Appendix E 拉格朗日乘子法（Lagrange Multipiers）

拉格朗日乘子法用来寻找多元函数在一组约束下的驻点方法。

先考虑一个简单的情形，设$\mathbf{x}=(x_1,x_2)^\top$为一个二维向量，要求在给定约束$g(x_1,x_2)=0$下求出$f(x_1,x_2)$的驻点。

$$g(\mathbf{x})=0 \tag{E.1}$$

一种简单的求解方法为，先通过$g(x_1,x_2)=0$将x_2表达为x_1的函数形式，即$x_2=h(x_1)$。通过对$f(x_1,h(x_1))$求导求得x_1^*，随即得到$x_2^*=h(x_1^*)$

上述方法的问题在于，有时候$x_2=h(x_1)$这个表达式很难构建出来，同时，这样的方法也破坏了变量之间的对称性。拉格朗日乘子法通过引入拉格朗日乘子λ，将等式约束融入到目标优化函数中。首先从几何角度来看，设D维向量$\mathbf{x}=(x_1,x_2,\ldots,x_D)$存在于一个D维空间中，$g(\mathbf{x})=0$表示该D维空间中的一个D-1维曲面，对$\mathbf{x}$点附近做泰勒展开，可以得到

$$g(\mathbf{x}+\boldsymbol{\epsilon})\simeq g(\mathbf{x})+\boldsymbol{\epsilon}^\top\nabla g(\mathbf{x}) \tag{E.2}$$

设 $\mathbf{x}+\boldsymbol{\epsilon}$ 为 $g(x)=0$ 曲面上的点，则$g(\mathbf{x}+\boldsymbol{\epsilon})=g(\mathbf{x})=0$，可以得到 $\boldsymbol{\epsilon}^\top\nabla g(\mathbf{x})\simeq\mathbf{0}$，故可以得到下面的结论

- 曲面 $g(\mathbf{x})=0$ 在 $\mathbf{x_0}$ 处的法向量为$\nabla_\mathbf{x}g(\mathbf{x_0})$
- 在寻找最优点$\mathbf{x}^*$时，该点的梯度方向 $\nabla_\mathbf{x}f(\mathbf{x^*})$ 一定也正交于曲面，即存在一个非零值λ满足

  $$\nabla f(\mathbf{x}^*)+\lambda\nabla g(\mathbf{x}^*)=0 \tag{E.3}$$

  这里的λ就是**拉格朗日乘子（Lagrange multiplier）**。这是显然的，因为如果$\nabla_\mathbf{x}f(\mathbf{x^*})$不垂直于曲面的话，可以通过沿着曲面与梯度相反的方向找到一个$\mathbf{x}_1^*$，使得$f(\mathbf{x}_1^*)>f(\mathbf{x}^*)$，通过沿着曲面与梯度相反的方向找到一个$\mathbf{x}_2^*$，使得$f(\mathbf{x}_2^*)>f(\mathbf{x}^*)$。

根据上述结论，我们引入拉格朗日表达式

$$L(\mathbf{x},\lambda)\equiv f(\mathbf{x})+\lambda g(\mathbf{x}) \tag{E.4}$$

显然，(E.3)可以表示为 $\nabla_\mathbf{x}L(\mathbf{x},\lambda)=\mathbf{0}$，而约束条件(E.1)可以表示为 $\frac{\partial L(\mathbf{x},\lambda)}{\partial\lambda}=0$

所以，带有(E.1)约束条件的极值问题转化为(E.5)的方程

$$\nabla_{\mathbf{x},\lambda}L(\mathbf{x},\lambda)=\mathbf{0} \tag{E.5}$$

考虑不等式约束$g(\mathbf{x})\geq 0$，寻找满足约束的$f(\mathbf{x})$的最大值，考虑以下两种情况

- $g(\mathbf{x})>0$，此时约束不起作用（**inactive**），直接通过$\nabla_\mathbf{x}f(\mathbf{x})=0$来求解最优点，对应于$\lambda=0$时拉格朗日函数优化，即$\nabla_\mathbf{x}L(\mathbf{x},\lambda=0)=\mathbf{0}$
- $g(\mathbf{x})=0$，约束起到作用，等价于上面讨论的等式约束，值得注意的是，在最优点$\mathbf{x}^*$处，$\nabla_\mathbf{x}g(\mathbf{x}^*)$方向必须与$\nabla_\mathbf{x}f(\mathbf{x}^*)$的方向必须相反，即$\lambda>0$。因为当方向相同时，只需将$\mathbf{x}^*$沿着$\nabla_\mathbf{x}f(\mathbf{x}^*)=-\lambda\nabla_\mathbf{x}g(\mathbf{x}^*)$的方向移动，设新的点为$\mathbf{x}^*+\boldsymbol{\epsilon}$，因为$\lambda<0$，新的点依然满足约束$g(\mathbf{x}+\boldsymbol{\epsilon})>g(\mathbf{x})\geq 0$，而此时由于沿着$f(\mathbf{x}^*)$的梯度方向移动，所以$f(\mathbf{x}+\boldsymbol{\epsilon})>f(\mathbf{x})$，即可以找到一个更优点来更新$\mathbf{x}^*$。

综合上述两个情况，在约束$g(\mathbf{x}) \geq 0$下对$f(\mathbf{x})$进行优化，可以转化成在如下条件下的拉格朗日函数（E.4）的优化问题

$$
\begin{aligned}
g(\mathbf{x}) &\geq 0 \\
\lambda &\geq 0 \\
\lambda g(\mathbf{x}) &= 0
\end{aligned}
\tag{E.6}
$$

这些条件被称为**KKT(Karush-Kuhn-Tucker)条件**。

上述方法可以推广到多个含等式和不等式的约束问题上，设我们希望最大化$f(\mathbf{x})$，其中包含J个等式约束$\{g_j(\mathbf{x}) = 0\}_{j=1}^{J}$和$K$个不等式约束$\{h_k(\mathbf{x}) \geq 0\}_{k=1}^{K}$，按照上述方法分别引入$J$个和$K$个拉格朗日乘子$\{\lambda_j\}_{j=1}^{J}, \{\mu\}_{k=1}^{K}$，优化的拉格朗日函数为

$$
L(\mathbf{x}, \{\lambda_j\}_{j=1}^{J}, \{\mu\}_{k=1}^{K}) = f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) + \sum_{k=1}^{K} \mu_k h_k(\mathbf{x}) \tag{E.7}
$$

其中引入的KKT条件$(k = 1, \ldots, K)$为

$$
\begin{aligned}
h_k(\mathbf{x}) &\geq 0 \\
\mu_k &\geq 0 \\
\mu_k h_k(\mathbf{x}) &= 0
\end{aligned}
\tag{E.8}
$$