运筹学

（约束优化问题）

王焕钢

清华大学自动化系

要点：约束优化问题最优性条件概述

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

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

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

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

约瑟夫·拉格朗日
Joseph-Louis Lagrange
1736~1813

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

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、 $\min \ (x_1-1)^2+(x_2-2)^2$

$$\begin{aligned} \text{s.t.}\ \ & x_1-x_2=0 \\ & x_1+x_2-2\le 0 \\ & x_1\ge 0,\ \ x_2\ge 0 \end{aligned}$$

最优解应该满足什么条件（方程组）？

例 2、 $\min\ (x_1-1)^2+(x_2-2)^2$

$$\begin{aligned} \text{s.t.}\ & x_1-x_2=0 \\ & x_1+x_2-2\le 0 \\ & x_1\ge 0,\ \ x_2\ge 0 \end{aligned}$$

最优解应该满足什么条件（方程组）？

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）

要点：不等式约束的分类

一般性不等式约束优化问题

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

线性不等式约束优化问题

$$\min\left\{ f(X) \,\middle|\, \text{s.t. } P_i^T X \geq b_i, 1 \leq i \leq l \right\}$$

将约束分类为不起作用约束和起作用约束

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

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

要点：线性不等式约束下的KT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？

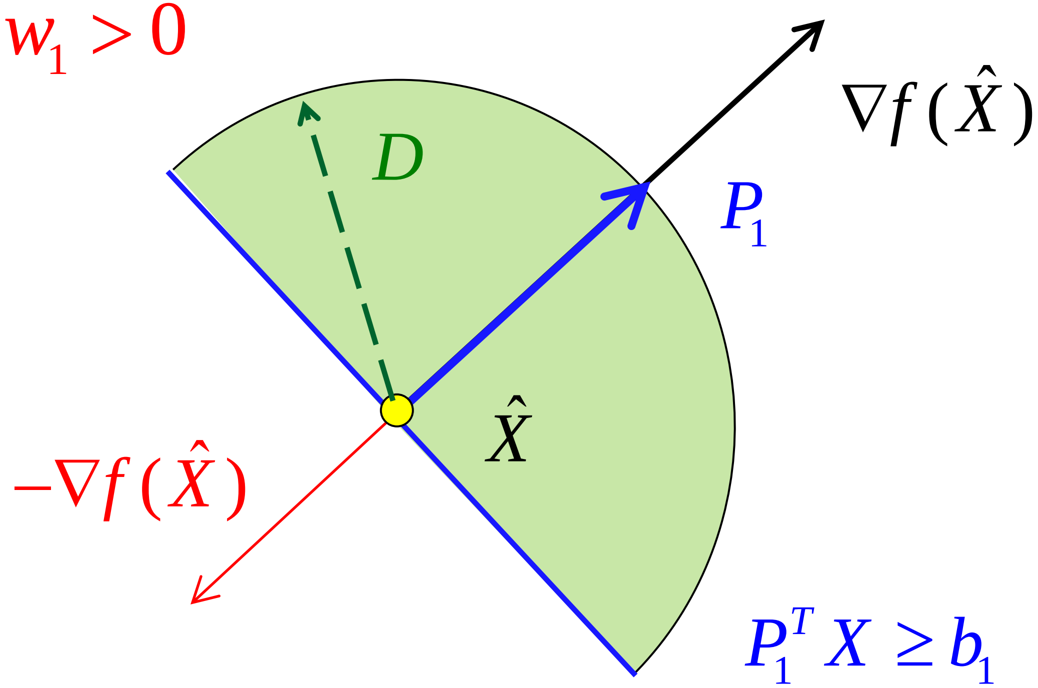
假设只有一个起作用约束

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

无解

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

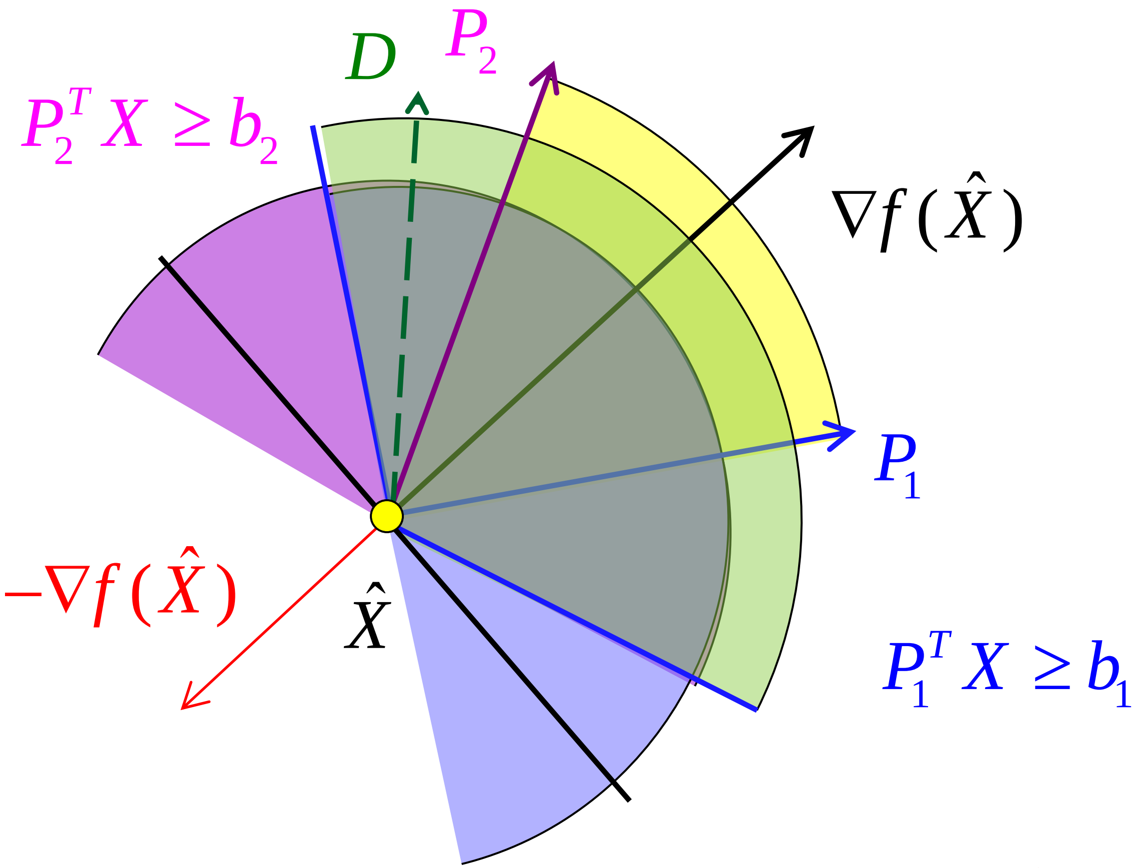
$$\nabla f(\hat{X}) = w_1 P_1,\ w_1 > 0$$



如果有两个起作用约束

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

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

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

无解

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

要点：线性等式约束处理方式

线性等式，一般性不等式约束的优化问题

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

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

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

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

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

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

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

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

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

不等式约束优问题局部最优解的必要条件

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

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

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

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

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

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

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

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

还原到 R^n
结果如何？

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

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

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

其中：

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

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

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

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

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

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

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

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

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

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

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

其中：

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

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

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

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

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

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

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

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

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

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

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

$$\begin{matrix} B^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Z}\alpha_i=0 \\ N^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{Z},\hat{Y}\right)}{\partial Y}\alpha_i=0 \end{matrix} \quad \Leftrightarrow \quad A^T\vec{\beta}+\sum_{i=1}^{\hat{l}}\frac{\partial g_i\left(\hat{X}\right)}{\partial X}\alpha_i=0$$

前面推导说明：

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

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

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

再利用 A^T 的列向量线性无关

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

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

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

要点：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（ $\bar{f}(Y)=f\left(B^{-1}\left(\vec{b}-NY\right),Y\right)$ ）

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

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

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

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

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

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

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：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：

$$\hat{Y} \geq 0$$

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

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

$$\Downarrow$$

可行方向

$$\hat{Y} \geq 0$$

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

$$D \neq 0$$

$$\Downarrow$$

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

要点：Karush-Kuhn-Tucker定理

Karush-Kuhn-Tucker定理

如果 $\hat{X}$ 是下述问题的局部最优解

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

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，则一定存在实数 $\lambda_j, 1\le j\le m$ 和 $w_i\ge 0, 1\le i\le l$ 满足

$$\nabla f(\hat{X})=\sum_{j=1}^{m}\nabla h_j(\hat{X})\lambda_j+\sum_{i=1}^{l}\nabla g_i(\hat{X})w_i$$

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

Karush-Kuhn-Tucker定理（文献常见的另外表述）

如果 $\hat{X}$ 是下述问题的局部最优解

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

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，则一定存在实数 $\hat{u}_j, 1\le j\le m$ 和 $\hat{v}_i \ge 0, 1\le i\le l$ 满足

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

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

Karush-Kuhn-Tucker定理（文献常见的另外表述）

如果 $\hat{X}$ 是下述问题的局部最优解

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

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，则一定存在实数 $\hat{u}_j, 1\le j\le m$ 和 $\hat{v}_i \ge 0, 1\le i\le l$ 满足

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

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

要点：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 $\min \ f(X)$

$$\begin{aligned} \text{s.t.} \quad & h_j(X)=0, \ \ j=1,2,\cdots,m \\ & g_i(X)\le 0, i=1,2,\cdots,l \end{aligned}$$

构造加上惩罚项的目标函数

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

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

外点法就是求解下述无约束问题逼近原问题的解

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

要点：障碍函数法

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

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

构造加上障碍函数项的目标函数，如

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

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

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

要点：KKT定理的构造性证明

Karush-Kuhn-Tucker 定理

如果 $\hat{X}$ 是下述问题的最优解

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

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

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

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

构造性证明

第一步、选择充分小的正数 ε 构造邻域

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

要求该邻域满足如下两个条件

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 $B\left(\hat{X},\varepsilon\right)$上的优化问题（简称罚问题）

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

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、通过求解罚问题得到收敛于$\hat{X}$ 的点列

用 X_k 表示罚问题的一个最优解

由 $\hat{X} \in B\left(\hat{X}, \varepsilon\right)$ 可得 $F_k^{(\hat{X})}(X_K) \le F_k^{(\hat{X})}(X)$，即

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

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

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

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

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、利用罚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 $t\ge\hat{t}$ 都成立 $\nabla F_{k_t}^{(\hat{X})}(X_{k_t})=0$，即

$$
\begin{aligned}
\nabla f(X_{k_t})&+2k_t\sum_{j=1}^{m}h_j(X_{k_t})\nabla h_j(X_{k_t})\\
&+2k_t\sum_{i=1}^{l}\max\left\{0,g_i(X_{k_t})\right\}\nabla g_i(X_{k_t})\\
&+2\left(X_{k_t}-\hat{X}\right)=0
\end{aligned}
$$

第四步继续

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

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

所以 $\max\left\{0, g_i(X_{k_t})\right\} = 0, \forall \hat{l}+1 \leq i \leq l$ ，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 $v_i^{k_t}\ge 0,\forall 1\le i\le \hat{l},t\ge\hat{t}$ ，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 $\nabla f\left(\hat{X}\right)+\sum_{j=1}^{m}\nabla h_j\left(\hat{X}\right)\hat{u}_j+\sum_{i=1}^{l}\nabla g_i\left(\hat{X}\right)\hat{v}_i=0$

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

要点：求KT解的一般性方法

直接求 KT 解的（理论上的）一般性方法

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

求 KT 解等价于求解下述等式和不等式方程

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

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

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

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

（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

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

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

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

求解：分别考虑各种 $v_i>0$ 和 $v_i=0$ 的组合情况

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

要点：凸优化问题KT解的性质

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

$$f(X)$$ 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：对于凸规划问题，Kuhn-Tucker条件是全局最优解的充分条件

要点：拉格朗日对偶

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

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

优化问题 $\min \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} x_{ij}$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, 2, \cdots, m$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \quad j = 1, 2, \cdots, n$$

$$x_{ij} \geq 0, \quad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n$$

对偶问题 $\max \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j$

$$\text{s.t.} \quad u_i + v_j \leq c_{ij}$$

$$i = 1, 2, \cdots, m \quad j = 1, 2, \cdots, n$$

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 $\rho\left(\hat{U},\hat{V}\right)\le f\left(\hat{X}\right)$

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

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

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