运筹学

12. 有约束的非线性规划

李 力
清华大学

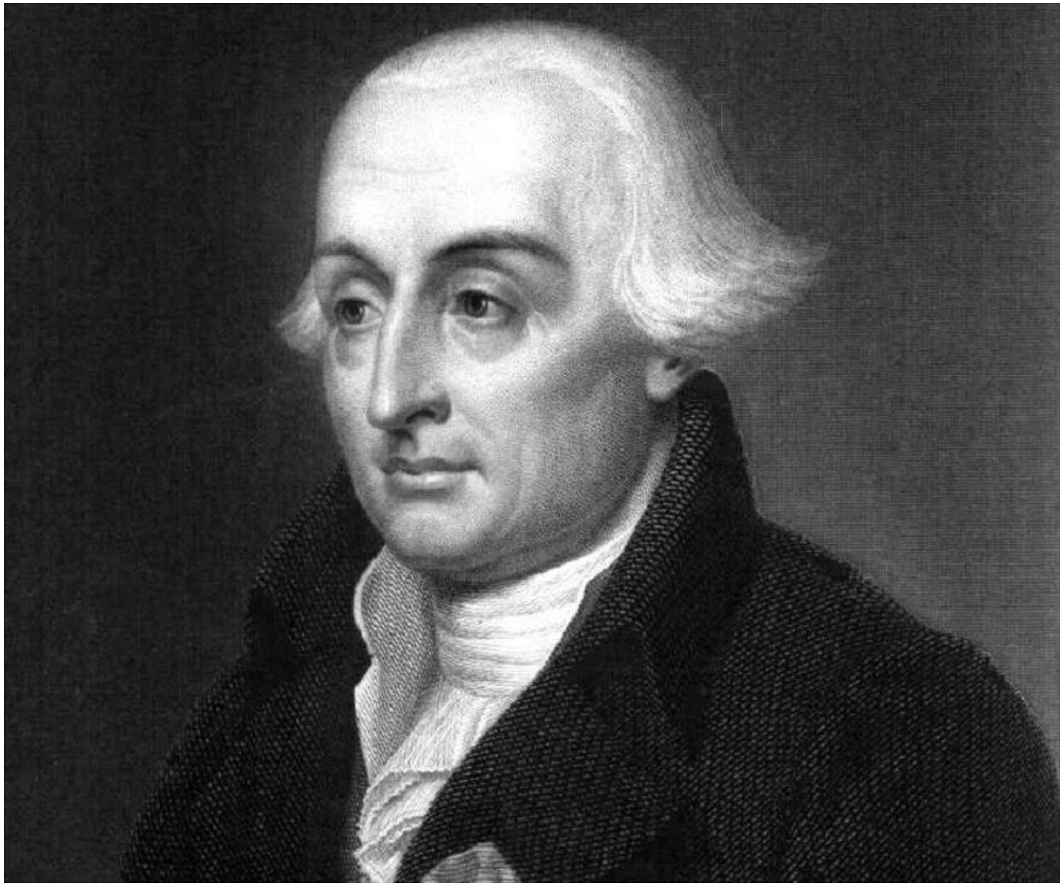
Email: li-li@tsinghua.edu.cn

2023.12.

在本科生《运筹学》中，我们主要关心：

1. **有约束的非线性规划最优解应该满足的性质（特别是KKT条件）**
2. **线性约束和线性不等式约束下的非线性规划问题解法**

拉格朗日条件

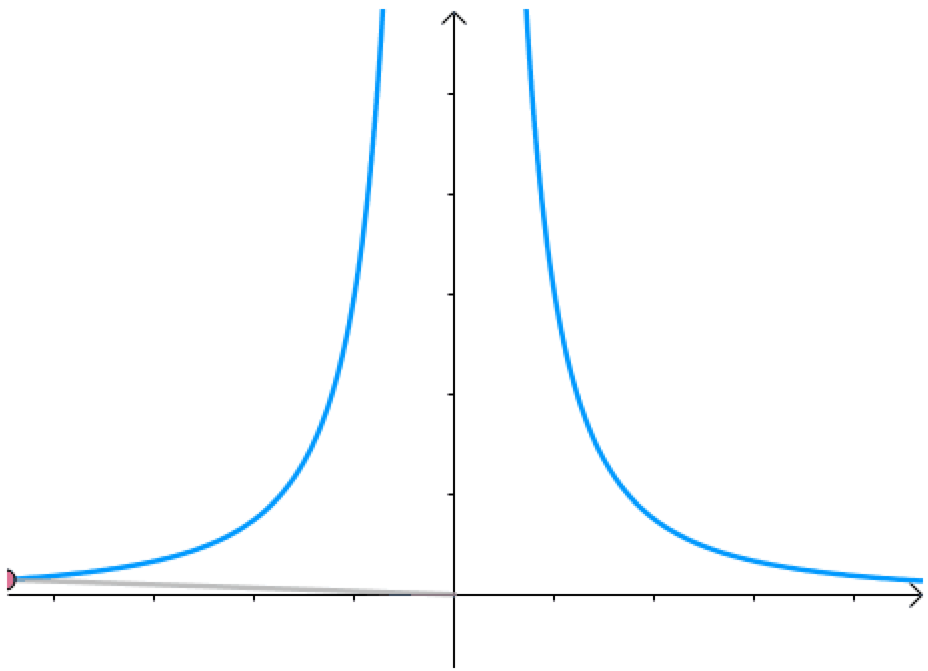


Joseph-Louis Lagrange

Jan. 25, 1736－April 11, 1813

拉格朗日条件的直观观察

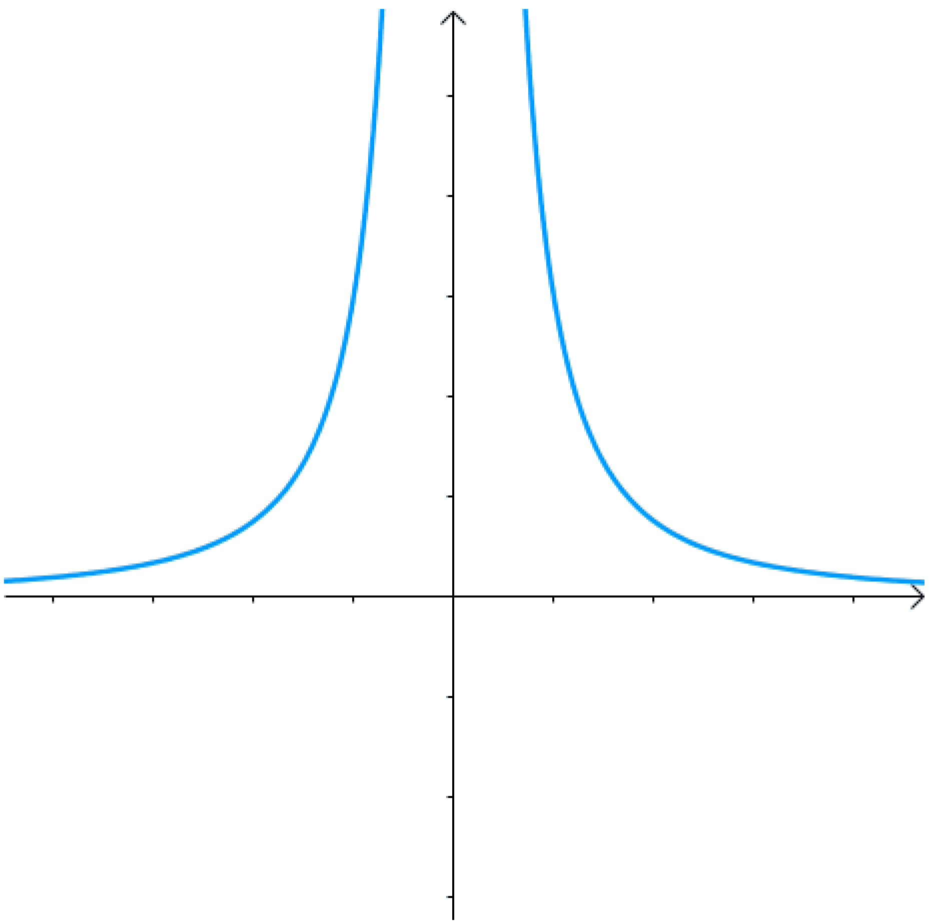
- 考虑x^2 * y-3=0 和原点的距离的最小值问题



$$\min \ x_1^2 + x_2^2$$
$$\text{s.t.} \ x_1^2 x_2 = 3$$

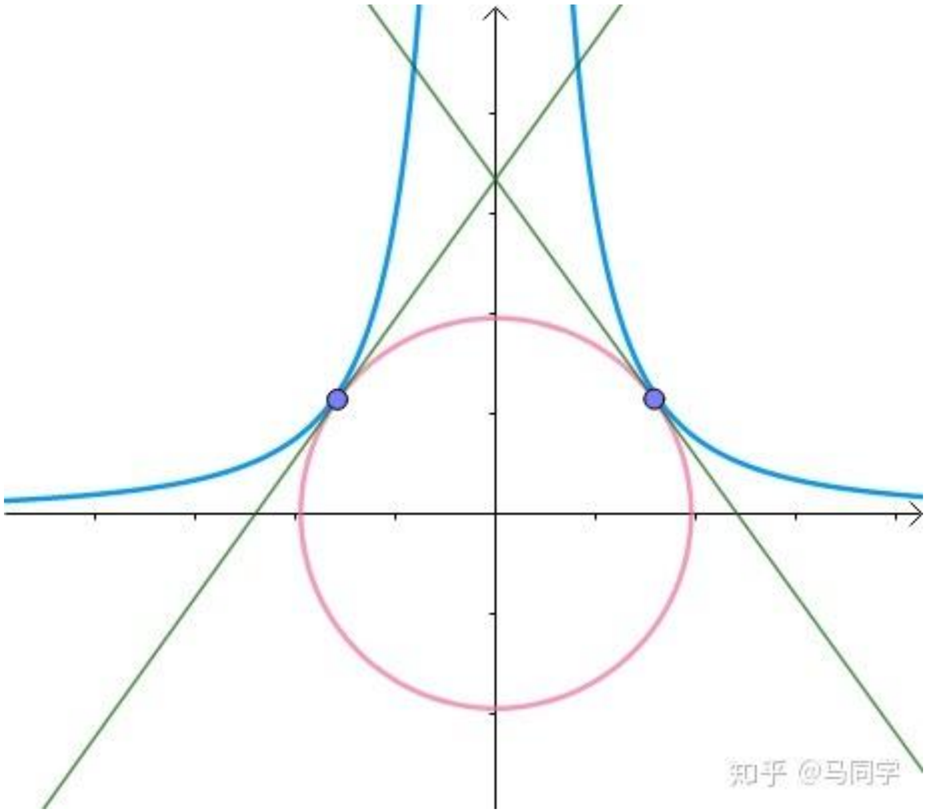
拉格朗日条件的直观观察

- 转为考虑x^2 * y-3=0 何时与以原点为圆心的圆相切



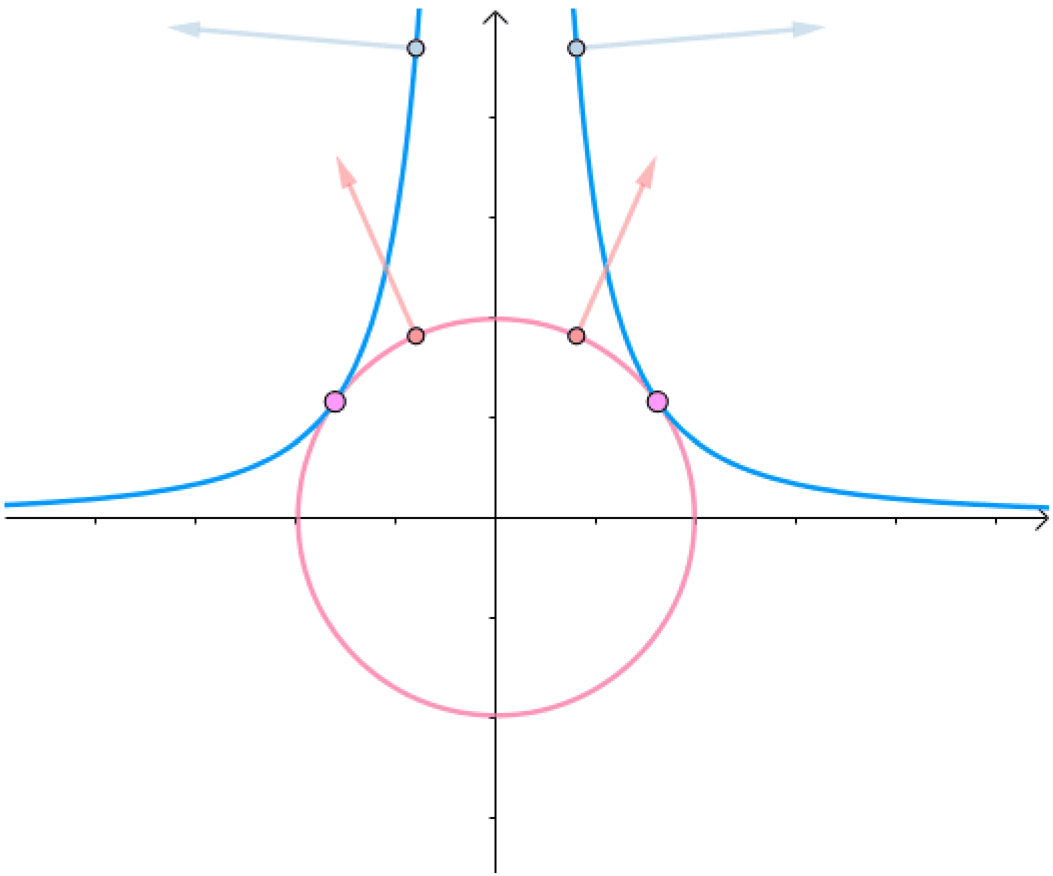
拉格朗日条件的直观观察

- $x^2 * y-3=0$ 与以原点为圆心的圆**相切时有共同切线**



拉格朗日条件的直观观察

- $x^2 * y-3=0$ 与以原点为圆心的圆**相切时法线共线**



等式约束的最优解的拉格朗日条件

$\min\left\{ f(X) \middle| \text{ s.t. } h_j(X)=0, 1\le j\le m \right\}$ 的最优解必须满足

$$\frac{\partial L(X,Y)}{\partial X}=0,\quad \frac{\partial L(X,Y)}{\partial Y}=0$$

其中 $L(X,Y)=f(X)+\sum_{j=1}^{m} y_j h_j(X)$ 被称为拉格朗日函数

只要求$\mathbf{y_j}$是实数

例 1、

$$\min \ (x_1-1)^2+(x_2-2)^2$$
$$\text{s.t.}\quad x_1-x_2=0$$

拉格朗日条件： $\dfrac{\partial L(X,Y)}{\partial X}=0,\ \dfrac{\partial L(X,Y)}{\partial Y}=0$

其中 $L(X,y)=(x_1-1)^2+(x_2-2)^2+y(x_1-x_2)$

最优解必须满足的方程

$$2(x_1-1)+y=0,\ 2(x_2-2)-y=0,\ x_1-x_2=0$$

解上述方程可得

$$x_1=1.5,\ x_2=1.5,\ y=-1$$

例 2、

$$
\begin{aligned}
\min\ & \left(x_1-1\right)^2+\left(x_2-2\right)^2 \\
\text{s.t.}\ & x_1-x_2=0 \\
& x_1+x_2-2\le 0 \\
& -x_1\le 0,\ -x_2\le 0
\end{aligned}
$$

最优解应该满足什么方程组？

KKT条件

等式约束的最优解的拉格朗日条件

$$\min\left\{ f(X) \middle| \text{ s.t. } h_j(X)=0, 1\le j\le m \right\}$$ 的最优解必须满足

$$\frac{\partial L(X,Y)}{\partial X}=0,\quad \frac{\partial L(X,Y)}{\partial Y}=0$$

其中 $L(X,Y)=f(X)+\sum_{j=1}^{m} y_j h_j(X)$ 被称为拉格朗日函数

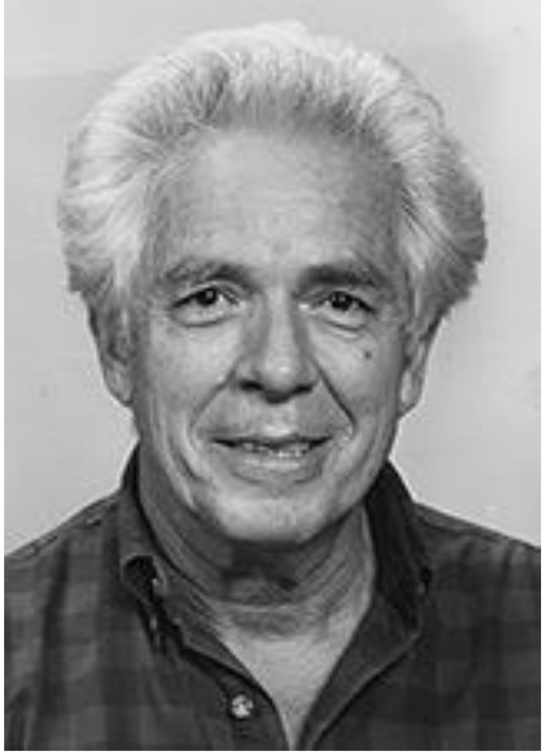
$$\Downarrow$$

包含不等式约束的最优解的Kuhn-Tucker条件

（上世纪50年代的工作）

William Karush

(1917–1997)



Harold W. Kuhn

(1925–2014)

Albert W. Tucker

(1905–1995)

答案：对于同时含有等式和不等式约束的问题

$$\min\left\{ f(X) \middle| \text{ s.t. } h_j(X)=0, 1\le j\le m,\ \ g_i(X)\le 0, 1\le i\le l \right\}$$

将其拉格朗日函数记为

$$L(X,U,V)=f(X)+\sum_{j=1}^{m}u_j h_j(X)+\sum_{i=1}^{l}v_i g_i(X)$$

其最优解（在一定条件下）必须满足

$$\frac{\partial L(X,U,V)}{\partial X}=0,\quad \frac{\partial L(X,U,V)}{\partial U}=0,\quad \frac{\partial L(X,U,V)}{\partial V}\le 0$$

$$v_i\ge 0,\ \ v_i\frac{\partial L(X,U,V)}{\partial v_i}=0,\quad i=1,2,\cdots,l$$

对于例 2

$$\begin{aligned} \min\ & (x_1-1)^2+(x_2-2)^2 \\ \text{s.t.}\ & x_1-x_2=0 \\ & x_1+x_2-2\le 0 \\ & -x_1\le 0,\ -x_2\le 0 \end{aligned}$$

拉格朗日函数

$$\begin{aligned} L(X,u,V)=&(x_1-1)^2+(x_2-2)^2+u(x_1-x_2) \\ &+v_1(x_1+x_2-2)-v_2x_1-v_3x_2 \end{aligned}$$

对应方程

$$2(x_1-1)+u+v_1-v_2=0,\ 2(x_2-2)-u+v_1-v_3=0$$

$$x_1-x_2=0,\ x_1+x_2-2\le 0,\ -x_1\le 0,\ -x_2\le 0$$

$$v_i\ge 0, i=1,2,3,\ v_1(x_1+x_2-2)=0,\ v_2x_1=0,\ v_3x_2=0$$

求解方程 $2(x_1-1)+u+v_1-v_2=0,\ 2(x_2-2)-u+v_1-v_3=0$

$$x_1-x_2=0,\ x_1+x_2-2\le 0,\ -x_1\le 0,\ -x_2\le 0$$

$$v_i\ge 0, i=1,2,3,\ v_1(x_1+x_2-2)=0,\ v_2x_1=0,\ v_3x_2=0$$

具体方法：考虑 v_i 大于和等于零的各种组合

例如，假设 $v_1>0, v_2=0, v_3=0$，可得等式方程组

$$2(x_1-1)+u+v_1-v_2=0,\ 2(x_2-2)-u+v_1-v_3=0$$

$$x_1-x_2=0,\ x_1+x_2-2=0, v_2=0, v_3=0$$

方程数等于变量数，可以求得 $x_1=x_2=1, v_1=1, u=-1$

由于满足 $-x_1\le 0, -x_2\le 0, v_1\ge 0$，所以确实是一个解

Kuhn-Tucker定理（Karush-Kuhn-Tucker定理）

如果 $\hat{X}$ 是下述问题的最优解

$$\min\left\{ f(X) \middle| \text{ s.t. } h_j(X)=0, 1\le j\le m,\ \ g_i(X)\le 0, 1\le i\le l \right\}$$

并且在 $\hat{X}$ 处等式约束和所有起作用的不等式约束的梯度线性无关，则一定存在 $\hat{u}_j, 1\le j\le m$ 和 $\hat{v}_i \ge 0, 1\le i\le l$ 满足

$$\nabla f\left(\hat{X}\right)+\sum_{j=1}^{m}\nabla h_j\left(\hat{X}\right)\hat{u}_j+\sum_{i=1}^{l}\nabla g_i\left(\hat{X}\right)\hat{v}_i=0 \qquad （梯度条件）$$

$$\hat{v}_i g_i(\hat{X})=0,\ \ \forall 1\le i\le l \qquad （互补松弛条件）$$

不等式约束在给定点的分类及其作用

设 $\hat{X}$ 满足一般性不等式约束，即

$$g_i(\hat{X}) \geq 0, \forall 1 \leq i \leq l$$

对任意的 $1 \leq j \leq l$，若 $g_j(\hat{X}) = 0$，称 $g_j(X) \geq 0$ 是 $\hat{X}$ 处起作用约束，若 $g_j(\hat{X}) > 0$，则称其是 $\hat{X}$ 处不起作用约束

如果 $g_j(X) \geq 0$ 是 $\hat{X}$ 处不起作用的约束，则对任意的 $D \in R^n$，都存在 $\hat{t} > 0$ 满足

$$g_j(\hat{X} + tD) > 0, \forall 0 \leq t \leq \hat{t}$$

所以，构造可行方向时不用考虑不起作用约束

对起作用约束指标集的约定

对任何满足一般性不等式（包括线性不等式）约束的可行解 $\hat{X}$，为讨论方便，只要不特别指明，我们总是假定其前 $\hat{l}$ 个约束是起作用约束，其它约束是不起作用约束，即

$$g_i(\hat{X}) = 0, \forall 1 \le i \le \hat{l}$$ 构造可行方向时需考虑

$$g_i(\hat{X}) > 0, \forall \hat{l} + 1 \le i \le l$$

有关术语

在 $\hat{X}$ 处起作用的不等式约束： $g_i(\hat{X}) = 0$ 的不等式约束

在 $\hat{X}$ 处不起作用的不等式约束：$g_i(\hat{X}) < 0$ 的不等式约束

对起作用约束指标集的约定：对任何可行解 $\hat{X}$ ，总是假定前 $\hat{l}$ 个约束是起作用约束，其它约束不起作用，即

$$g_i(\hat{X}) = 0, \forall 1 \le i \le \hat{l}, g_i(\hat{X}) < 0, \forall \hat{l} + 1 \le i \le l$$

定理中的有关梯度线性无关是指以下向量组线性无关：

$$\nabla h_j\left(\hat{X}\right), 1 \le j \le m, \nabla g_i\left(\hat{X}\right), 1 \le i \le \hat{l}$$

$$\nabla f\left(\hat{X}\right)+\sum_{j=1}^{m}\nabla h_j\left(\hat{X}\right)\hat{u}_j+\sum_{i=1}^{l}\nabla g_i\left(\hat{X}\right)\hat{v}_i=0$$

（<u>梯度条件</u>）

$$\hat{v}_i g_i(\hat{X})=0,\ \forall 1\le i\le l$$

（<u>互补松弛条件</u>）

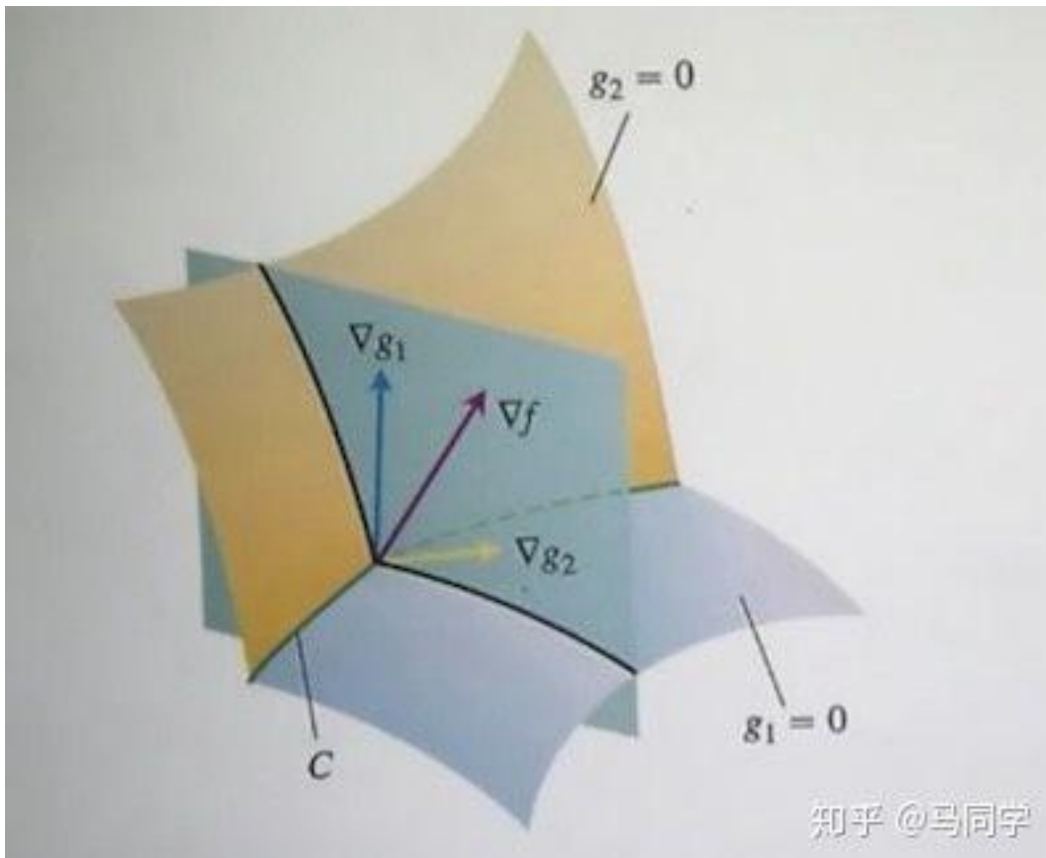
简单说，就是在极值点处， f 的梯度是一系列等式约束 h_j 的梯度和不等式约束 g_i 的梯度的线性组合。在这个线性组合中，等式约束梯度的权值 λ_j 没有要求；不等式约束梯度的权值 μ_i 是非负的，并且如果某个 $g_i(x^*)$ 严格小于0，那这个约束不会出现在加权式中，因为对应的权值 μ_i 必须为0。换句话说，只有 x^* 恰好在边界 $g_i=0$ 上的那些 g_i 的梯度才会出现在加权式中。如果去掉不等式约束的部分，那么上式就是拉格朗日乘子法的精确表述。

$$\nabla f\left(\hat{X}\right)+\sum_{j=1}^{m}\nabla h_j\left(\hat{X}\right)\hat{u}_j+\sum_{i=1}^{l}\nabla g_i\left(\hat{X}\right)\hat{v}_i=0$$

（梯度条件）

$$\hat{v}_i g_i(\hat{X})=0,\ \forall 1\le i\le l$$

（互补松弛条件）



给定一个优化问题，我们把**满足所有约束条件的n维空间区域称为可行域。**从可行域中的每一个点 x 朝某个方向 v 出发走一点点，如果还在可行域中，或者偏离可行域的程度很小，准确地说，偏移量是行进距离的高阶无穷小量，那么我们就说 v 是一个**可行方向**。我们用 $F(x)$ 表示点 x 的所有可行方向的集合。

对于可行域中的一个极大值点 x^* ，它的可行方向集合为 $F(x^*)$ ，从 x^* 朝 $F(x^*)$ 中的某个方向走一小步，那么落点仍然（近似）在可行域中。 x^* 是局部最大值点就意味着在这些可行方向上目标函数 $f(x)$ 不能增大，从而我们得到下面这样一个结论：

在极值点 x^* ，让目标函数增大的方向不能在 $F(x^*)$ 中。

把这句话精确表达出来，就得到了KKT条件的数学表述。

满足等式 $f(x)=0$ 的点 x ，在n维空间里构成了一个(通常是光滑的)曲面，

这个曲面是n-1维的，因为自变量的n个分量只满足一个方程，因而曲面上有n-1个自由度。例如在二维空间里 $y-x=0$ 是一条一维的直线，在三维空间中 $x^2+y^2+z^2-1=0$ 是一张二维曲面。

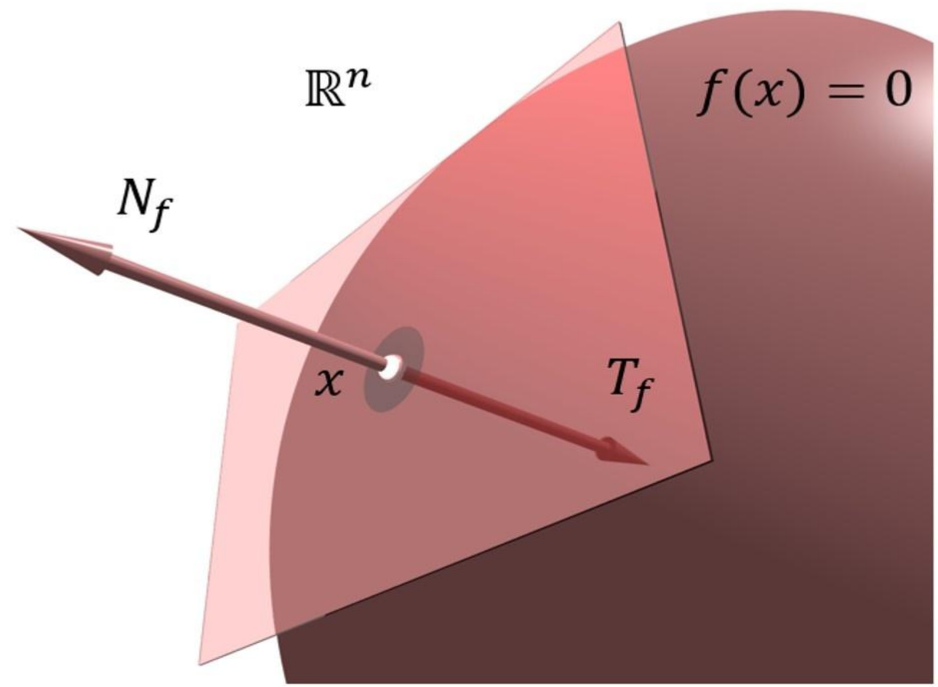
在这些曲面上每一点 x 都存在一个梯度方向 $\nabla f(x)$ ，沿着这个方向函数值 $f(x)$ 增大，沿着它的逆方向 $-\nabla f(x)$ 函数值减小，这两个方向我们记为

$$N_f^+(x)=\{\lambda\nabla f(x),\lambda\geq 0\}$$

$$N_f^-(x)=\{\lambda\nabla f(x),\lambda\leq 0\}$$

$$N_f=N_f^+\cup N_f^-$$

N_f 是两条射线的并，是由梯度方向张成的一维子空间。与这个一维子空间正交的n-1一维空间，被称为切空间或者切平面 $T_f(x)$ ，自变量在 $T_f(x)$ 中的方向上移动时，函数值 f 变化不大。我们都知道对于光滑的曲面，如果把它在一点 x 上无限放大，最终会近似成为平面，这个平面就是 $x+T_f(x)$ 。



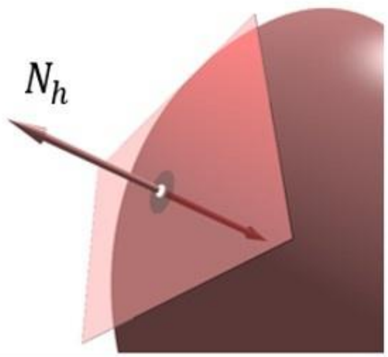
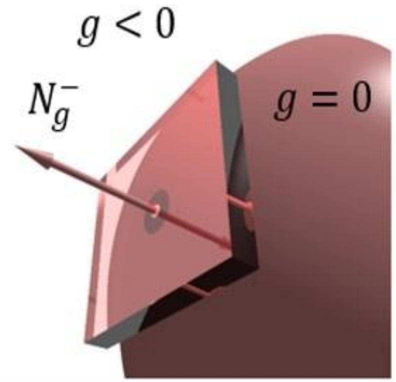
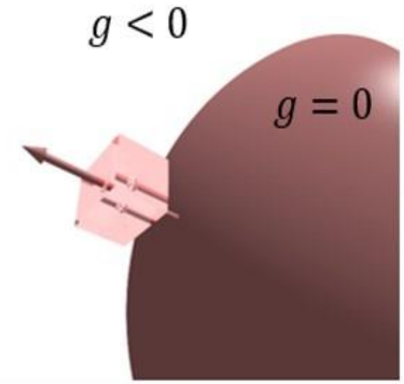
下面来看看等式约束和不等式约束的可行方向，也就是这样一些方向，朝它们移动一点，约束依然近似满足。

1 如果点 x^* 满足等式约束 $h(x^*)=0$ ，那么在切平面方向 T_h 上滑动，这个等式约束依然近似满足，所以可行方向为 T_h 。

2 如果点 x^* 满足不等式约束 $g(x^*)<0$ ，说明它在 曲面 $g(x)=0$ 的某一侧，并且不在边界上，所以无论朝哪个方向跑，它仍然满足约束 $g(x)\leq 0$ ，可行方向为 R^n 。

3 如果点 x^* 满足 $g(x^*)=0$ ，因为约束条件是 $g(x)\leq 0$ ， x^* 可以朝切平面 T_g 中的方向滑动，保持在 $g(x)=0$ 上，也可以往 g 的负梯度方向 N_g^- 跑，让 g 值减小，不等式约束依然成立。但是不能往正梯度方向跑，因为会造成 $g(x)>0$ ，违背不等式约束。所以可行方向是 $N_g^-\cup T_g$ ，这是n维空间的一半，或者n-1维子空间与一条垂直于它的射线的并。

总结一下就是：假设一个点满足某约束，可行方向是这样一些方向，在这些方向上移动一点点，点仍然近似满足约束。对等式约束，可行方向是切空间；对不等式约束，如果点正好在边界上，可行方向是半空间，如果不在边界上，可行方向是全空间。

约束	$h=0$	$g\leq 0$	$g\leq 0$
点$\boldsymbol{x}$满足条件	$h(\boldsymbol{x})=0$	$g(\boldsymbol{x})=0$	$g(\boldsymbol{x})<0$
可行方向	T_h	$T_g\cup N_g^-$	$T_g\cup N_g\ =\mathbb{R}^n$
可行方向维数	n-1维子空间 	n维半空间 	全空间 
可行方向的补	N_h	N_g^+	0

KKT条件证明

（一种）构造性证明

第一步、选择充分小的正数 ε 构造邻域

$$B\left(\hat{X},\varepsilon\right)=\left\{X\in R^n \,\middle|\, \left\|X-\hat{X}\right\|\le\varepsilon\right\}$$

要求该邻域满足，1）局部最优性：

$$f\left(\hat{X}\right)\le f\left(X\right),\forall X\in B\left(\hat{X},\varepsilon\right)$$

2）不起作用约束集不变：

$$g_i(X)<0,\forall X\in B(\hat{X},\varepsilon),\hat{l}+1\le i\le l$$

由 $\hat{X}$ 和 $\hat{l}$ 的定义可知存在这样的 ε

第二步、构造 $B\left(\hat{X},\varepsilon\right)$ 上的（无约束）优化问题

$$\min_{X\in B\left(\hat{X},\varepsilon\right)} \quad F_k^{(\hat{X})}(X)$$

其中，目标函数 $F_k^{(\hat{X})}(X)$ 定义为

$$f(X)+k\sum_{j=1}^{m}\varphi\left(h_j(X)\right)+k\sum_{i=1}^{l}\rho\left(g_i(X)\right)+\left\|X-\hat{X}\right\|^2$$

k 是给定的正整数，$\varphi(\cdot)$ 和 $\rho(\cdot)$ 分别是对等式约束和不等式约束的（可导）罚函数，定义为

$$\varphi(u)=u^2,\quad \rho(u)=\left(\max\{0,u\}\right)^2=\begin{cases}u^2 & \forall u>0\\ 0 & \forall u\le 0\end{cases}$$

第三步、通过求解（无约束）问题得到收敛于$\hat{X}$ 的点列

用 X_k表示（无约束）问题的一个最优解

由 $\hat{X} \in B\left(\hat{X}, \varepsilon\right)$ 可得

$$f(X_k) + k\sum_{j=1}^{m}\varphi\left(h_j(X_k)\right) + k\sum_{i=1}^{l}\rho\left(g_i(X_k)\right) + \left\|X_k - \hat{X}\right\|^2$$

$$\leq f(\hat{X}) + k\sum_{j=1}^{m}\varphi\left(h_j(\hat{X})\right) + k\sum_{i=1}^{l}\rho\left(g_i(\hat{X})\right) + \left\|\hat{X} - \hat{X}\right\|^2$$

$$= f(\hat{X})$$

由此又可得 $f(X_k) + \left\|X_k - \hat{X}\right\|^2 \leq f(\hat{X}), \forall k$

$$\lim_{k\to\infty}\varphi\left(h_j(X_k)\right) = 0, \forall j \quad \lim_{k\to\infty}\rho\left(g_i(X_k)\right) = 0, \forall i$$

第三步继续

由于 $X_k \in B\left(\hat{X},\varepsilon\right), \forall k$，在 $\{X_k\}_{k=1}^{\infty}$ 中一定存在

子列收敛于某个 $\bar{X} \in B\left(\hat{X},\varepsilon\right)$，记为 $\left\{X_{k_t}\right\}_{t=1}^{\infty}$

利用 $f(X_{k_t}) + \left\|X_{k_t} - \hat{X}\right\|^2 \le f(\hat{X}), \forall t$

$$\lim_{t\to\infty} \rho\left(g_i(X_{k_t})\right) = 0, \forall i \quad \lim_{t\to\infty} \phi\left(h_j(X_{k_t})\right) = 0, \forall j$$

可以得到 $f(\bar{X}) + \left\|\bar{X} - \hat{X}\right\|^2 \le f(\hat{X})$

$$\rho\left(g_i(\bar{X})\right) = 0, \forall i, \ \varphi\left(h_j(\bar{X})\right) = 0, \forall j$$

由此可知：$\bar{X}$ 是原问题的可行解

第三步继续

因为 $\bar{X} \in B\left(\hat{X},\varepsilon\right)$ 是原问题的可行解，$\hat{X}$ 是原问题在 $B\left(\hat{X},\varepsilon\right)$ 中的最优解，所以 $f(\hat{X}) \le f(\bar{X})$，再结合前面得到 $f(\bar{X}) + \left\|\bar{X} - \hat{X}\right\|^2 \le f(\hat{X})$，可得

$$f(\bar{X}) + \left\|\bar{X} - \hat{X}\right\|^2 \le f(\bar{X})$$

由此可知 $\left\|\bar{X} - \hat{X}\right\|^2 = 0$，因此 $\bar{X} = \hat{X}$，这意味着

$$\lim_{t\to\infty} X_{k_t} = \hat{X}$$

其中每个 X_{k_t} 都是 $\min\limits_{X\in B\left(\hat{X},\varepsilon\right)} F_{k_t}^{(\hat{X})}(X)$ 的最优解

第四步、利用（无约束）问题必要条件得到所需结果

因为 $\hat{X}$ 是 $B\left(\hat{X},\varepsilon\right)$ 的内点，所以存在正整数 $\hat{t}$，满足所有 $\left\{X_{k_t}\right\}_{t=\hat{t}}^{\infty}$ 都是 $B\left(\hat{X},\varepsilon\right)$ 的内点，因此对所有的 $t \geq \hat{t}$ 都成立 $\nabla F_{k_t}^{(\hat{X})}(X_{k_t})=0$，即

$$\nabla f(X_{k_t})+2k_t\sum_{j=1}^{m}h_j(X_{k_t})\nabla h_j(X_{k_t})$$

$$+2k_t\sum_{i=1}^{l}\max\left\{0,g_i(X_{k_t})\right\}\nabla g_i(X_{k_t})$$

$$+2\left(X_{k_t}-\hat{X}\right)=0$$

第四步继续

注意到 $X_{k_t} \in B(\hat{X},\varepsilon)$, $\forall t \ge \hat{t}$，根据定义又有

$$g_i(X) < 0,\ \forall X \in B(\hat{X},\varepsilon),\ \hat{l}+1 \le i \le l$$

所以 $\max\left\{0, g_i(X_{k_t})\right\} = 0,\ \forall \hat{l}+1 \le i \le l$，前面等式可写成

$$\nabla f(X_{k_t}) + 2k_t \sum_{j=1}^{m} h_j(X_{k_t}) \nabla h_j(X_{k_t})$$

$$+ 2k_t \sum_{i=1}^{\hat{l}} \max\left\{0, g_i(X_{k_t})\right\} \nabla g_i(X_{k_t})$$

$$+ 2\left(X_{k_t} - \hat{X}\right) = 0$$

第四步继续

记 $A_{k_t} = \left(\nabla h_1(X_{k_t}),\cdots,\nabla h_m(X_{k_t}),\nabla g_1(X_{k_t}),\cdots,\nabla g_{\hat{l}}(X_{k_t})\right)$

$$u_j^{k_t} = 2k_t h_j(X_{k_t})$$

$$v_i^{k_t} = 2k_t \max\left\{0, g_i(X_{k_t})\right\}$$

$$Y_{k_t} = \left(u_1^{k_t},\cdots,u_{\hat{l}}^{k_t},v_1^{k_t},\cdots,v_m^{k_t}\right)^T$$

则有 $v_i^{k_t} \geq 0, \forall 1 \leq i \leq \hat{l}, t \geq \hat{t}$，前面的等式可写成

$$\nabla f(X_{k_t}) + \sum_{j=1}^{m} u_j^{k_t} \nabla h_j(X_{k_t}) + \sum_{i=1}^{\hat{l}} v_i^{k_t} \nabla g_i(X_{k_t}) + 2\left(X_{k_t} - \hat{X}\right)$$

$$= \nabla f(X_{k_t}) + A_{k_t} Y_{k_t} + 2\left(X_{k_t} - \hat{X}\right) = 0$$

第四步继续

由 $\nabla f(X_{k_t})+A_{k_t}Y_{k_t}+2\left(X_{k_t}-\hat{X}\right)=0$ 可得

$$A_{k_t}{}^T A_{k_t}Y_{k_t}=-A_{k_t}{}^T\nabla f(X_{k_t})-2A_{k_t}{}^T\left(X_{k_t}-\hat{X}\right)$$

记 $\hat{A}=\left(\nabla g_1(\hat{X}),\cdots,\nabla g_{\hat{l}}(\hat{X}),\nabla h_1(\hat{X}),\cdots,\nabla h_m(\hat{X})\right)$

根据给定条件，$\hat{A}$是列满秩矩阵，$\left(\hat{A}^T\hat{A}\right)^{-1}$存在，又因

$\lim\limits_{t\to\infty}A_{k_t}=\hat{A}$，所以对充分大的 t 存在 $\left(A_{k_t}{}^T A_{k_t}\right)^{-1}$，此时

$$Y_{k_t}=-\left(A_{k_t}{}^T A_{k_t}\right)^{-1}\left(A_{k_t}{}^T\nabla f(X_{k_t})+2A_{k_t}{}^T\left(X_{k_t}-\hat{X}\right)\right)$$

由此可知 $\lim\limits_{t\to\infty}Y_{k_t}$ 存在，记为 $\hat{Y}=\left(\hat{u}_1,\cdots,\hat{u}_{\hat{l}},\hat{v}_1,\cdots,\hat{v}_m\right)^T$

第四步继续

在 $\nabla f(X_{k_t}) + \sum_{j=1}^{m} u_j^{k_t} \nabla h_j(X_{k_t}) + \sum_{i=1}^{\hat{l}} v_i^{k_t} \nabla g_i(X_{k_t}) + 2\left(X_{k_t} - \hat{X}\right) = 0$

左边令 $t \to \infty$ ，可得

$$\nabla f(\hat{X}) + \sum_{j=1}^{m} \hat{u}_j \nabla h_j(\hat{X}) + \sum_{i=1}^{\hat{l}} \hat{v}_i \nabla g_i(\hat{X}) = 0$$

其中 $\hat{v}_i = \lim_{t\to\infty} 2k_t \max\left\{0, g_i(X_{k_t})\right\} \geq 0$，再令

$$\hat{v}_i = 0, \forall \hat{l} + 1 \leq i \leq l$$

最终可得 $\nabla f\left(\hat{X}\right) + \sum_{j=1}^{m} \nabla h_j\left(\hat{X}\right)\hat{u}_j + \sum_{i=1}^{l} \nabla g_i\left(\hat{X}\right)\hat{v}_i = 0$

$$\hat{v}_i g_i(\hat{X}) = 0, \ \forall 1 \leq i \leq l$$

KKT条件的应用

直接求KT解的（理论上的）一般性方法

问题 $\min\left\{ f(X) \middle| \text{ s.t. } g_i(X) \le 0,\ 1 \le i \le l, h_j(X) = 0, 1 \le j \le m \right\}$

求KT解等价于求解下述等式和不等式方程

$$\nabla f(X) + \sum_{j=1}^{m} \nabla h_j(X) u_j + \sum_{i=1}^{l} \nabla g_i(X) v_i = 0$$

$$h_j(X) = 0, \quad j = 1, 2, \cdots, m$$

$$v_i \ge 0, \quad g_i(X) \le 0, \quad i = 1, 2, \cdots, l$$

$$v_i g_i(X) = 0, \quad i = 1, 2, \cdots, l$$

（理论上的）求解方法：

1）假定 $\hat{l}$ 个 $v_i > 0$（$g_i(x) = 0$），其余 $v_i = 0$，如

$$v_i > 0,\ i = 1, 2, \cdots, \hat{l};\ v_i = 0,\ i = \hat{l}+1, \hat{l}+2, \cdots, l$$

2）求解 $n+m+\hat{l}$ 个变量的 $n+m+\hat{l}$ 个等式方程

3）验证所求得的解是否满足其余 $l-\hat{l}$ 个不等式

分别考虑起作用的不等式约束的所有组合情况，求得所有的KT解，或者确定不存在KT解

例、直接用KT条件求解下述问题

解同一个问题，
直接拉格朗日和
KKT基本一致

$$\min\ (x_1-1)^2+(x_2-2)^2$$
$$\text{s.t.}\quad -x_1+x_2-1=0$$
$$x_1+x_2-2\le 0,\ -x_1\le 0,\ -x_2\le 0$$

拉格朗日函数 $L(X,u,V)=(x_1-1)^2+(x_2-2)^2+u(-x_1+x_2-1)$
$$+v_1(x_1+x_2-2)+v_2(-x_1)+v_3(-x_2)$$

梯度条件 $$\frac{\partial L(X,u,V)}{\partial X}=\begin{pmatrix}2(x_1-1)-u+v_1-v_2\\2(x_2-2)+u+v_1-v_3\end{pmatrix}=0$$

互补松弛条件 $v_1(x_1+x_2-2)=0,\ v_2x_1=0,\ v_3x_2=0$

求解：<u>分别考虑各种 $v_i>0$ 和 $v_i=0$ 的组合情况</u>

例如：$v_1>0,\ v_2=0,\ v_3=0$ ，相当于 $x_1+x_2-2=0,\ -x_1\le 0,\ -x_2\le 0$

由四个等式方程可求得 x_1,x_2,u,v_1，再验证有关不等式

凸优化问题KT解的性质

$\min\limits_{X\in\Omega} f(X)$ 是凸规划问题的条件

f 是凸函数，Ω 是凸集

等式和不等式约束描述的问题

$$\min \left\{ f(X) \middle| \text{ s.t. } h_j(X)=0,\ \ j=1,\cdots,m,\ \ g_i(X)\le 0, i=1,\cdots,l \right\}$$

是凸规划问题的条件

f 凸函数，g_i 都是凸函数，h_j 都是线性（仿射）函数

$$h_j(X)=P_j^T X-b_j,\ \forall j$$

此时可行集是凸集

设 $\hat{X}$ 是满足KT条件的可行解，X是任意可行解

$$g_i(X) \le 0,\ g_i(\hat{X}) = 0,\ \forall 1 \le i \le \hat{l}$$

$$P_j^T X = b_j,\ P_j^T \hat{X} = b_j,\ \forall 1 \le j \le m$$

利用可导凸（凹）函数的性质可以得到

$$(X - \hat{X})^T \nabla f(\hat{X}) \le f(X) - f(\hat{X})$$

$$(X - \hat{X})^T \nabla g_i(\hat{X}) \le g_i(X) - g_i(\hat{X}) \le 0,\ \forall 1 \le i \le \hat{l}$$

又因为 $h_j(X) = P_j^T X - b_j,\ \forall j$，可知

$$(X - \hat{X})^T \nabla h_j(\hat{X}) = (X - \hat{X})^T P_j = 0,\ \forall 1 \le j \le m$$

因为 $\hat{X}$ 是满足KT条件的可行解，所以存在

可正可负的 $u_j, j=1,2,\cdots,m$ 和 $v_i \geq 0, i=1,2,\cdots,\hat{l}$

满足 $$\nabla f(\hat{X}) = -\sum_{j=1}^{m} u_j \nabla h_j(\hat{X}) - \sum_{i=1}^{\hat{l}} v_i \nabla g_i(\hat{X})$$

利用 $$(X-\hat{X})^T \nabla h_j(\hat{X}) = 0, \forall 1 \leq j \leq m$$

可得 $$(X-\hat{X})^T \nabla f(\hat{X}) = -\sum_{i=1}^{\hat{l}} v_i (X-\hat{X})^T \nabla g_i(\hat{X})$$

再利用 $$(X-\hat{X})^T \nabla f(\hat{X}) \leq f(X) - f(\hat{X})$$

又可得 $$f(X) - f(\hat{X}) \geq -\sum_{i=1}^{\hat{l}} v_i (X-\hat{X})^T \nabla g_i(\hat{X})$$

最后，由于 $v_i \geq 0, i = 1, 2, \cdots, \hat{l}$

$$(X - \hat{X})^T \nabla g_i(\hat{X}) \leq 0, \ \forall 1 \leq i \leq \hat{l}$$

$$f(X) - f(\hat{X}) \geq -\sum_{i=1}^{\hat{l}} v_i (X - \hat{X})^T \nabla g_i(\hat{X})$$

可以得到 $f(X) \geq f(\hat{X})$

由于 X 是任意可行解，所以 $\hat{X}$ 是全局最优解

结论：<u>对于凸规划问题，**Kuhn-Tucker**条件是全局最优解的充分条件</u>

- KKT主要是针对带约束的可微分的优化问题，凸优化研究的对象是目标函数为凸函数，约束为凸集的优化问题。两者研究的对象，有交集，也各有不同。

- 带约束的可微分凸优化问题同时具备　凸优化和可微分性，此时KKT为全局最优解的充要条件。

- 可微分的但不是凸优化的问题也很多，多采用基于梯度的算法来求解，例如对深度神经网络的训练问题。采用梯度法仅仅能保证收敛到局部最优的必要条件而已。

KKT条件的应用条件

- KKT 条件在使用时是有要求的：在某些规范性条件（Constraint Qualification，简写作CQ）成立的时候，优化问题的最优解将满足 KKT 条件。即：CQ => KKT condition holds at optimal solution

- 最常用的 CQ 是 Slater’s condition, Linearity constraint qualification（简写为 LCQ）和Linear independence constraint quality （简写为 LICQ）

- Slater‘s condition也叫 strict feasibility，意思是说对于一个凸优化问题，如果存在一个点x，使得所有约束都成立，并且其中非线性的不等式约束都严格成立，那么最优解处KKT条件成立。（研究生凸优化算法会详细讲解Slater条件的证明）

- LCQ是说，对于一个优化问题，如果等式约束和不等式约束都是线性的（形如Ax+b的仿射函数），那么最优解处KKT条件成立。

- LICQ是说，对于一个优化问题，如果目标函数，等式约束和不等式约束都是可微且一阶导数连续，起作用的 g(x) 函数（即 g(x) 相当于等式约束的情况）和 h(x) 函数在极值点x*处的梯度线性无关，那么最优解处KKT条件成立。

- 针对LICQ的反例如下，min x, s.t. x^2 <=0。▽f=1，▽g=0， KKT的第一个等式不可能成立。

- Fritz John 条件

$$\mu_0 \nabla f\left(x^*\right)+\sum_{i=1}^{m} \mu_i^* \nabla g_i\left(x^*\right)+\sum_{i=1}^{n} \lambda_i^* \nabla h_i\left(x^*\right)=0$$

$$\begin{aligned}
& h_i\left(x^*\right)=0, && i=1,2,\cdots,n \\
& g_i\left(x^*\right) \leq 0, && i=1,2,\cdots,m \\
& \mu_i^* \geq 0, && i=1,2,\cdots,m \\
& \mu_i^* g_i\left(x^*\right)=0, && i=1,2,\cdots,m
\end{aligned}$$

仔细对比KKT条件和Fritz John条件，发现两者的区别是Fritz John conditions多了一个对目标函数的乘子，恰恰可以把上面那种退化的情况涵盖进去。

- KKT条件的用途

- 可以比较容易的用KKT条件验证任意的一个解是不是最优解，帮助我们把可行域里边很多不是最优的解排除掉，方便我们进一步寻找真正的最优解。

- 帮助我们理解约束问题的本质，设计更好的算法（研究生凸优化算法会详细讲解如何基于KKT条件设计内点法）

拉格朗日对偶问题

拉格朗日对偶理论

原问题

$$\min\left\{ f(X) \middle|\ \text{s.t. } h_j(X)=0, 1\le j\le m, g_i(X)\le 0, 1\le i\le l \right\}$$

拉格朗日对偶问题

$$\max\left\{ \rho(U,V) \middle|\ \text{s.t. } V\ge 0 \right\}$$

其中对偶目标函数为

$$\rho(U,V)=\min_{X\in R^n} L(X,U,V)=f(X)+\sum_{j=1}^{m}h_j(X)u_j+\sum_{i=1}^{l}g_i(X)v_i$$

注意：拉格朗日对偶问题永远是凸优化问题！

例、标准线性规划问题的拉格朗日对偶问题

原问题 $-\min\left\{-C^T X \,\middle|\, \text{s.t. } -X \le 0,\ AX = \vec{b}\right\}$

（优化问题的）拉格朗日函数

$$L(X,U,V) = -C^T X + U^T\left(AX - \vec{b}\right) + V^T(-X)$$

$$= \left(-C + A^T U - V\right)^T X - \vec{b}^T U$$

对偶目标函数

$$\rho(U,V) = \min_{X\in R^n} L(X,U,V) = \begin{cases} -\vec{b}^T U & \text{if } -C + A^T U - V = 0 \\ -\infty & \text{if } -C + A^T U - V \ne 0 \end{cases}$$

对偶问题 $-\max\left\{-\vec{b}^T U \,\middle|\, \text{s.t. } -C + A^T U - V = 0,\ V \ge 0\right\}$

价表示 $\min\left\{\vec{b}^T U \,\middle|\, \text{s.t. } A^T U \ge C\right\}$

弱对偶定理 如果$\left(\hat{X},\hat{U},\hat{V}\right)$分别是原对偶可行解，即

$$h_j(\hat{X})=0,1\le j\le m,\ \ g_i(\hat{X})\le 0,1\le i\le l,\quad \hat{V}\ge 0$$

则成立 $\rho\left(\hat{Y},\hat{Z}\right)\le f\left(\hat{X}\right)$

证明 $$\rho\left(\hat{U},\hat{V}\right)=\min_{X\in R^n} f\left(X\right)+\sum_{j=1}^{m}h_j\left(X\right)\hat{u}_j+\sum_{i=1}^{l}g_i\left(X\right)\hat{v}_i$$

$$\le f\left(\hat{X}\right)+\sum_{j=1}^{m}h_j\left(\hat{X}\right)\hat{u}_j+\sum_{i=1}^{l}g_i\left(\hat{X}\right)\hat{v}_i\le f\left(\hat{X}\right)$$

推论 上述$\left(\hat{X},\hat{U},\hat{V}\right)$若满足$\rho\left(\hat{U},\hat{V}\right)=f\left(\hat{X}\right)$，则分别是原对偶问题的最优解

该推论直观的给出了定义拉格朗日对偶问题的理由

强对偶和 KKT 条件的关系

- 对于任何一个优化问题，假如强对偶性成立，那么问题的最优解将满足 KKT 条件；For any optimization problem with strong duality, KKT condition is necessary condition

- 对于凸优化问题，CQ 可以导出强对偶性，即：convex + CQ => strong duality holds特别地，如果选择 Slater's condition，上述 CQ 就相当于 strict feasibility 成立。

- 对于严格可行的凸优化问题，满足 KKT 条件的点也是原问题和对偶问题的最优解。For convex optimization with Slater’s condition, KKT condition is also sufficient condition for primal-dual optimality.

线性不等式约束下的KKT条件

线性不等式约束可行方向的充要条件

对于线性不等式约束 $P_i^T X \geq b_i, 1 \leq i \leq l$ ， $D \in R^n$

是可行解 $\hat{X}$ 处可行方向的充要条件是

$$P_i^T D \geq 0, \forall 1 \leq i \leq \hat{l}$$

证明 $P_i^T(\hat{X} + tD) \geq b_i \quad \Leftrightarrow \quad tP_i^T D \geq b_i - P_i^T \hat{X}$

因为 $P_i^T \hat{X} = b_i, \forall 1 \leq i \leq \hat{l},\ t > 0$ 起作用约束

所以 $tP_i^T D \geq b_i - P_i^T \hat{X}, \forall 1 \leq i \leq \hat{l}$

$$\Leftrightarrow \quad P_i^T D \geq 0, \forall 1 \leq i \leq \hat{l}$$

线性不等式约束可行方向的充要条件

对于线性不等式约束 $P_i^T X \geq b_i, 1 \leq i \leq l$，$D \in R^n$ 是可行解 $\hat{X}$ 处可行方向的充要条件是

$$P_i^T D \geq 0, \forall 1 \leq i \leq \hat{l}$$

下降方向的充分条件

对任意的 $\hat{X} \in R^n, D \in R^n$，如果 $\nabla^T f(\hat{X}) D < 0$，$D$ 就是 $f(X)$ 在 $\hat{X}$ 处的下降方向

线性不等式约束的非线性规划：

可行方向

$$P_i^T D \geq 0, \quad \forall 1 \leq i \leq \hat{l}$$

下降方向

$$\nabla^T f(\hat{X}) D < 0$$

可行下降方向

$$\begin{cases} P_i^T D \geq 0, \forall 1 \leq i \leq \hat{l} \\ \nabla^T f(\hat{X}) D < 0 \end{cases}$$

如果 $\hat{X}$ 是局部最优点，则 $\hat{X}$ 处没有可行下降方向

$$\begin{cases} P_i^T D \geq 0, \quad \forall 1 \leq i \leq \hat{l} \\ \nabla^T f(\hat{X}) D < 0 \end{cases} \quad \text{无解}$$

显然 $\nabla f(\hat{X}) = 0$ 时，上式无解；当 $\nabla f(\hat{X}) \neq 0$ 时：

若 $D \in R^1$ 显然 ∇f 与 p_i 同号时上式无解，而 $D \in R^n$

上式无解的含义是什么?

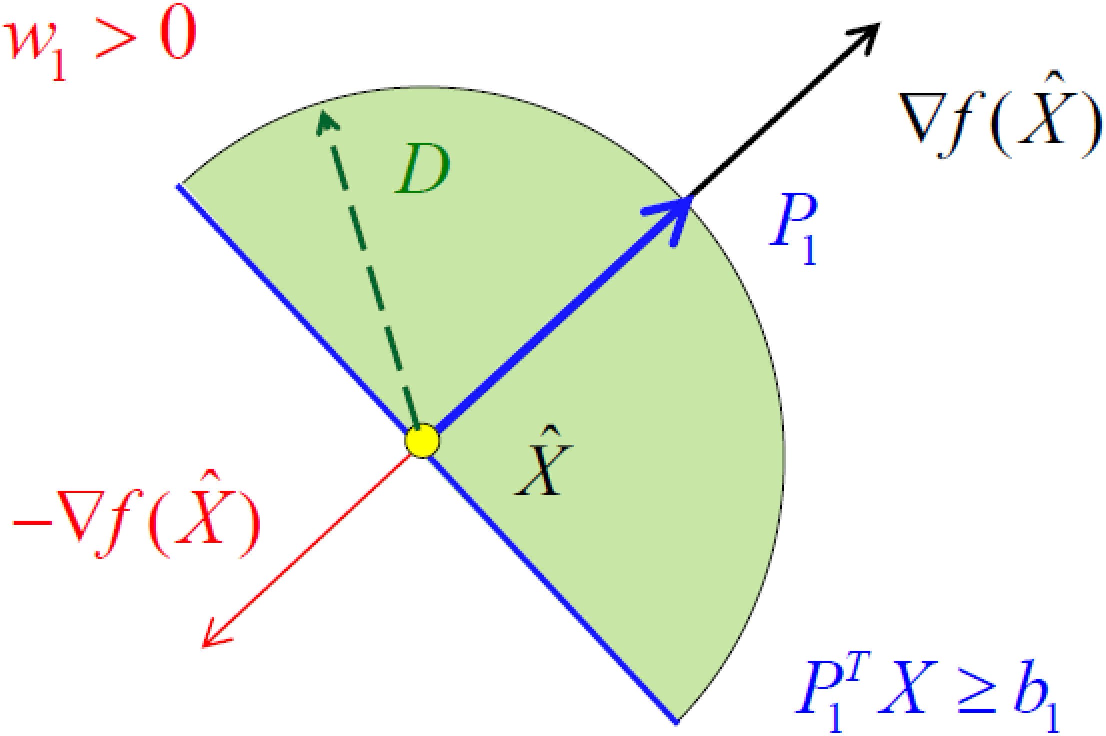
假设只有一个起作用约束

$$\begin{cases} P_1^T D \geq 0 \\ \nabla^T f(\hat{X}) D < 0 \end{cases}$$

无解

显然只要 $\nabla f(\hat{X})$ 与 P 同方向即能确保上式无解

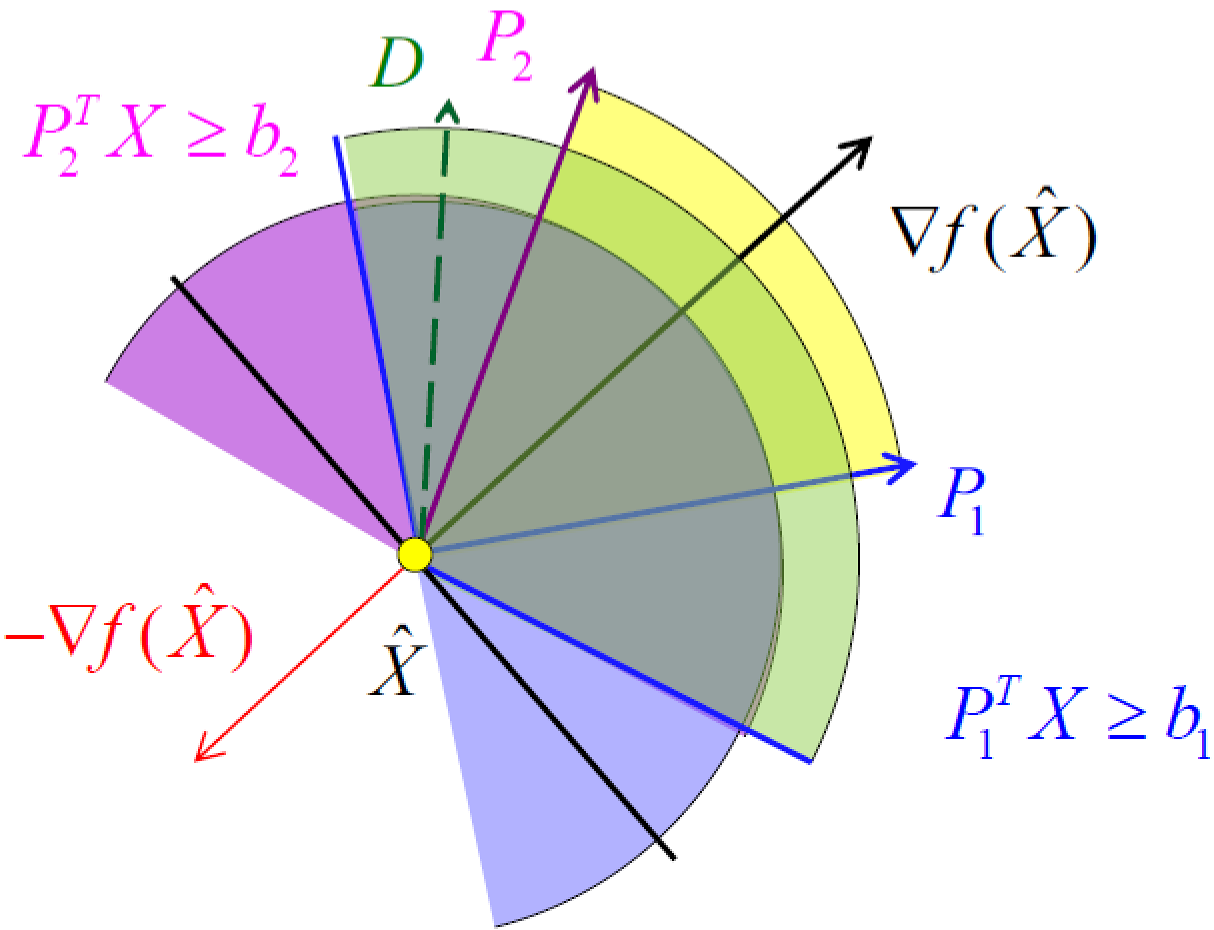
$$\nabla f(\hat{X}) = w_1 P_1,\ w_1 > 0$$



如果有两个起作用约束

$$\begin{cases} P_1^T D \ge 0, P_2^T D \ge 0 \\ \nabla^T f(\hat{X}) D < 0 \end{cases}$$

无解



$\nabla f(\hat{X}) = w_1 P_1 + w_2 P_2$, w_1, w_2 是不全为零的非负数

如果 $\hat{X}$ 是局部最优点，则 $\hat{X}$ 处没有可行下降方向

$$\begin{cases} P_i^T D \geq 0, \quad \forall 1 \leq i \leq \hat{l} \\ \nabla^T f(\hat{X}) D < 0 \end{cases}$$

无解

Farkas引理

上式无解等价于

$$\nabla f(\hat{X}) - (w_1 P_1 + w_2 P_2 + \cdots + w_{\hat{l}} P_{\hat{l}}) = 0$$

$$\Rightarrow \nabla f(\hat{X}) - \left(P_1, P_2, \cdots, P_{\hat{l}}\right)\left(w_1, w_2, \cdots, w_{\hat{l}}\right)^T = 0$$

其中 w_i 是不全为零的非负数

含有线性等式和非线性不等式约束下的KKT条件

线性等式，一般性不等式约束的优化问题

$$\min\left\{ f(X) \middle| \text{ s.t. } AX-\vec{b}=0,\ g_i(X)\ge 0,\ 1\le i\le l \right\}$$

其中 $A\in R^{m\times n}, n>m$

假定： A 是行满秩矩阵 $\Rightarrow$ $A=(B,N)$， B^{-1} 存在

令 $X=\begin{pmatrix} Z \\ Y \end{pmatrix}, Z\in R^{m}, Y\in R^{n-m}$

上述问题可写成

$$\min\left\{ f(Z,Y) \middle| \text{ s.t. } BZ+NY-\vec{b}=0,\ g_i(Z,Y)\ge 0,\ 1\le i\le l \right\}$$

约束 $BZ+NY-\vec{b}=0$ ，B^{-1} 存在， Z 可以用 Y 表示

$$Z=B^{-1}\left(\vec{b}-NY\right)=F(Y)$$

$$f(Z,Y)=f\left(F(Y),Y\right)=\bar{f}(Y)$$

$$g_i(Z,Y)=g_i\left(F(Y),Y\right)=\bar{g}_i(Y)$$

所以，求解

$$\min\left\{f(Z,Y)\middle|\ \text{s.t. } BZ+NY-\vec{b}=0,\ g_i(Z,Y)\geq 0, 1\leq i\leq l\right\}$$

可以等价转换为求解仅含变量 Y 的不等式约束问题

$$\min\left\{\bar{f}(Y)\middle|\ \text{s.t. } \bar{g}_i(Y)\geq 0, 1\leq i\leq l\right\}$$

不等式约束优问题局部最优解的必要条件

问题 $\min\left\{ f(X) \mid \text{s.t. } g_i(X) \ge 0, 1 \le i \le l \right\}$

前提 $g_i(\hat{X}) = 0, 1 \le i \le \hat{l},\ g_i(\hat{X}) > 0, \hat{l}+1 \le i \le l$

$\nabla g_1(\hat{X}), \nabla g_2(\hat{X}), \cdots, \nabla g_{\hat{l}}(\hat{X})$ 线性无关

结论 如果 $\hat{X}$ 是上述问题的局部最优解，则必满足

$$\nabla f(\hat{X}) = \sum_{i=1}^{\hat{l}} \nabla g_i(\hat{X}) w_i$$

LICQ+KKT

其中 w_i 是不全为零的非负数

设 $\hat{Y} \in R^{n-m}$ 是下述问题的一个可行解

$$\min\left\{\bar{f}(Y) \middle| \text{ s.t. } \bar{g}_i(Y) \geq 0, 1 \leq i \leq l\right\}$$

满足前提：1） $\bar{g}_i\left(\hat{Y}\right) = 0, 1 \leq i \leq \hat{l}, \bar{g}_i\left(\hat{Y}\right) > 0, \hat{l}+1 \leq i \leq l$

2） $\nabla\bar{g}_1\left(\hat{Y}\right), \nabla\bar{g}_2\left(\hat{Y}\right), \cdots, \nabla\bar{g}_{\hat{l}}\left(\hat{Y}\right)$ 线性无关

由不等式约束的 K-T 条件， $\hat{Y}$ 是最优解的必要条件是：存在 $w_i \geq 0, 1 \leq i \leq \hat{l}$ 满足

$$\nabla\bar{f}\left(\hat{Y}\right) = \sum_{i=1}^{\hat{l}} \nabla\bar{g}_i\left(\hat{Y}\right) w_i$$

求 $\bar{f}(Y)=f(F(Y),Y)$，$\bar{g}_i(Y)=g_i(F(Y),Y)$ 的梯度

$$\nabla\bar{f}(Y)=\frac{\partial F^T(Y)}{\partial Y}\frac{\partial f(Z,Y)}{\partial Z}+\frac{\partial f(Z,Y)}{\partial Y}$$

$$\nabla\bar{g}_i(Y)=\frac{\partial F^T(Y)}{\partial Y}\frac{\partial g_i(Z,Y)}{\partial Z}+\frac{\partial g_i(Z,Y)}{\partial Y}$$

代入 K-T 条件的等式 $\nabla\bar{f}(\hat{Y})=\sum_{i=1}^{\hat{l}}\nabla\bar{g}_i(\hat{Y})\,w_i$ 可得

$$\frac{\partial F^T(\hat{Y})}{\partial Y}\frac{\partial f(\hat{Z},\hat{Y})}{\partial Z}+\frac{\partial f(\hat{Z},\hat{Y})}{\partial Y}$$

其中：

$\hat{Z}=F(\hat{Y})$

$$=\sum_{i=1}^{\hat{l}}\left(\frac{\partial F^T(\hat{Y})}{\partial Y}\frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Z}+\frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Y}\right)w_i$$

$$\frac{\partial F^T\left(\hat{Y}\right)}{\partial Y}\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}+\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}$$

$$=\sum_{i=1}^{\hat{l}}\left(\frac{\partial F^T\left(\hat{Y}\right)}{\partial Y}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}+\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}\right)w_i$$

$$\Rightarrow \quad \frac{\partial F^T\left(\hat{Y}\right)}{\partial Y}\left(\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i\right)$$

$$=-\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i$$

因为 $F(Y)=B^{-1}\left(\vec{b}-N\hat{Y}\right)$ 所以 $\partial F^T\left(\hat{Y}\right)/\partial Y=-N^T B^{-T}$

$$\frac{\partial F^T\left(\hat{Y}\right)}{\partial Y}\left(\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i\right)$$

$$=-\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i$$

$$\Rightarrow \quad -N^T B^{-T}\left(\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i\right)$$

$$=-\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i$$

令 $B^{-T}\left(\dfrac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\dfrac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i\right)=\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_m\end{pmatrix}=\vec{\lambda}$

结合 $-N^T B^{-T}\left(\dfrac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\dfrac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i\right)$

$$=-\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i$$

$$\Rightarrow \begin{aligned}&\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i=B^T\vec{\lambda}\\ &-N^T\vec{\lambda}=-\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i\end{aligned}$$

$$\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}-\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i=B^T\vec{\lambda}$$

$$-N^T\vec{\lambda}=-\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i$$

其中：

$$\hat{X}=\begin{pmatrix}\hat{Z}\\\hat{Y}\end{pmatrix}$$

$$A^T=\begin{pmatrix}B^T\\N^T\end{pmatrix}$$

$$\Rightarrow\begin{array}{l}\dfrac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z}=\displaystyle\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}w_i+B^T\vec{\lambda}\\\dfrac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}=\displaystyle\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}w_i+N^T\vec{\lambda}\end{array}$$

$$\Rightarrow\frac{\partial f\left(\hat{X}\right)}{\partial X}=\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{X}\right)}{\partial X}w_i+A^T\vec{\lambda}$$ （K-T条件）

线性等式、不等式约束的 K-T 条件

$$\nabla f(\hat{X}) = \sum_{i=1}^{\hat{l}} \nabla g_i(\hat{X}) w_i + A^T \vec{\lambda}$$

一般不等式约束的 K-T 条件

$$\nabla f(\hat{X}) = \sum_{i=1}^{\hat{l}} \nabla g_i(\hat{X}) w_i$$

下面考虑如何用原函数表示K-T定理需要满足的前提

条件 1 $\bar{g}_i\left(\hat{Y}\right)=0, 1\le i\le \hat{l}, \quad \bar{g}_i\left(\hat{Y}\right)>0, \hat{l}+1\le i\le l$

因为 $\hat{Z}=F(\hat{Y}), A=\left(B,N\right), \quad \hat{X}=\begin{pmatrix}\hat{Z}\\ \hat{Y}\end{pmatrix}$

$$BF(\hat{Y})+N\hat{Y}-\vec{b}=0, \quad \bar{g}_i\left(\hat{Y}\right)=g_i\left(F(\hat{Y}),\hat{Y}\right)$$

以上条件显然等价于

$$A\hat{X}-\vec{b}=0$$

$$g_i\left(\hat{X}\right)=0, 1\le i\le \hat{l}, \quad g_i\left(\hat{X}\right)>0, \hat{l}+1\le i\le l$$

条件 2 $\nabla \bar{g}_1(\hat{Y}), \nabla \bar{g}_2(\hat{Y}), \cdots, \nabla \bar{g}_{\hat{l}}(\hat{Y})$ 线性无关

因为 $$\nabla \bar{g}_i(\hat{Y}) = \frac{\partial F^T(\hat{Y})}{\partial Y} \frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Z} + \frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Y}$$

$$= -N^T B^{-T} \frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Z} + \frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Y}$$

$$\sum_{i=1}^{\hat{l}} \nabla \bar{g}_i(\hat{Y}) \alpha_i = 0$$

$$\Leftrightarrow \quad -N^T B^{-T} \sum_{i=1}^{\hat{l}} \frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Z} \alpha_i + \sum_{i=1}^{\hat{l}} \frac{\partial g_i(\hat{Z},\hat{Y})}{\partial Y} \alpha_i = 0$$

$$\text{令}\ -B^{-T}\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}\alpha_i=\vec{\beta}=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_m\end{pmatrix}\Leftrightarrow\ -\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}\alpha_i=B^T\vec{\beta}$$

可得如下结果

$$-N^TB^{-T}\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}\alpha_i+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}\alpha_i=0\quad\Leftrightarrow$$

$$B^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}\alpha_i=0$$

$$N^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}\alpha_i=0$$

$$\Leftrightarrow\quad A^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{X}\right)}{\partial X}\alpha_i=0$$

前面推导说明：

$$\nabla\overline{g}_1\left(\hat{Y}\right),\nabla\overline{g}_2\left(\hat{Y}\right),\cdots,\nabla\overline{g}_{\hat{l}}\left(\hat{Y}\right)$$ 线性无关

$$\Leftrightarrow\ \sum_{i=1}^{\hat{l}}\nabla\overline{g}_i\left(\hat{Y}\right)\alpha_i=0$$ 是否有非零解

$$\Leftrightarrow\ A^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{X}\right)}{\partial X}\alpha_i=0$$ 的 α_i 是否有非零解

再利用 A^T 的列向量线性无关

$\Leftrightarrow$ A^T 的列向量和 $\nabla\overline{g}_1\left(\hat{X}\right),\cdots,\nabla\overline{g}_{\hat{l}}\left(\hat{X}\right)$ 一起线性无关

小结：对于线性等式一般性不等式约束的优化问题

$$\min\left\{ f(X) \middle| \text{ s.t. } AX-\vec{b}=0,\ g_i(X)\geq 0,\ 1\leq i\leq l \right\}$$

如果 $\hat{X}$ 是该问题的局部最优解，且满足：

1） $A\hat{X}-\vec{b}=0,\ \ g_i\left(\hat{X}\right)=0, 1\leq i\leq \hat{l},\ \ g_i\left(\hat{X}\right)>0, \hat{l}+1\leq i\leq l$

2） A^T 的列向量和 $\nabla g_1\left(\hat{X}\right),\cdots,\nabla g_{\hat{l}}\left(\hat{X}\right)$ 一起线性无关

那么，一定存在 $w_i\geq 0, 1\leq i\leq \hat{l}$ 和 $\vec{\lambda}\in R^m$ 成立

$$\nabla f\left(\hat{X}\right)=\sum_{i=1}^{\hat{l}}\nabla g_i\left(\hat{X}\right)w_i+A^T\vec{\lambda}$$

标准线性约束的简约梯度法

标准线性约束优化问题（可表示任意线性约束）

$$\min\left\{ f(X) \middle| \text{ s.t. } AX=\vec{b},\ X\geq 0\right\}$$

已知可行解 $\hat{X}=\begin{pmatrix}\hat{Z}\\ \hat{Y}\end{pmatrix}$ 满足以下条件：

1） $A\hat{X}=B\hat{Z}+N\hat{Y}=\vec{b},\ B^{-1}$ 存在

2） $\hat{Z}$ 的每个分量都大于零（非退化情况）

于是 $\hat{Y}$ 是下述问题可行解（ $\overline{f}(Y)=f\left(B^{-1}\left(\vec{b}-NY\right),Y\right)$）

$$\min\left\{ \overline{f}(Y) \middle| \text{ s.t. } B^{-1}\left(\vec{b}-NY\right)\geq 0,\ Y\geq 0\right\}$$

并且，$B^{-1}\left(\vec{b}-N\hat{Y}\right)>0$（对应的约束是不起作用约束）

求简约梯度

$$\overline{f}(Y) = f\left(Z, Y\right),\ Z = F(Y) = B^{-1}\left(\vec{b} - NY\right)$$

$$\Rightarrow \quad \nabla\overline{f}\left(Y\right) = \frac{\partial F^{T}(Y)}{\partial Y}\frac{\partial f\left(Z, Y\right)}{\partial Z} + \frac{\partial f\left(Z, Y\right)}{\partial Y}$$

$$= -N^{T}B^{-T}\frac{\partial f\left(Z, Y\right)}{\partial Z} + \frac{\partial f\left(Z, Y\right)}{\partial Y}$$

$$\Rightarrow \quad \nabla\overline{f}\left(\hat{Y}\right) = -N^{T}B^{-T}\frac{\partial f\left(\hat{Z}, \hat{Y}\right)}{\partial Z} + \frac{\partial f\left(\hat{Z}, \hat{Y}\right)}{\partial Y}$$

对于线性不等式约束的优化问题

$$\min\left\{\bar{f}(Y)\,\middle|\,\text{s.t. } -B^{-1}\left(\vec{b}-NY\right)\le 0, -Y\le 0\right\}$$

已知：$-B^{-1}\left(\vec{b}-N\hat{Y}\right)<0,\ -\hat{Y}\le 0$

$$\begin{pmatrix} r_1 \\ \vdots \\ r_{n-m} \end{pmatrix} = \nabla\bar{f}\left(\hat{Y}\right) = -N^T B^{-T}\frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Z} + \frac{\partial f\left(\hat{Z},\hat{Y}\right)}{\partial Y}$$

令 $D=\left(d_1,\cdots,d_{n-m}\right)^T,\ d_i=\begin{cases} -r_i & \text{if } r_i\le 0 \\ -\hat{y}_i r_i & \text{if } r_i>0 \end{cases}$

定理4.5.3（教材134页）：如果 $D=0$，$\hat{Y}$ 是上述问题的**KT**解，否则 D 是 $\hat{Y}$ 处的可行下降方向

KT解的理由：

$$D=0$$

$$d_i=\begin{cases}-r_i & \text{if } r_i\le 0\\ -\hat{y}_i r_i & \text{if } r_i>0\end{cases} \quad\Rightarrow\quad r_i\ge 0,\ \forall i \quad\Rightarrow\quad \begin{matrix} r_i=0 & \text{if } \hat{y}_i>0\\ r_i\ge 0 & \text{if } \hat{y}_i=0\end{matrix}$$

令 $w_i=r_i,\ \forall \hat{y}_i=0$

$$\Rightarrow\quad \nabla\bar{f}\left(\hat{Y}\right)=\begin{pmatrix} r_1\\ \vdots\\ r_{n-m}\end{pmatrix}=\sum_{\hat{y}_i=0}\vec{e}_i r_i=-\sum_{\hat{y}_i=0}\frac{\partial\left(-y_i\right)}{\partial Y}w_i$$

$$w_i\ge 0,\ \forall \hat{y}_i=0$$

$\Rightarrow\quad \hat{Y}$ **是KT解**

可行下降方向的理由：

$$\hat{Y} \ge 0$$

$$d_i = \begin{cases} -r_i & \text{if } \ r_i \le 0 \\ -\hat{y}_i r_i & \text{if } \ r_i > 0 \end{cases} \Rightarrow \hat{y}_i + t d_i \ge 0, \ \forall 1 \le i \le n-m, t > 0$$

$$B^{-1}\left(\vec{b} - N\hat{Y}\right) > 0$$

$$\Downarrow$$

可行方向

$$\hat{Y} \ge 0$$

$$D^T \nabla \overline{f}\left(\hat{Y}\right) = -\sum_{r_i \le 0} \left(r_i\right)^2 - \sum_{r_i > 0} \hat{y}_i \left(r_i\right)^2 \Rightarrow D^T \nabla \overline{f}\left(\hat{Y}\right) \le 0$$

$$D \ne 0$$

$$\Downarrow$$

下降方向 $\Leftarrow$ $D^T \nabla \overline{f}\left(\hat{Y}\right) < 0$

标准线性约束优化问题的简约梯度法

1）确定初始可行解 $\hat{X}$

2）选择 $\hat{X}$ 的前 m 个最大的正分量为 Z 向量，确定

$$\min\left\{\bar{f}(Y)\,\middle|\,\text{s.t. } B^{-1}\left(\vec{b}-NY\right)\geq 0, Y\geq 0\right\}$$ 及可行解 $\hat{Y}$

3）计算简约梯度 $r_i, 1\leq i\leq n-m$ 和可行下降方向 D

如果 $D=0$，停止，$\hat{X}$ 已是KT解

4）在 $\hat{Y}$ 处沿方向 D 进行一维搜索确定 $\hat{t}>0$，然后用 $\hat{Y}+\hat{t}D$ 替换 $\hat{Y}$，用 $\hat{Z}=B^{-1}\left(\vec{b}-N\hat{Y}\right)$ 和 $\hat{Y}$ 更换 $\hat{X}$，再回到 2）继续迭代

转化为无约束问题的方法

一般性优化问题的罚函数法（外点法）

对于一般性优化问题 $\min \; f(X)$

$$\text{s.t.} \quad h_j(X)=0, \quad j=1,2,\cdots,m$$

$$g_i(X)\le 0, \; i=1,2,\cdots,l$$

构造加上惩罚项的目标函数

$$F_k(X)=f(X)+k\sum_{j=1}^{m}\varphi\left(h_j(X)\right)+k\sum_{i=1}^{l}\rho\left(g_i(X)\right)$$

其中 $\varphi(u)=u^2, \rho(u)=\left(\max\{0,u\}\right)^2=\begin{cases} u^2 & \forall u>0 \\ 0 & \forall u\le 0 \end{cases}$

外点法就是求解下述无约束问题逼近原问题的解

$$\min_{X\in R^n} \; F_k(X) \quad \Rightarrow \quad \hat{X}(k) \quad \Rightarrow \quad X^*=\lim_{k\to\infty}\hat{X}(k)$$

不等式约束优化问题的障碍函数法（内点法）

对于一般性不等式约束优化问题

$$\min \ f(X)$$
$$\text{s.t.} \ g_i(X) \le 0, i = 1, 2, \cdots, l$$

构造加上障碍函数项的目标函数，如

$$\hat{f}_k(X) = f(X) - \frac{1}{k}\sum_{i=1}^{l}\log(-g_i(X))$$

内点法就是从可行集的某个内点开始求解下述

无约束优化问题逼近原问题的解

$$\min_{X \in R^n} \hat{f}_k(X) \quad \Rightarrow \quad \bar{X}(k) \quad \Rightarrow \quad X^* = \lim_{k \to \infty} \bar{X}(k)$$

非线性优化方法小结

1、凸函数的判断

1）多元函数转化为一元函数

2）多元函数的一阶充要条件

3）多元函数的二阶充分条件和必要条件

2、直线搜索方法（原理和性质）

1）0.618法

2）斐波那契法

3）利用一阶二阶导数的方法

4）非精确搜索方法

3、无约束优化问题的下降方向（原理、性质和计算）

1）最速下降方向

2）共轭梯度方向（二次目标函数的有限终止性）

3）各种下降方向的优缺点

4、约束优化问题

1）等式不等式约束的KT定理（导出过程与条件）

2）求KT解的方法（解方程~~或设计算法~~）

3）Lagrange对偶问题

4）简约梯度法