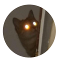
形象理解拉格朗日乘子法



Cat food

本文分三种情况讲解拉格朗日乘子法：等式约束情形、不等式约束情形、混合情形。

1.等式情形

1.1. 简要回顾拉格朗日（Lagrange）乘子法

拉格朗日乘子法是一种将约束优化问题转化为无约束优化问题的方法，我们现在要求解如下优化（最小化）问题：

$$\min f(\boldsymbol{x}) \quad \text{s.t. } g(\boldsymbol{x})=0$$

式中 $\boldsymbol{x}$ 可以是多元向量 $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)$ 。s.t.是subject to的缩写，意思是 受制于，后跟约束条件。

这个约束很烦，它限制了解空间。但是如果引入如下拉格朗日函数：

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

式中 λ 称为拉格朗日乘子，函数L本身就没有约束了。原问题的最优解 $\boldsymbol{x}^*$ 可以通过解这个方程组得到：

$$\begin{cases}\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\lambda)=0\\ g(\boldsymbol{x})=0\end{cases}$$

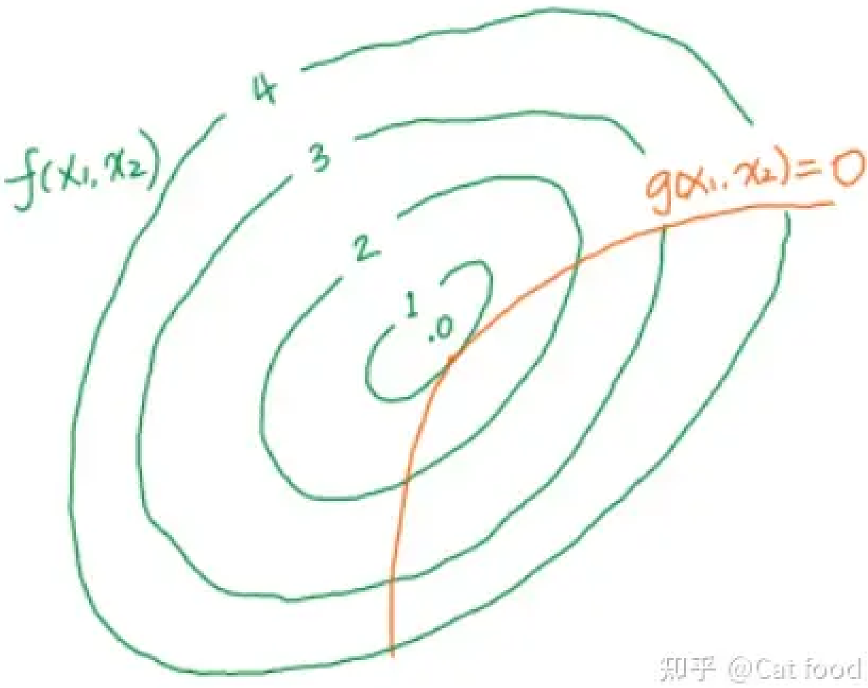
$\nabla_{\boldsymbol{x}} L=0$ 即L对 $\boldsymbol{x}$ 的各分量的偏导都等于0： $\partial L/\partial x_i=0$ 。但注意此方程组只是必要条件，即这个方程组求出来的解不一定都是最优解（例如存在鞍点），但是最优解$\boldsymbol{x}^*$一定在里面。在一些特殊情况下，如f是凸函数，这个方程组的解就才一定是最优解。

1.2. 如何理解拉格朗日乘子法?

现在已经知其然，还要知其所以然。为什么最优解一定在 $\nabla_{\boldsymbol{x}} L=0,\quad g(\boldsymbol{x})=0$ 的解集里？我们从图形层面去说明拉格朗日乘子法的原理。为了方便可视化，我们令自变量是二维的 $\boldsymbol{x}=(x_1,x_2)$ ，问题如下：

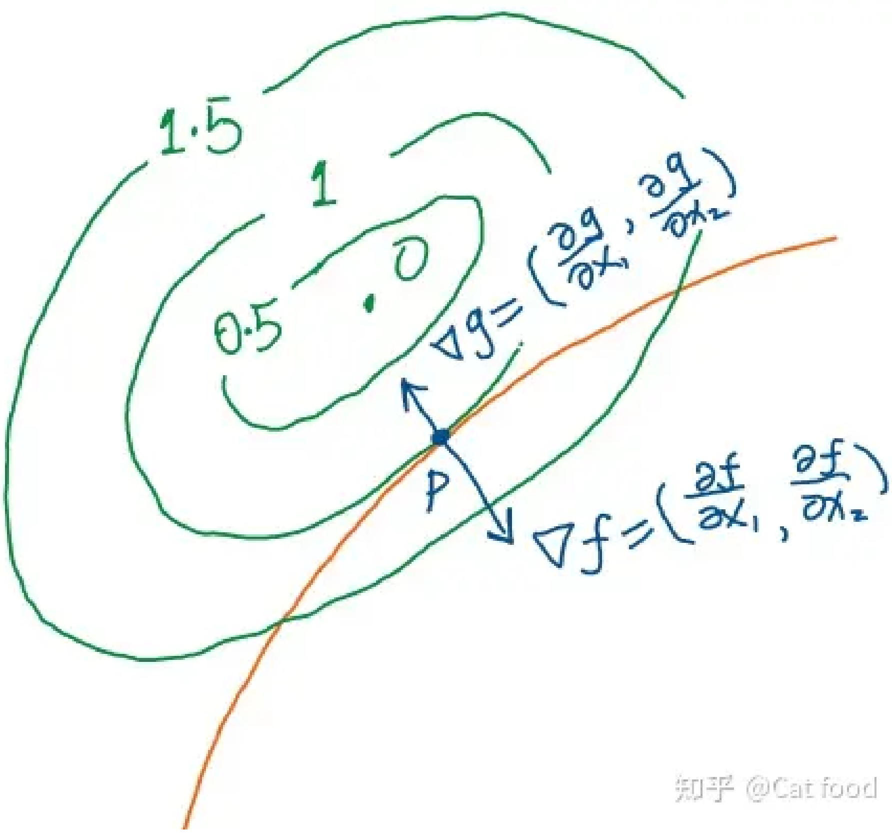
$$\min f(x_1,x_2) \quad \text{s.t. } g(x_1,x_2)=0$$

目标函数 $f(x_1,x_2)$ 应该是曲面，可在XY坐标系中用等高线地图（绿线）来表示；约束方程 $g(x_1,x_2)=0$则是曲线，用黄线画出，如下：



这样一看，该最优化问题相当于在黄线上找海拔最低的点

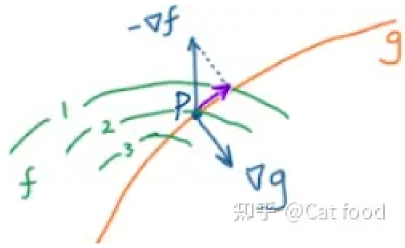
仔细想想可以发现：我们所求的在黄线约束 $g(x_1, x_2) = 0$ 下的最优点P一定是**约束曲线g=0与目标函数f的某一条等值线的切点**，也就是最优点P处约束曲线的法向量 ∇g 一定与该处的目标函数的梯度 ∇f 共线（同向或反向，因为∇g的方向可正可负）。如下图所示：



约束曲线法向量 与 目标函数梯度 在最优点处必然共线

再仔细体会体会~如果还是不能理解为什么的话，这里只好搬出反证法咯：

如下图所示，假设最优点P处，目标函数梯度∇f与约束的法向量 ∇g 不共线，因此负梯度 $-\nabla f$（表示f下降最快的方向）与∇g也不会共线，这样一来负梯度 $-\nabla f$ 在 约束曲线g 上的切向上就存在 紫色的分量，这就表明黄线上的P点沿此方向再挪一点，目标函数值还能进一步下降，所以当前的P点并不是最优点，与假设矛盾。



另外，有人可能会问为什么∇g是g的法向量？有这个问题的朋友可能不清楚函数梯度和曲线（面）方程法向量的关系，它俩的求法一模一样，都是对各分量的偏导数。这也是有直观的几何解释的，参看法向量和梯度的关系。

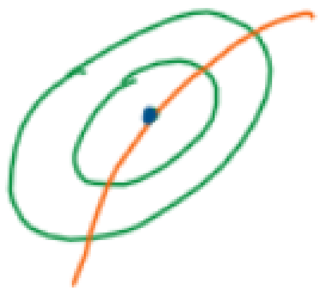
回到我们的结论：最优点P处约束曲线的法向量 ∇g 一定与该处的目标函数的梯度 ∇f 共线。

这句话写成数学式子就是：

$\exists \lambda \in R,$ 使得$\nabla f + \lambda \nabla g = 0$

所谓的拉格朗日乘子 λ 就是待求的一个伸缩系数。令 $L(x, \lambda) = f(x) + \lambda g(x)$ 后，上式正是 $\nabla_x L = 0$。同时别忘了也要满足约束条件 $g(x) = 0$。这就从几何层面上解释了拉格朗日乘子法的原理。

特别地，如果碰巧解得 $\lambda = 0$ ，说明∇f=0，这表明约束曲线g=0恰好经过了f的梯度为0的点（极值点或鞍点）



上面的讨论可以推广到多个等式约束下的多元函数优化问题：

$$\min f(x_1, x_2 \cdots x_n) \quad s.t. \begin{cases} g_1(x_1, \cdots x_n) = 0 \\ g_2(x_1, \cdots, x_n) = 0 \\ \vdots \\ g_m(x_1, \cdots x_n) = 0 \end{cases}$$

令 $\mathcal{L}(x_1, \cdots x_n; \lambda_1, \cdots \lambda_m) = f + \lambda_1 g_1 + \lambda_2 g_2 + \cdots + \lambda_m g_m = f + \sum_{i=1}^{m} \lambda_i g_i \quad (\lambda_i \in R)$

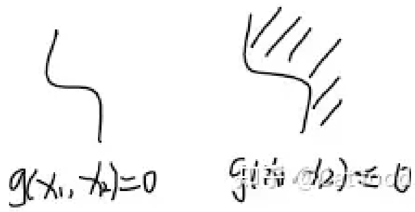
最优解 $x^* = (x_1^*, \cdots x_n^*)^T$ 的必要条件是

$$\begin{cases} \nabla_x L = \nabla_x f + \sum_{i=1}^{m} \lambda_i \nabla_x g_i = 0 \iff \begin{cases} \frac{\partial L}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial L}{\partial x_n} = 0 \end{cases} \quad (n\text{个方程}) \\ g_i = 0 \quad (m\text{组原约束}) \end{cases}$$

2. 不等式情形

拉格朗日乘子法还可用于不等式约束，这里仍采用类似的叙述思路。我们还是以二维情形来说明：

设有优化问题: $\min f(x_1,x_2) \quad \text{s.t. } g(x_1,x_2) \le 0$。由于约束变成了不等式，所以可行域相应地变成了二维区域：

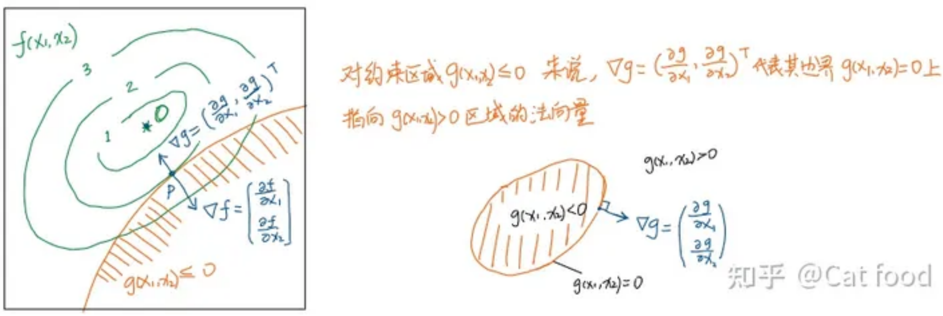


> 注意：这里不等式约束以≤0为例，若此处约束是≥号，下文相应地都要变号

这里我们根据最优点所处的位置，分两种情况来讨论：

2.1 情况A：最优点 (x_1^*, x_2^*) 在约束区域边界上取得，即 $g(x_1^*, x_2^*) = 0$

此时与上一篇文章讲到的等式约束的情况很像，但也有不同，那就是在最优点P处，约束区域边界的法向量 ∇g 与目标函数梯度 ∇f 的方向**必然相反**，原因在下图中：



写成数学语言即 $\exists \lambda > 0, \quad \nabla f + \lambda \nabla g = 0$

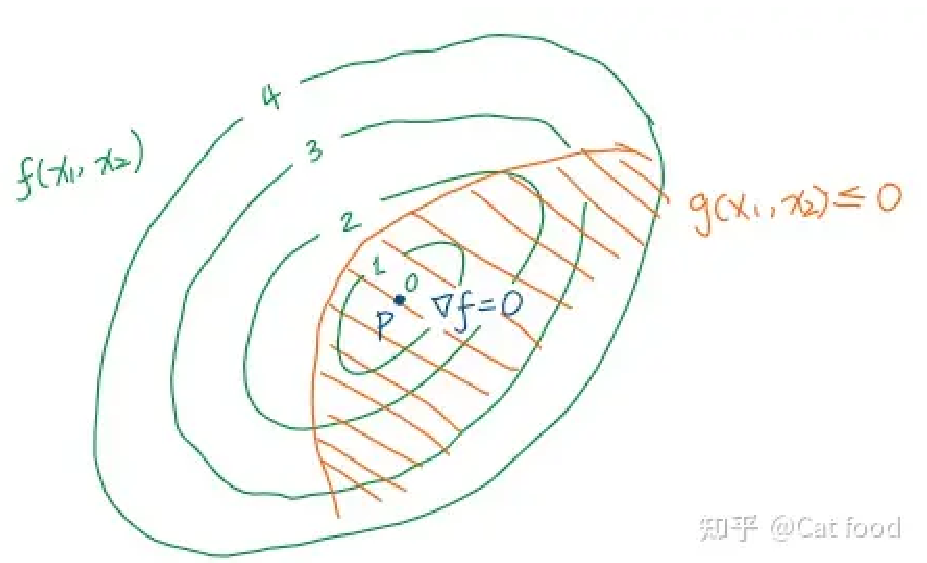
该情况A下最优解的必要条件可以写为下面这个方程组：

$$\begin{cases} \nabla L(x_1,x_2,\lambda) = \nabla f(x_1,x_2) + \lambda \nabla g(x_1,x_2) = 0 \quad \text{【}\nabla f\text{与}\nabla g\text{反向】} \\ \lambda > 0 \\ g(x_1,x_2) = 0 \quad \text{【题设，最优点在边上取得】} \end{cases}$$

接下来我们看第二中情况：

2.2 情况B：最优点 (x_1^*, x_2^*) 在约束区域内部 $g(x_1,x_2)=0$ 取得，即 $g(x_1^*, x_2^*) < 0$

这种情况相当于约束g没起作用，仅靠求 $\nabla f = 0$ 即可求得最优点，如下：



但是，为了能统一写成情况A的形式，即 $\nabla f + \lambda \nabla g = 0$，我们不得不让 $\lambda = 0$，此时约束g没有任何实际作用，只是来凑个形式。所以情况B下最优解的必要条件可以写为下面这个方程组：

$$\begin{cases} \nabla L(x_1, x_2, \lambda) = \nabla f(x_1, x_2) + \lambda \nabla g(x_1, x_2) = 0 & \text{【其实相当于} \nabla f = 0 \text{】} \\ \lambda = 0 & \text{【} g \text{没有作用】} \\ g(x_1, x_2) < 0 & \text{【题设，最优点在内部取得】} \end{cases}$$

2.3 合并情况A+B

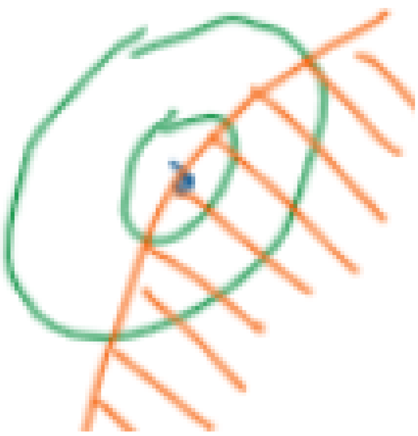
现综合考虑以上情况A和B，最优解满足的式子都可以写为：

$$\nabla f(x_1, x_2) + \lambda \nabla g(x_1, x_2) = 0 \quad (\lambda \geq 0)$$

其中：

- 最优点在边界上时，有 $g(x_1, x_2) = 0$，此时 $\lambda > 0$
- 最优点在内部时，有 $g(x_1, x_2) < 0$，此时相当于没有g，$\lambda = 0$

所以g与 λ 这两者至少有一个必为0，即 $\lambda g = 0$，这就是**互补松弛条件**。这是源于最优点“不在边界上那就是在界内“的对立事件的自然逻辑结果。如果两者同时为0，说明非常巧，约束边界恰好经过函数本身的最优点（严格来说是驻点）。



\lambda 和 g(x1,x2) 同时为零

综合以上两种情况，得到不等式约束下最优解的必要条件为：

$$\begin{cases} \nabla L(x_1,x_2,\lambda)=\nabla f(x_1,x_2)+\lambda\nabla g(x_1,x_2)=0 \\ \lambda\geq 0 \\ \lambda g(x_1,x_2)=0 \quad \text{【互补松弛】} \\ g(x_1,x_2)\leq 0 \quad \text{【原约束】} \end{cases}$$

以上讨论可以推广到多元目标函数，多个不等式约束的最优化问题：

$$\min f(x_1,x_2,\cdots x_n) \quad s.t. \begin{cases} g_1(x_1,\cdots x_n)\leq 0 \\ g_2(x_1,\cdots x_n)\leq 0 \\ \vdots \\ g_m(x_1,\cdots x_n)\leq 0 \end{cases}$$

$$\text{令 } L(x_1,\cdots x_n;\lambda_1,\cdots\lambda_m)=f+\lambda_1 g_1+\lambda_2 g_2+\cdots+\lambda_m g_m=f+\sum_{i=1}^{m}\lambda_i g_i$$

最优解 $x^*=(x_1^*,\cdots x_n^*)^T$ 的必要条件(KKT)是：

$$\begin{cases} \nabla L=\nabla f+\sum_{i=1}^{m}\lambda_i\nabla g_i=0 \Leftrightarrow \begin{cases} \frac{\partial L}{\partial x_1}=\frac{\partial f}{\partial x_1}+\sum_{i=1}^{m}\lambda_i\frac{\partial g_i}{\partial x_1} \\ \vdots \quad \cdots \\ \frac{\partial L}{\partial x_n}=\frac{\partial f}{\partial x_n}+\sum_{i=1}^{m}\lambda_i\frac{\partial g_i}{\partial x_n} \end{cases} \text{(m个方程)} \\ \lambda_i\geq 0 \quad (i=1,2,\cdots,m) \Leftrightarrow \begin{cases} \lambda_1\geq 0 \\ \vdots \\ \lambda_m\geq 0 \end{cases} \\ \lambda_i g_i=0 \Leftrightarrow \begin{cases} \lambda_1 g_1=0 \\ \lambda_2 g_2=0 \\ \vdots \\ \lambda_m g_m=0 \end{cases} \text{(m个互补松弛条件)} \\ g_i\leq 0 \quad (i=1,2,\cdots,m) \Leftrightarrow \begin{cases} g_1\leq 0 \\ g_2\leq 0 \\ \vdots \\ g_m\leq 0 \end{cases} \text{(m组原约束)} \end{cases}$$

3 混合情形

一般约束优化问题就是既有等式约束也有不等式约束。现在直接用矢量 $\boldsymbol{x}$ 代表多元向量， $f(\boldsymbol{x})$ 代表多元目标函数，存在m个等式约束和n个不等式约束：

$$\min f(\boldsymbol{x}) \quad \text{s.t.} \begin{cases} g_1(\boldsymbol{x})=0 \\ \cdots \\ g_m(\boldsymbol{x})=0 \\ h_1(\boldsymbol{x})\leq 0 \\ \cdots \\ h_n(\boldsymbol{x})\leq 0 \end{cases}$$

定义拉格朗日函数

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})=f(\boldsymbol{x})+\sum_{i=1}^{m}\lambda_i g_i(\boldsymbol{x})+\sum_{j=1}^{n}\mu_j h_j(\boldsymbol{x})$$

式中 $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_m]$ 是等式们的拉格朗日乘子。 $\boldsymbol{\mu}=[\mu_1,\ldots,\mu_n]$ 是不等式们的拉格朗日乘子。

综合等式情形和不等式情形，可得混合情形最优解的必要条件是：

$$\begin{cases}\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = \nabla f(\boldsymbol{x}) + \sum_{i=1}^{m}\lambda_i\nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{n}\mu_j\nabla h_j(\boldsymbol{x}) = 0 \\ g_i(\boldsymbol{x}) = 0 \quad (i=1,2,\dots,m) \quad \text{原等式约束} \\ h_j(\boldsymbol{x}) \le 0 \quad (j=1,2,\dots,n) \quad \text{原不等式约束} \\ \lambda_i \in R \quad (i=1,2,\dots,m) \quad \text{等式约束的乘子可正可负，此式可省略} \\ \mu_j \ge 0 \quad (j=1,2,\dots,n) \quad \text{不等式约束的乘子不能为负} \\ \mu_j h_j(\boldsymbol{x}) = 0 \quad (j=1,2,\dots,n) \quad \text{不等式约束专属互补松弛，每组}h\text{和}\mu\text{至少有一个是}0\end{cases}$$

这个式子也成为**KKT条件**。前面多次强调，它只是（一阶）必要条件，即满足KKT条件的点未必是局部（全局）最优点（还可能是局部极大和鞍点），但局部（全局）最优点必然满足KKT条件。特别地，对于凸规划问题，满足KKT条件的解直接就是全局最优。

例题：

例. 利用最优性条件求解约束优化问题

$$\min \quad f(x) = (x_1-3)^2 + (x_2-1)^2$$

$$s.t. \quad c_1(x) = -x_1^2 + x_2 \ge 0,$$

$$c_2(x) = 2x_1 + x_2 - 3 = 0.$$

解：$L(x,\lambda,\mu) = f(x) - \lambda c_1(x) - \mu c_2(x)$

$= (x_1-3)^2 + (x_2-1)^2 - \lambda(-x_1^2+x_2) - \mu(2x_1+x_2-3)$

$\frac{\partial L}{\partial x_1} = 2(x_1-3) + 2\lambda x_1 - 2\mu,$

$\frac{\partial L}{\partial x_2} = 2(x_2-1) - \lambda - \mu,$

利用$K-T$条件和约束条件即得

$$\begin{cases} 2(x_1-3)+2\lambda x_1 - 2\mu = 0, \\ 2(x_2-1)-\lambda-\mu=0, \\ \lambda(-x_1^2+x_2)=0, \\ -x_1^2+x_2\ge 0, \\ 2x_1+x_2-3=0, \\ \lambda \ge 0. \end{cases}$$

i)若$\lambda=0$,则

$$\begin{cases} 2(x_1-3)-2\mu=0, & (1) \\ 2(x_2-1)-\mu=0, & (2) \\ -x_1^2+x_2\ge 0, & (3) \\ 2x_1+x_2-3=0, & (4) \end{cases}$$

由（1），（2）得$(x_1-3)=2(x_2-1)$,

结合（4）得$x_1=\frac{7}{5}, x_2=\frac{1}{5}$不满足(3),从而 $\lambda\ne 0$.

ii)$\lambda\ne 0, -x_1^2+x_2=0$即$x_2=x_1^2$代入K-T条件即 得

$x_1=1, x_2=1, \lambda=1, \mu=-1,$

或$x_1=-3, x_2=9, \lambda=-11, \mu=27$（舍）

因此K-T 点只有（1,1）^{T}

又f是凸函数，c_1是凹函数，因此问题为凸规划，故（1，1）是全局最优解。

编辑于 2021-12-18 10:12