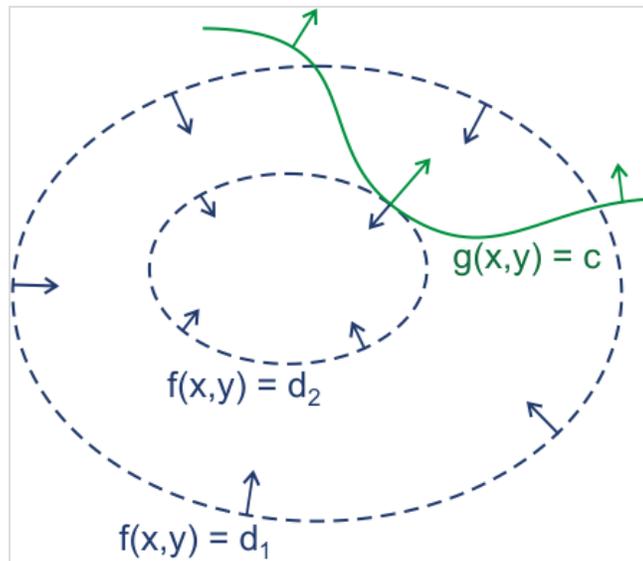


图1：绿线标出的是约束 $g(x, y) = c$ 的点的轨迹。蓝线是 f 的等高线。箭头表示梯度，和等高线的法线平行。

微积分中最常见的问题之一是求一个函数的极大极小值（极值）。但是很多时候找到极值函数的显式表达是很困难的，特别是当函数有先决条件或约束时。拉格朗日乘数则提供了一个非常便利方法来解决这类问题，而避开显式地引入约束和求解外部变量。

先看一个三维的例子：假设有函数： $f(x, y)$ ，要求其极值（最大值/最小值），且满足条件

$$g(x, y) = c$$

c 为常数。對不同 d_n 的值，不难想像出

$$f(x, y) = d_n$$

的等高线。而方程 g 的可行集所构成的线正好是 $g(x, y) = c$ 。想像我们沿着 $g = c$ 的可行集走；因为大部分情况下 f 的等高线和 g 的可行集线不会重合，但在有解的情况下，这两条线会相交。想像此时我们移动 $g = c$ 上的点，因为 f 是连续的方程，我们因此能走到 $f(x, y) = d_n$ 更高或更低的等高线上，也就是说 d_n 可以变大或变小。只有当 $g(x, y) = c$ 和 $f(x, y) = d_n$ 相切，也就是说，此时，我们正同时沿着 $g(x, y) = c$ 和 $f(x, y) = d_n$ 走。这种情况下，会出现极值或鞍点。

气象图中就很常出现这样的例子，当温度和气压两列等高线同时出现的时候，切点就意味着约束极值的存在。

用向量的形式来表达的话，我们说相切的性质在此意味着 f 和 g 的切线在某点上平行，同时也意味着两者的梯度向量有如下关系（引入一个未知标量 λ ）：

$$\nabla f(x, y) = -\lambda \nabla (g(x, y) - c)$$

即：

$$\nabla [f(x, y) + \lambda (g(x, y) - c)] = \mathbf{0}.$$

一旦求出 λ 的值，将其套入下式，易求在无约束条件下的极值和对应的极值点。

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

新方程 $F(x, y, \lambda)$ 在达到极值时与 $f(x, y)$ 相等。因为 $F(x, y, \lambda)$ 达到极值时 $\nabla F(x, y, \lambda) = \mathbf{0}$ ，而 $\frac{\partial F}{\partial \lambda} = g(x, y) - c$ ，也就是说 $g(x, y) - c$ 等于零。

证明

设函数 $f(x, y)$ 在 A 点处有极值 κ ，且在 A 点的邻域内连续。则在 A 点处有

$$f(x, y) = \kappa$$

另有一常值函数

$$g(x, y) = c$$

两函数在A点处的全微分为

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0$$

$$dg = \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy = 0$$

當中的等式右方等於零是由於極值會發生在兩者函數斜率完全相同且 $f(x, y)$ 在區段內有極值的時候

由於 dx 和 dy 是任取的无穷小量，故该线性方程组的系数成比例，有

$$\frac{\frac{\partial f}{\partial x}}{\frac{\partial g}{\partial x}} = \frac{\frac{\partial f}{\partial y}}{\frac{\partial g}{\partial y}} = -\lambda$$

即

$$\frac{\partial f}{\partial x} + \lambda \cdot \frac{\partial g}{\partial x} = 0$$

$$\frac{\partial f}{\partial y} + \lambda \cdot \frac{\partial g}{\partial y} = 0$$

將上二式分別乘以 dx 和 dy ，再相加并积分，得到一新函数

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda \cdot g(x, y)$$

那麼，求原函数极值的问题就转化为求该函数极值的问题。

类似地，这种求极值的方法也可以推广到多维函数 $f(x_1, \dots, x_n)$ 。

拉格朗日乘数的运用方法

如 f 定义为在 \mathbb{R}^n 上的方程，约束为 $g_k(\mathbf{x}) = c_k$ （或将约束左移得到 $g_k(\mathbf{x}) - c_k = 0$ ）。定义拉格朗日 Λ 为

$$\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = f + \sum_k \lambda_k (g_k - c_k)。$$

注意极值的条件和约束现在就都被记录到一个式子里了：

$$\nabla \Lambda = 0 \Leftrightarrow \nabla f = - \sum_k \lambda_k \nabla g_k,$$

和

$$\nabla_{\lambda} \Lambda = 0 \Leftrightarrow g_k = c_k.$$

拉格朗日乘数常被用作表达最大增长值。原因是从式子：

$$-\frac{\partial \Lambda}{\partial c_k} = \lambda_k.$$

中我们可以看出 λ_k 是当方程在被约束条件下，能够达到的最大增长率。拉格朗日力学就使用到这个原理。

拉格朗日乘数法在卡罗需-库恩-塔克条件被推广。

例子

- 很简单的例子

求此方程的最小值：

$$f(x, y) = x^2 y$$

同时未知数满足

$$x^2 + y^2 = 1$$

因为只有一个未知数的限制条件，我们只需要用一个乘数 λ .

$$g(x, y) = x^2 + y^2 - 1$$

$$\Phi(x, y, \lambda) = f(x, y) + \lambda g(x, y) = x^2 y + \lambda(x^2 + y^2 - 1)$$

将所有 Φ 方程的偏微分设为零，得到一个方程组，最小值是以下方程组的解中的一个：

$$2xy + 2\lambda x = 0$$

$$x^2 + 2\lambda y = 0$$

$$x^2 + y^2 - 1 = 0$$

■ 另一个例子

求此离散分布的最大熵：

$$f(p_1, p_2, \dots, p_n) = - \sum_{k=1}^n p_k \log_2 p_k。$$

所有概率的总和是1，因此我们得到的约束是 $g(\mathbf{p}) = 1$ 即

$$g(p_1, p_2, \dots, p_n) = \sum_{k=1}^n p_k = 1。$$

可以使用拉格朗日乘数找到最高熵（概率的函数）。对于所有的 k 从1到 n ，要求

$$\frac{\partial}{\partial p_k} (f + \lambda(g - 1)) = 0,$$

由此得到

$$\frac{\partial}{\partial p_k} \left(- \sum_{k=1}^n p_k \log_2 p_k + \lambda \left(\sum_{k=1}^n p_k - 1 \right) \right) = 0。$$

计算出这 n 个等式的微分，我们得到：

$$- \left(\frac{1}{\ln 2} + \log_2 p_k \right) + \lambda = 0。$$

这说明 p_k 都相等（因为它们都只是 λ 的函数）。解出约束 $\sum_k p_k = 1$ ，得到

$$p_k = \frac{1}{n}。$$

因此，使用均匀分布可得到最大熵的值。

经济学

约束最优化在经济学占有很重要的地位。例如一个消费者的选择问题可以被视为一个求效用方程在预算约束下的最大值问题。拉格朗日乘数在经济学中被解释为影子价格，设定在某种约束下，在这里即收入的边际效用。

拉格朗日乘数就是效用函数在最优解处对收入的偏导数，也就是在最优解处增加一个单位收入带来的效用增加，或者说在最优解处有效用衡量收入的价值，称之为收入的边际效用。

在企业生产问题中，拉格朗日乘数用来衡量要素投入变动所带来的收入变动， $du/dm=\lambda_u$ 表示效用函数或生产函数， m 表示收入或要素投入。

在具体数学推导中还可以运用包络定理的内容。

参考

- 卡罗需－库恩－塔克条件：拉格朗日乘数的推广。
- 拉格朗日方程式
- 哈密頓原理
- 作用量

对外链接

参考拉格朗日原作或方法的命名：

- Earliest known uses of some of the words of mathematics: L (<https://web.archive.org/web/19991003023640/http://members.aol.com/jeff570/l.html>)

更深入的介绍和互动applet：

- Applet (<https://web.archive.org/web/20070101204006/http://www-math.mit.edu/18.02/applets/LagrangeMultipliersTwoVariables.html>)
- Tutorial and applet (<https://web.archive.org/web/20061110031805/http://www.math.gatech.edu/~carlen/2507/notes/lagMultipliers.html>)
- Conceptual introduction (<http://www.slimy.com/~steuard/teaching/tutorials/Lagrange.html>) (页面存档备份 (<https://web.archive.org/web/20061222050610/http://www.slimy.com/~steuard/teaching/tutorials/Lagrange.html>)，存于互联网档案馆) (概念介绍和对于拉格朗日乘数方法在变分法以及物理中的运用)
- Lagrange Multipliers without Permanent Scarring (<http://www.cs.berkeley.edu/~klein/papers/lagrange-multipliers.pdf>) (页面存档备份 (<https://web.archive.org/web/20061207085838/http://www.cs.berkeley.edu/~klein/papers/lagrange-multipliers.pdf>)，存于互联网档案馆) (tutorial by Dan Klein)

检索自“<https://zh.wikipedia.org/w/index.php?title=拉格朗日乘数&oldid=84899620>”