拉格朗日乘子法——带等式约束项的函数优化

Dezeming Family

2021 年 8 月 28 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。

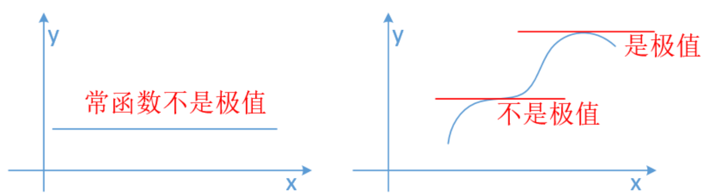
20210903：完成第一版。

目录

一 多元函数的无条件极值

类比一元函数，我们可以想象的到，多元函数的极值条件应该一定要满足偏导为 0。我们可以以二元函数为例，函数在 x 方向的变化率为 0，且在 y 方向变化率为 0，就说明函数有可能达到极值。

当然偏导为 0 不表示一定是极值，因为存在特例，比如对于一元函数，在某点 x_0 处偏导为 0，但可能存在下述情况，故不是极值点：



二 多元函数条件极值

有时候，我们求函数 $f(x,y)$ 时我们需要保证 x,y 满足一些关系。这里的关系有很多种，比如等式关系，不等式关系。不等式关系会困难不少，涉及运筹学线性规划等知识，我们这里讨论等式关系约束项。

例如，我们要求二元函数 $z=f(x,y)$ 在约束项 $\varphi(x_0,y_0)=0$ 下的条件极值。首先设在 (x_0,y_0) 处取得极值，并设 f 和 φ 其在邻域处都有连续一阶偏导（没有的话我们就没法求极值了）。

设方程 φ 变形为 $y=\psi(x)$，代入 f 中，就能得到 $z=f[x,\psi(x)]$，也就是说 z 在 x_0 处取得极值：

$$\frac{dz}{dx}\Big|_{x=x_0}=f'_x(x_0,y_0)+f'_y(x_0,y_0)\frac{dy}{dx}\Big|_{x=x_0}=0 \tag{二.1}$$

根据一元函数隐函数求导公式，由 $\varphi(x,y)=0$ 可得：

$$\left.\frac{dy}{dx}\right|_{x=x_0}=-\frac{\varphi_x'(x_0,y_0)}{\varphi_y'(x_0,y_0)} \tag{二.2}$$

其中 $\varphi_y'(x_0,y_0)\neq 0$。

把上式代入到前面可得：

$$\left.\frac{dz}{dx}\right|_{x=x_0}=f_x'(x_0,y_0)-f_y'(x_0,y_0)\frac{\varphi_x'(x_0,y_0)}{\varphi_y'(x_0,y_0)}=0 \tag{二.3}$$

令 $\frac{f_y'(x_0,y_0)}{\varphi_y'(x_0,y_0)}=-\lambda_0$，于是可以得到：

$$\frac{f_x'(x_0,y_0)}{\varphi_x'(x_0,y_0)}=\frac{f_y'(x_0,y_0)}{\varphi_y'(x_0,y_0)}=-\lambda_0 \tag{二.4}$$

所以就可以联立三个式子，就能得到我们想要的最优解。

$$\begin{cases}\frac{f_x'(x_0,y_0)}{\varphi_x'(x_0,y_0)}=-\lambda_0\\ \frac{f_y'(x_0,y_0)}{\varphi_y'(x_0,y_0)}=-\lambda_0\\ \varphi(x,y)=0\end{cases} \tag{二.5}$$

综上所述,我们可以引入辅助函数 $L(x,y,\lambda)=f(x,y)+\lambda\varphi(x,y)$,于是前两个式子其实就是 $L_x'(x_0,y_0)=0$ 和 $L_y'(x_0,y_0)=0$。λ 为待求常数，我们称为拉格朗日乘子：$L_\lambda'=\varphi(x,y)=0$。

三　多个变量单限定式

设需要取极值的函数为 $t=f(x,y,z)$ 满足 $\varphi(x,y,z)=0$。根据该式我们可以得到两组关系：$y=y(x,z)$ 以及 $z=z(x,y)$。

为了后面的叙述方便，这里我们直接使用“函数 + 下标”表示偏导，省略函数变量值 (x_0,y_0,z_0)。

我们可以得到两组方程：

$$\begin{cases}u=f(x,y,z(x,y)) & u_x=0,u_y=0\\ u=f(x,y(x,z),z) & u_x=0,u_z=0\end{cases} \tag{三.1}$$

我们只说一下第一行的求法：

$$\begin{cases}u_x=f_x+f_z\frac{\partial z}{\partial x}=0\\ u_y=f_y+f_z\frac{\partial z}{\partial y}=0\end{cases} \tag{三.2}$$

根据隐函数求导法可以得到：

$$\begin{cases}\frac{\partial z}{\partial x}=-\frac{\varphi_x}{\varphi_z}\\ \frac{\partial z}{\partial y}=-\frac{\varphi_y}{\varphi_z}\end{cases} \tag{三.3}$$

代入，重复上一节的步骤：令 $\gamma=-\frac{f_z}{\varphi_z}$，得到联立式：

$$\begin{cases}u_x=f_x'(x_0,y_0,z_0)+\gamma\varphi_x'(x_0,y_0,z_0)=0\\ u_y=f_y'(x_0,y_0,z_0)+\gamma\varphi_y'(x_0,y_0,z_0)=0\\ \varphi(x,y,z)=0\end{cases} \tag{三.4}$$

同理，第二行求法可以得到：

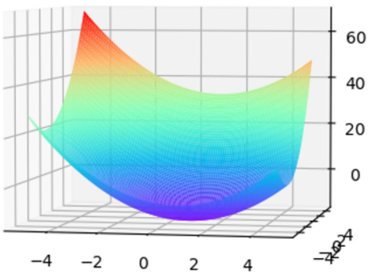
$$\begin{cases}u_x=f_x'(x_0,y_0,z_0)+\gamma_2\varphi_x'(x_0,y_0,z_0)=0\\ u_z=f_z'(x_0,y_0,z_0)+\gamma_2\varphi_z'(x_0,y_0,z_0)=0\\ \varphi(x,y,z)=0\end{cases} \tag{三.5}$$

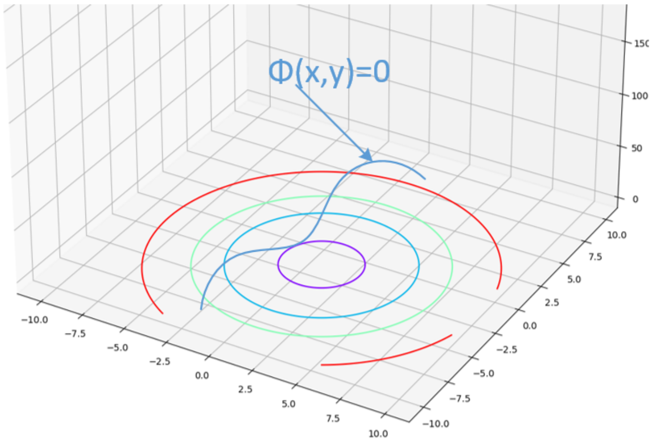
又因为 $\gamma=-\frac{f_z}{\varphi_z}$，$\gamma_2=-\frac{f_y}{\varphi_y}$。根据前面的 $u_y=f'_y+-\frac{f_z}{\varphi_z}\varphi'_y=0$，我们得到 $\gamma=\gamma_2$，所以我们就得到了总的联立式，化为拉格朗日方程即：

$$\begin{cases} L=f(x,y,z)+\gamma\varphi(x,y,z) \\ L_x=L_y=L_z=0 \\ \varphi(x,y,z)=0 \end{cases} \tag{三.6}$$

四　二元函数单约束优化的图示

现在用图示法来讲解一下拉格朗日乘子法的几何意义。如果您对函数梯度并没有那么熟悉，可以参考 DezemingFamily 的《函数的切线与梯度》。下图中，左边表示待优化的函数，$z=x^2-2x+y^2-4y$，约束项为 $\Phi(x,y)=0$；右边是函数的等高线，可以看出，中心函数值较低，越往外函数值越高。





根据上图，我们可以看出，如果约束项与等高线某一环相交，说明该环并不会取得极值，而只有相切时，函数值才会取得极值。

也就是说，当 $\nabla z(x,y)=\lambda\nabla\Phi(x,y)$，即梯度方向一样时，取得最优值。这里的 λ 的值正负未知，因此可以移到另一边，得到通常情况下表示的拉格朗日函数：

$$L=f(x,y)+\gamma\varphi(x,y) \tag{四.1}$$

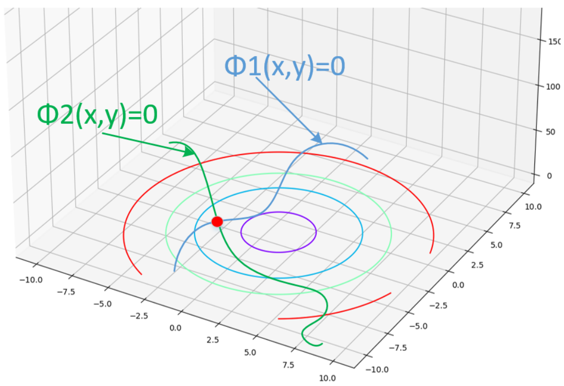
五　多个等式条件

当有多个等式限定条件时，我们可以联立更多的式子：比如需要同时满足 $\varphi_1(x,y,z)=0$，$\varphi_2(x,y,z)=0$。设需要取极值的函数为 $t=f(x,y,z)$。

因为有两个限定条件，所以我们分别得到两组关系：

$$\begin{cases} z=z_1(x,y) \\ y=y_1(x,z) \end{cases} \quad \begin{cases} z=z_2(x,y) \\ y=y_2(x,z) \end{cases} \tag{五.1}$$

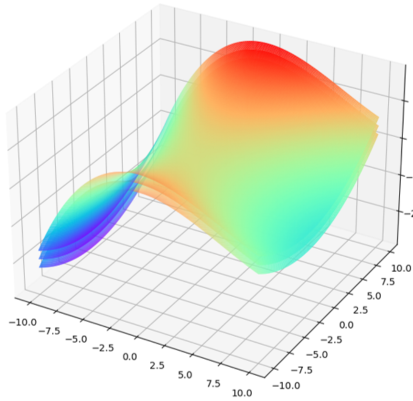
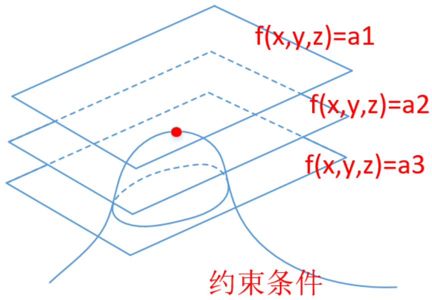
而这两组条件需要同时满足。如果从单约束扩展到多约束，这里我们需要思考一下，首先假如我们的优化目标是二元函数，约束条件的图示如下：



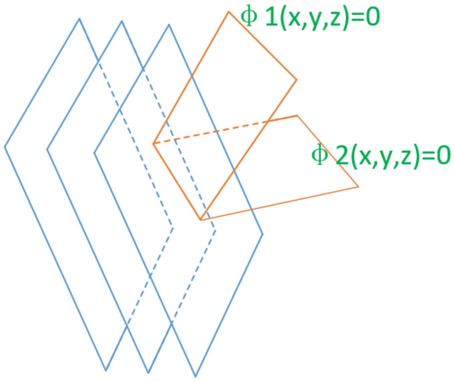
我想这就没的说了，同时满足两个约束的点只有一个，所以求出 $\Phi 1$ 和 $\Phi 2$ 两个约束项的交点，就是该二元函数能取到的最小值了（尽管只有这一个值）。

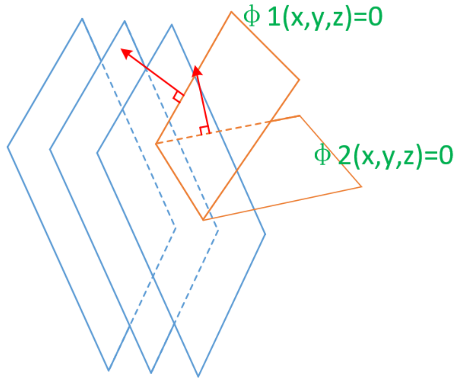
但是当函数扩展到三元函数、四元函数甚至更高维，多个约束项共同约束下可以取到的值就会大于 1 个，例如三维空间里的约束项函数 $\Phi_1(x,y,z)$ 是一个曲面，另一个约束项函数 $\Phi_2(x,y,z)$ 也是一个曲面，因此两个曲面的如果相交，相交处就是一条曲线（除非两个曲面正好相切，例如两个相切的球体曲面）。当然，如果是三元函数对应于三个约束条件，那么这三个约束条件的公共区域也很可能只有一个取值点（大家可以进行类推）。

因此我们现在以优化三元函数为例：$s = f(x,y,z)$，约束项为 $\Phi 1(x,y,z) = 0$ 和 $\Phi 2(x,y,z) = 0$。注意到二元函数值可以表示为等值线，因此三元函数的值就可以表示为等值面，比如下面左图的表示：

右图表示当我们单独观察很小的区域时，等值面可以近似为一个平面。当区域非常微小时，约束条件也是一个微小的平面，在两个约束条件相交处我们可以使用如下图左表示：





而可以取值的 $p(x,y,z)$ 在两个约束条件的曲面相交处移动。

在该局部区域可行域任意一点 p_0，其他可以取值的点在这个局部区域构成的向量 $p - p_0$ 一定会垂直于所有的约束条件面局部区域的法向量，即上图右的法向量。

对于曲面 $\Phi(x,y,z) = 0$，曲面上某点 (x_0,y_0,z_0) 的法向量为：

$$\left(\Phi'_x(x_0,y_0,z_0), \Phi'_y(x_0,y_0,z_0), \Phi'_z(x_0,y_0,z_0)\right) \tag{五.2}$$

我们要求的函数 s 的最小值是在相交线上运动的，而在可取到的最大值附近，$p - p_0$ 必须要平行于 s 当前位置的等值面，否则 p 就可以顺着移动方向到达更大值处。

我们根据图示可以明白，约束条件相交处就是可取值处，可取值处在一小段范围内可以看做是直的线段，该线段不但与待优化函数 s 垂直，同时还与这多个约束条件垂直，我们设该线段向量为 $\vec{l}$，得到如下关系：

$$\vec{l} \perp \nabla\Phi 1 \tag{五.3}$$

$$\vec{l} \perp \nabla\Phi 2 \tag{五.4}$$

$$\vec{l} \perp \nabla f(x,y,z) \tag{五.5}$$

因此，目标函数的法向量就是约束函数法向量的线性组合（约束函数法向量张开的向量空间中）。而曲面方程的法向量在这里应该用梯度来表示（注意区分曲面方程与多元函数，多元函数的梯度表示函数值

增大的方向，见 DezemingFamily 的《函数的切线与梯度》)，我们得到公式：

$$\nabla f(x,y,z)=\lambda_1\nabla\Phi1+\lambda_2\nabla\Phi2 \tag{五.6}$$

因此可以展开成：

$$L(x,y,z,\lambda_1,\lambda_2)=f(x,y,z)+\lambda_1\Phi1(x,y,z)+\lambda_2\Phi2(x,y,z) \tag{五.7}$$

$$\begin{cases} L'_x=f'_x+\lambda_1\Phi1'_x+\lambda_2\Phi2'_x=0 \\ L'_y=f'_y+\lambda_1\Phi1'_y+\lambda_2\Phi2'_y=0 \\ L'_z=f'_z+\lambda_1\Phi1'_z+\lambda_2\Phi2'_z=0 \\ L'_{\lambda_1}=\Phi1(x,y,z)=0 \\ L'_{\lambda_2}=\Phi2(x,y,z)=0 \end{cases} \tag{五.8}$$

计算完以后我们还得思考是不是合理的极值最优点，需要考虑函数的实际意义（比如约束条件下求立方体长宽高时，需要让其值都是正数）。

参考文献

[1] 吴臻分册, 吴臻, 刘建亚. 微积分二 [M]. 山东大学出版社, 2004.

[2] https://zhuanlan.zhihu.com/p/29525538