## What

- 对于含有等式和不等式的最优化问题,

$$\min f(x)$$

$$\text{s.t. } g_j(x) \le 0 \quad (j=1,2,\cdots,m)$$

$$h_k(x)=0 \quad (k=1,2,\cdots,l)$$

KKT条件是判断$x^*$是否为最优解的必要条件,即:

$$\begin{cases} \dfrac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \dfrac{\partial g_j}{\partial x_i} + \sum_{k=1}^{l} \lambda_k \dfrac{\partial h_k}{\partial x_i} = 0, \quad (i=1,2,\cdots,n) \\ h_k(x)=0 \quad (k=1,2,\cdots,l) \\ \mu_j g_j(x) = 0 \quad (j=1,2,\cdots,m) \\ \mu_j \ge 0. \end{cases}$$

# 等式约束优化问题

参考文献 • 同济7版 高数下册（P.116-118）

what • $\min f(x_1, x_2, \cdots, x_n)$

$\text{s.t. } h_k(x_1, x_2, \cdots, x_n) = 0$

令 $L(x,\lambda) = f(x) + \sum_{k=1}^{l} \lambda_k h_k(x)$，函数 $L(x,y)$ 称为Lagrange函数，参数 $\lambda$ 称为Lagrange乘子.

再联立方程组
$$\begin{cases} \frac{\partial L}{\partial x_i} = 0 & (i=1,2,\cdots,n) \\ \frac{\partial L}{\partial \lambda_k} = 0 & (k=1,2,\cdots,l) \end{cases},$$

得到的解为可能的极值点，由于是必要条件，具体是否是极值点需根据问题本身的具体情况检验，

# 不等式约束优化问题

转换 • 不等式约束问题

$\downarrow$ 引入松弛变量

等式约束问题

$\downarrow$ 引入Lagrange乘子

无约束优化

举例 • $\min f(x)$ s.t. $\begin{cases} g_1(x) = a - x \le 0 \\ g_2(x) = x - b \le 0 \end{cases}$ 都是"$\le$"号

对于约束$g_1$与$g_2$加入松弛变量$a_1^2, b_1^2$有 $\begin{cases} h_1(x, a_1) = g_1 + a_1^2 = 0 \\ h_2(x, b_1) = g_2 + b_1^2 = 0 \end{cases}$

加入$a_1^2, b_1^2$而非$a_1, b_1$是因为$g_1, g_2$这两个不等式的左边必须加上一个正数才能将不等式变为等式.

对于等式约束有

$$L(x, a_1, b_1, \mu_1, \mu_2) = f(x) + \mu_1(a - x + a_1^2) + \mu_2(x - b + b_1^2)$$

$$\begin{cases} \frac{\partial L}{\partial x} = \frac{\partial f(x)}{\partial x} - \mu_1 + \mu_2 = 0 \qquad \frac{\partial L}{\partial \mu_1} = a - x + a_1^2 = 0 \\ \frac{\partial L}{\partial a_1} = 2\mu_1 a_1 = 0 \quad \frac{\partial L}{\partial b_1} = 2\mu_2 b_1 = 0 \qquad \frac{\partial L}{\partial \mu_2} = x - b + b_1^2 = 0 \end{cases}$$

对于$\mu_1 a_1 = 0$有 $\begin{cases} \mu_1 = 0, a_1 \ne 0; \text{ 此时 } g_1 \text{ 不起作用} \\ \mu_1 \ge 0, a_1 = 0, \text{ 此时 } g_1(x) = a - x = 0 \text{ 且 } \mu_1 > 0, \text{ 即 } g_1 \text{ 起作用, 且 } g_1(x) = 0 \end{cases}$

由此,方程组转化为

$$\begin{cases} \frac{df}{dx} + \mu_1 \frac{dg_1}{dx} + \mu_2 \frac{dg_2}{dx} = 0, \\ \mu_1 g_1(x) = 0, \quad \mu_2 g_2(x) = 0 \\ \mu_1 \ge 0, \quad \mu_2 \ge 0 \end{cases}$$

# 为什么KKT的乘子≥0

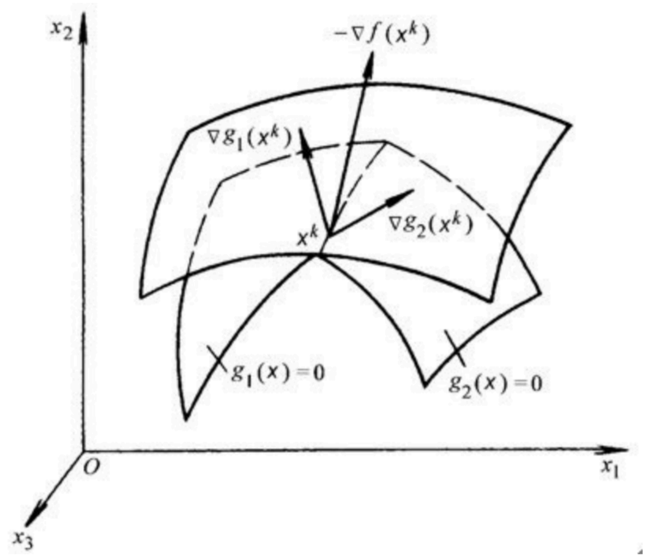
- 由于

$$\frac{\partial f(x^*)}{\partial x_i} + \sum_{j=1}^{m} \mu_j \frac{\partial g_j(x^*)}{\partial x_i} = 0 \quad (i=1,2,\cdots,n)$$

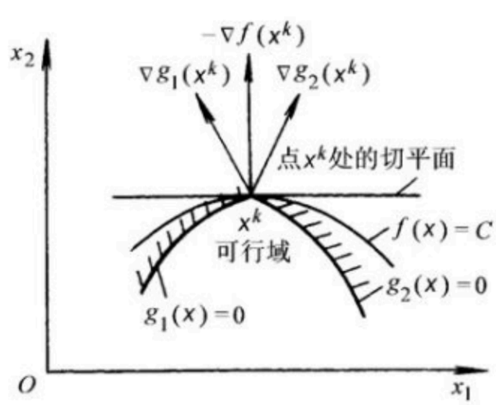
$$-\frac{\partial f(x^*)}{\partial x_i} = \sum_{j=1}^{m} \mu_j \frac{\partial g_j(x^*)}{\partial x_i}.$$

注意到梯度为向量，上式表示在约束极小值点$x^*$处，函数$f(x^*)$的负梯度一定可以表示成：所有起作用约束在该点的梯度的线性组合，

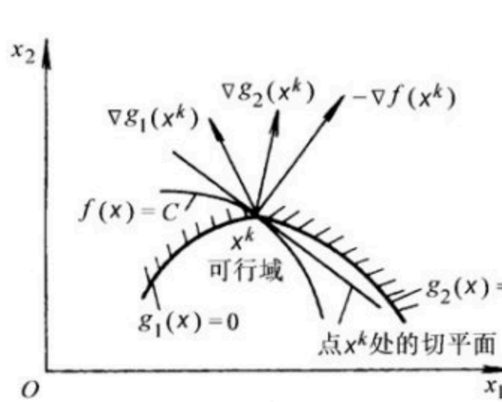
为方便作图，假设现在**只有两个起作用约束**，我们作出图形如下图.注意我们上面推导过，约束起作用时 $g_j(\mathbf{x})=0$，所以此时约束在几何上应该是一簇**约束平面**.我们假设在 $\mathbf{x}^*$ 取得极小值点，若同时满足 $g_1(\mathbf{x})=0$ 和 $g_2(\mathbf{x})=0$，则 $\mathbf{x}^k$ 一定在这两个平面的交线上，且 $-\nabla f(\mathbf{x}^*)=\sum_{j\in J}\mu_j\nabla g_j(\mathbf{x}^*)$，即 $-\nabla f(\mathbf{x}^k)$ 、$\nabla g_1(\mathbf{x}^k)$ 和 $\nabla g_2(\mathbf{x}^k)$ 共面.




下图是在点 $\mathbf{x}^k$ 处沿 $x_1Ox_2$ 面的截面，过点 $\mathbf{x}^k$ 作目标函数的负梯度 $-\nabla f(\mathbf{x}^k)$，它垂直于目标函数的等值线 $f(\mathbf{x})=C$（高数课本：一点的梯度与等值线相互垂直），且指向目标函数 $f(\mathbf{x})$ 的最速减小方向.再作约束函数 $g_1(\mathbf{x})=0$ 和 $g_2(\mathbf{x})=0$ 的梯度 $\nabla g_1(\mathbf{x}^k)$ 和 $\nabla g_2(\mathbf{x}^k)$，它们分别垂直 $g_1(\mathbf{x})=0$ 和 $g_2(\mathbf{x})=0$ 两曲面在 $\mathbf{x}^k$ 的切平面，并形成一个锥形夹角区域.此时，可能有a、b两种情形：

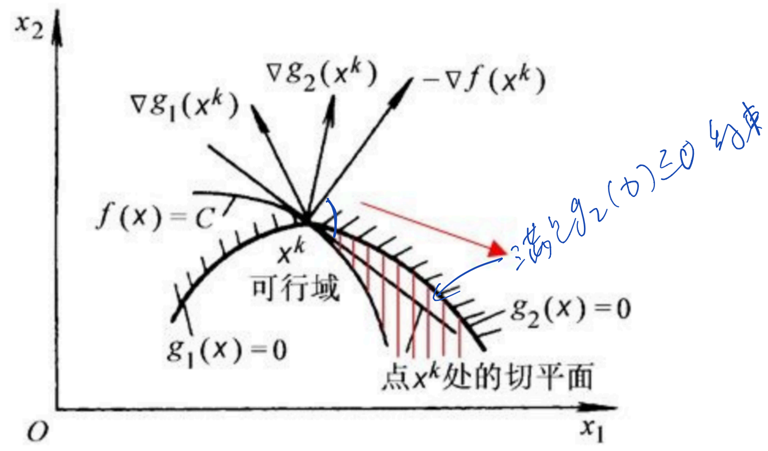



a)




b)

我们先来看情形b：若3个向量的位置关系如b所示，即 $-\nabla f$ 落在 $\nabla g_1$ 和 $\nabla g_2$ 所形成的锥角区外的一侧. 此时，作等值面 $f(\mathbf{x})=C$ 在点 $\mathbf{x}^k$ 的切平面（它与 $-\nabla f\left(\mathbf{x}^k\right)$ 垂直），我们发现：**沿着与负梯度 $-\nabla f$ 成锐角的方向移动（如下图红色箭头方向）**，只要在红色区域取值，目标函数 $f(\mathbf{x})$ 总能减小.而红色区域是可行域（ $f(\mathbf{x})=C$，C取不同的常数能得到不同的等值线，因此能取到红色区域），因此既可减小目标函数值，又不破坏约束条件. 这说明 $\mathbf{x}^k$ 仍可沿约束曲面移动而不破坏约束条件，且目标函数值还能够减小.所以 $\mathbf{x}^k$ 不是稳定的最优点，即不是局部极值点.



b)

反过头来看情形a：$-\nabla f$ 落在 $\nabla g_1$ 和 $\nabla g_2$ 形成的锥角内. 此时，同样作 $f(\mathbf{x})=C$ 在点 $\mathbf{x}^k$ 与 $-\nabla f$ 垂直的切平面. 当从 $\mathbf{x}^k$ 出发**沿着与负梯度 $-\nabla f$ 成锐角的方向移动**时，**虽然能使目标函数值减小，但此时任何一点都不在可行区域内.** 显然，此时 $\mathbf{x}^k$ 就是局部最优点 $\mathbf{x}^*$，再做任何移动都将破坏约束条件，故它是稳定点.

why $\geq 0$ ?

由于 $-\nabla f(\mathbf{x}^*)$ 和 $\nabla g_1(\mathbf{x}^*)$、$\nabla g_2(\mathbf{x}^*)$ 在一个平面内，所以前者可看成是后两者的线性组合. 又由上面的几何分析知，$-\nabla f(\mathbf{x}^*)$ 在 $\nabla g_1(\mathbf{x}^*)$ 和 $\nabla g_2(\mathbf{x}^*)$ 的夹角之间，所以线性组合的系数为正，有

$$-\nabla f(\mathbf{x}^*)=\mu_1\nabla g_1(\mathbf{x}^*)+\mu_2\nabla g_2(\mathbf{x}^*)\text{，且 }\mu_1>0\text{，}\mu_2>0\,.$$

这就是 $\mu_j>0$ 的原因. 类似地，当有多个不等式约束同时起作用时，要求 $-\nabla f(\mathbf{x}^*)$ 处于 $\nabla g_j(\mathbf{x}^*)$ 形成的超角锥（高维图形，我姑且称之为“超”）之内.

### 3.总结：同时包含等式和不等式约束的一般优化问题

$$\min f(\mathbf{x})$$
$$s.t.\, g_j(\mathbf{x}) \leq 0(j=1,2,\cdots,m)$$
$$h_k(\mathbf{x}) = 0(k=1,2,\cdots,l)$$

**KKT条件（$\mathbf{x}^*$ 是最优解的必要条件）为**

$$\begin{cases} \frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \mu_j \frac{\partial g_j}{\partial x_i} + \sum_{k=1}^{l} \lambda_k \frac{\partial h_k}{\partial x_i} = 0,\ (i=1,2,\ldots,n) \\ h_k(\mathbf{x}) = 0, (k=1,2,\cdots,l) \\ \mu_j g_j(\mathbf{x}) = 0, (j=1,2,\cdots,m) \\ \mu_j \geq 0. \end{cases}$$

注意，**对于等式约束的Lagrange乘子，并没有非负的要求！以后求其极值点，不必再引入松弛变量，直接使用KKT条件判断！**

这样，我们就推导完了KKT条件。各位看官可以自己分别罗列并比较一下：无约束优化、等式约束优化和等式+不等式约束优化条件下某点为局部最优解（极值点）的必要条件。