最优化方法

哈尔滨工业大学
数学学院

杨　畅

第 6 章

有约束最优化方法

问题 (fgh)
$$\begin{cases} \min & f(x) \\ \text{s.t.} & g(x) \le 0 \\ & h(x)=0 \end{cases}$$

分量形式略

约束集 $S=\{x|g(x) \le 0, h(x)=0\}$

6.1 Kuhn-Tucker 条件

1.等式约束性问题的最优性条件：

考虑

$$(fh) \begin{cases} \min & f(x) \\ \text{s.t.} & h(x)=0 \end{cases}$$

回顾高等数学中所学的条件极值：

问题 求$z=f(x,y)$极值

在$\Phi(x,y)=0$的条件下，

即 $\begin{cases} \min & f(x,y) \\ \text{S.t.} & \Phi(x,y)=0 \end{cases}$

引入Lagrange乘子：λ

Lagrange函数 $L(x,y;\lambda)= f(x,y)+ \lambda \Phi(x,y)$

若(x^*,y^*)是条件极值，则存在λ^*，使

$$f_x(x^*,y^*)+\lambda^*\Phi_x(x^*,y^*)=0$$
$$f_y(x^*,y^*)+\lambda^*\Phi_y(x^*,y^*)=0$$
$$\Phi(x^*,y^*)=0$$

推广到多个等式约束，可得到对于(*fh*)的情况：

分量形式：
$$\begin{cases} \min & f(x) \\ \text{s.t.} & h_j(x)=0 \qquad j=1,2,\ldots,l \end{cases}$$

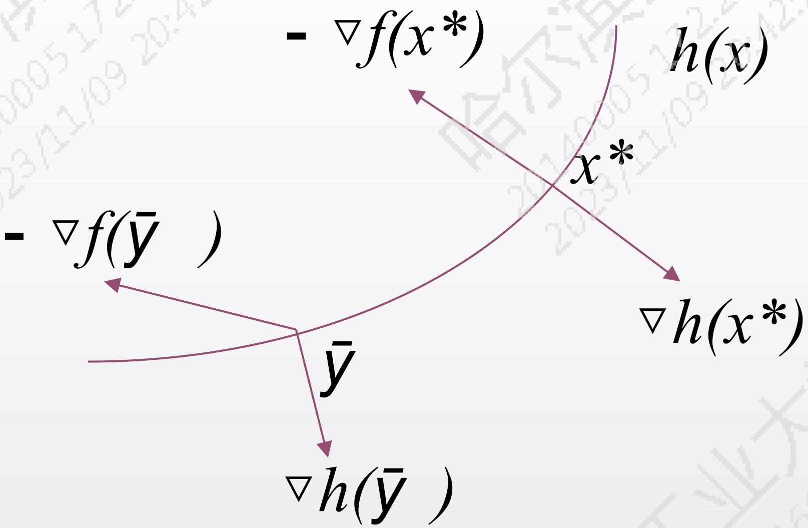
若x^*是(*fh*)的l.opt. ，则存在$\upsilon^*\in R^l$使

$$\nabla f(x^*)+\sum_{j=1}^{l}\upsilon_j^*\nabla h_j(x^*)=0$$

矩阵形式：

$$\nabla f(x^*)+\frac{\partial h(x^*)}{\partial x}\upsilon^*=0$$

几何意义是明显的：考虑一个约束的情况：



这里 x^* —l.opt.，$\nabla f(x^*)$与$\nabla h(x^*)$ 共线，而$\bar{y}$非l. opt.，$\nabla f(\bar{y})$与$\nabla h(\bar{y})$不共线。

最优性条件即

$$\nabla f(x^*) = -\sum_{j=1}^{h} \upsilon_j^* \nabla h_j(x^*)$$

2.不等式约束问题的Khun-Tucker条件

考虑问题 $\min \quad f(x)$

$\text{s.t.} \quad g_i(x) \le 0, \ i=1,2,\ldots,m$

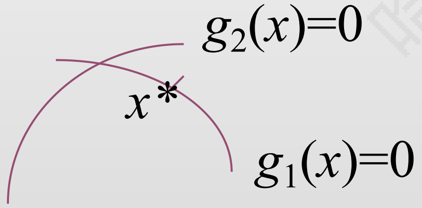
设 $x^* \in S=\{x \mid g_i(x) \le 0, \ i=1,2,\ldots,m\}$

令

$$I=\{i \mid g_i(x^*)=0, \ i=1,2,\ldots,m\}$$

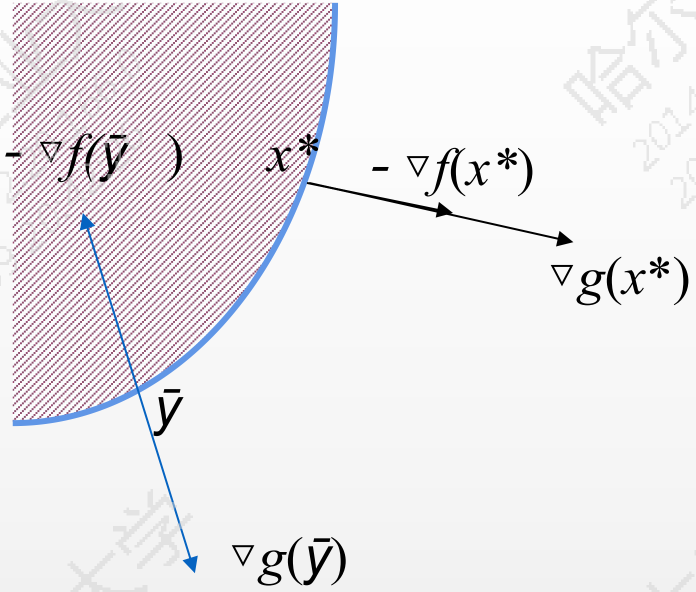
称I为 x^*点处的起作用集（紧约束集）。

如果x^*是l.opt. ,对每一个约束函数来说，只有当它是起作用约束时，才产生影响，如：



$g_1(x^*)=0$, g_1为起作用约束

特别地，有如下特征：



在x^*点使$f(x)$下降的方向（$-\nabla f(x^*)$方向）指向约束集合边界：$\nabla f(x^*)+u^*\nabla g(x^*)=0$，$u^*>0$

在$\bar{y}$点使$f(x)$下降的方向（$-\nabla f(\bar{y})$方向）指向约束集合内部，因此$\bar{y}$不是l.opt.。

定理（最优性必要条件 ）：（K-T条件）

问题(fg), 设$S=\{x|g_i(x)\le 0\}$, $x^*\in S$, I为x^*点处的起作用集，设f, $g_i(x)$,$i\in I$在x^*点可微， $g_i(x)$, $i\notin I$ 在x^*点连续。向量组$\{\nabla g_i(x^*), i\in I\}$线性无关。若$x^*$是l.opt. 那么存在$u_i\ge 0$, $i\in I$使

$$\nabla f(x^*)+\sum_{i\in I}u_i\nabla g_i(x^*)=0.$$

若f, $g_i(x)$,$\forall i$在x^*处可微，那么

$$\nabla f(x^*)+\sum_{i=1}^{m}u_i\nabla g_i(x^*)=0$$

$$u_i\ge 0, i=1,\cdots,m$$

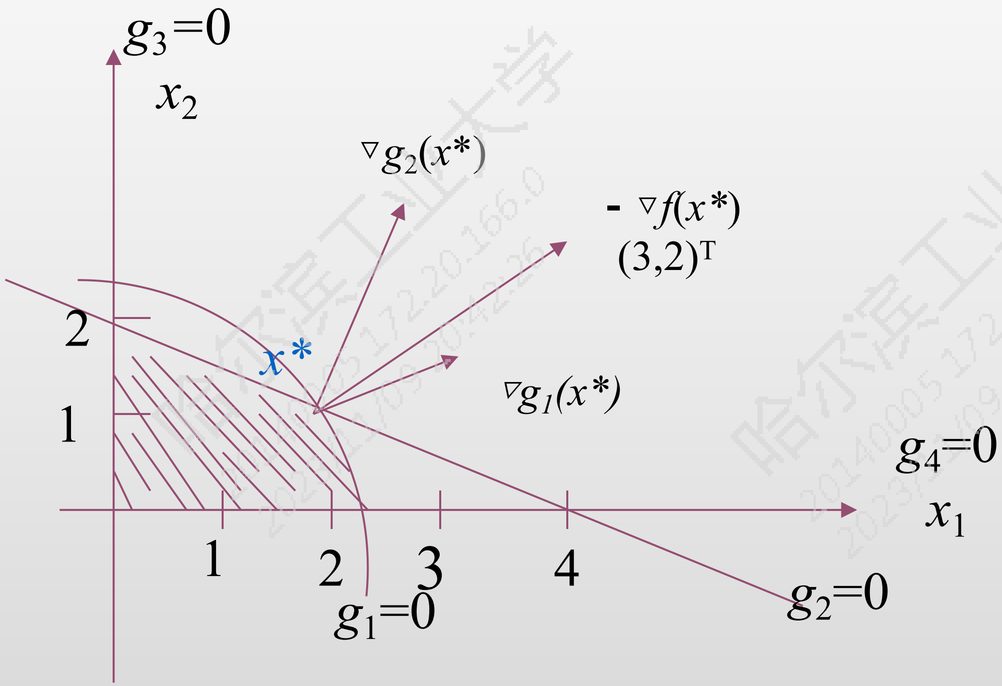
$$u^T g(x^*)=0, \text{(互补松弛条件)}$$

$$\leftrightarrow \quad u_i g_i(x^*)=0, i=1,\cdots,m$$

满足K-T条件的点x^*被称为K-T点。

例

$$
\begin{cases}
\text{Min} \quad f(x_1,x_2)=(x_1-3)^2+(x_2-2)^2 \\
\text{s.t.} \quad g_1(x_1,x_2)=x_1^2+x_2^2-5\le 0 \\
\qquad g_2(x_1,x_2)=x_1+2x_2-4\le 0 \\
\qquad g_3(x_1,x_2)=-x_1\le 0 \\
\qquad g_4(x_1,x_2)=-x_2\le 0
\end{cases}
$$



在 x^*点

$$\begin{cases} g_1(x_1,x_2)=0 \\ g_2(x_1,x_2)=0 \end{cases}$$ 交点（ $2,1)^T$ 起作用集 $I=\{1,2\}$

$$\nabla g_1(x^*)=(2x_1^*,2x_2^*)^T=(4,2)^T$$

$$\nabla g_2(x^*)=(1,2)^T$$

$$\nabla f(x^*)=(2(x_1^*-3),2(x_2^*-2))^T=(-2,-2)^T$$

计算可得 $u_1^*=\frac{1}{3}$ $u_2^*=\frac{2}{3}$ 使

$$\nabla f(x^*)+\tfrac{1}{3}\nabla g_1(x^*)+\tfrac{2}{3}\nabla g_2(x^*)=0$$

下面从使用K-T条件的角度来求解

$$\nabla f(x)=\begin{pmatrix}2(x_1-3)\\2(x_2-2)\end{pmatrix},\nabla g_1(x)=\begin{pmatrix}2x_1\\2x_2\end{pmatrix},\nabla g_2(2)=\begin{pmatrix}1\\2\end{pmatrix}$$

$$\nabla g_3(x)=\begin{pmatrix}-1\\0\end{pmatrix},\nabla g_4=\begin{pmatrix}0\\-1\end{pmatrix}$$

$$
\begin{cases}
\nabla f(x) + \sum_{i}^{m} u_i \nabla g_i(x) = 0 \\
u_i \geq 0, i = 1, 2, \cdots, m \\
u_i g_i(x) = 0
\end{cases}
$$

$$
\begin{cases}
2(x_1 - 3) + u_1 2x_1 + u_2 - u_3 = 0 \cdots\cdots (1) \\
2(x_2 - 2) + u_1 2x_2 + 2u_2 - u_4 = 0 \cdots\cdots (2) \\
u_1, u_2, u_3, u_4 \geq 0 \\
u_1(x_1^2 + x_2^2 - 5) = 0 \cdots\cdots (3) \\
u_2(x_1 + 2x_2 - 4) = 0 \cdots\cdots (4) \\
u_3 x_1 = 0 \cdots\cdots (5) \\
u_4 x_2 = 0 \cdots\cdots (6)
\end{cases}
$$

6个方程　6个未知量

可能的K-T点出现在下列情况：

①两约束曲线的交点：g_1与g_2，g_1与g_3，g_1与g_4，g_2与g_3，g_2与g_4，g_3与g_4。

②目标函数与一条曲线相交的情况：g_1，g_2，g_3，g_4

对上述每一个情况容易求得满足(1)~(6)的点$(x_1,x_2)^T$及乘子u_1,u_2,u_3,u_4，当满足$u_i \geq 0$时，即为K-T点。

下面举几个情况：

- g_1与g_2交点：$x=(2,1)^T \in S$，$I=\{1,2\}$ 则$u_3=u_4=0$

解

$$\begin{cases} 2(x_1-3)+2u_1x_1+u_2=0 \\ 2(x_2-2)+2u_1x_2+2u_2=0 \end{cases}$$

得$u_1=\frac{1}{3}, u_2=\frac{2}{3}>0$

故$x=(2,1)^T$是K－T点。

➢

g_1与g_3交点：$\begin{cases} x_1^2 + x_2^2 - 5 = 0 \\ x_1 = 0 \end{cases}$ 得 $x = (0, \pm\sqrt{5})^{\mathrm{T}}$

$(0, -\sqrt{5})^{\mathrm{T}} \notin S$，故不是 K－T点；

$(0, \sqrt{5})^{\mathrm{T}} \notin S$，不满足 $g_2 \le 0$，故不是 K－T点。

➢

g_3, g_4交点：$x = (0,0)^{\mathrm{T}} \in S$ $I = \{3,4\}$故$u_1 = u_2 = 0$

解$\begin{cases} 2(0-3) - u_3 = 0 \\ 2(0-2) - u_4 = 0 \end{cases}$得$u_3 = -6 < 0, u_4 = -4 < 0$

故非K－T点。

目标函数$f(x)$与$g_1(x)=0$相切的情况：

$I=\{1\}$，则$u_2=u_3=u_4=0$

解$\begin{cases} 2(x_1-3)+2x_1u_1=0 \\ 2(x_2-2)+2x_2u_1=0 \\ x_1^2+x_2^2-5=0 \end{cases}$ 得$(\pm\sqrt{\frac{45}{13}},\pm\sqrt{\frac{20}{13}})\notin S$

故均不是K－T点 $g_2(x_1,x_2)=\sqrt{\frac{45}{13}}+2\sqrt{\frac{20}{13}}-4=7\sqrt{\frac{5}{13}}-4=0.34>0$

3.一般约束问题的Kuhn-Tucker 条件

$$(fgh)\begin{cases} \min \quad f(x) \\ \text{s.t.} \quad g_i(x)\le 0,\ i=1,2,\cdots,m \\ \qquad h_j(x)=0,\ j=1,2,\cdots,l \end{cases}$$

定理：问题(fgh)，$x^* \in S=\{x \mid g(x)\le 0, h(x)=0\}$，$I$为起作用集。

设$g_i(x)(i\in I)$在x^*可微，$g_i(x)(i\notin I)$在x^*连续，h_j （$j=1,2,\cdots,l$）

在x^*的某邻域内连续可微。（CQ，约束规格）。

向量组$\{\cdots, \nabla g_i(x^*)(i\in I),\cdots,\nabla h_1(x^*),\cdots,\nabla h_l(x^*)\}$线性无关。

如果 x^*- l.opt. 那么 $\exists u_i^* \geq 0, i \in I,$

$v_j^* \in R, j=1,2,\cdots,l$

$$\nabla f(x^*)+\sum_{i\in I} u_i^* \nabla g_i(x^*)+\sum_{j=1}^{l} v_j^* \nabla h_j(x^*)=0$$

如果还有 $g_i(x)(i \notin I)$在x^*亦可微，那么

$$\begin{cases} \nabla f(x^*)+\sum_{i=1}^{m} u_i^* \nabla g_i(x^*)+\sum_{j=1}^{l} v_j^* \nabla h_j(x^*)=0 \\ u_i^* \geq 0 \qquad\qquad i=1,2,\cdots,m \\ u_i^* g_i(x^*)=0 \end{cases}$$

4.凸规划的一阶充分性条件

定理：当（fgh）为凸规划，$f(x), g_i(x), i=1,\ldots,m$为可微凸函数，$h(x)=Ax-b$, $A\in\mathbb{R}^{l\times n}, b\in\mathbb{R}^{l}$. 再设

$$x^*\in S=\{x\in\mathbb{R}^n \mid g(x)\le 0, h(x)=0\}$$

满足K-T条件，即

$$\nabla f(x^*)+\sum_{i\in I} u_i\nabla g_i(x^*)+\sum_{j=1}^{l} v_j\nabla h_j(x^*)=0$$

其中$u_i\ge 0$，$i\in I$, $v_j\in\mathbb{R}, j=1,\cdots,l$. 则$x^*$为（fgh）的g.opt.

考虑（LP）的标准形式：

$$\begin{cases} \max & z = c^T x \\ s.t. & Ax = b \\ & x \geq 0 \end{cases}$$

问题：（LP）的K-T点是否为最优解？

5.约束优化问题的二阶条件

例：（fh）$\begin{cases} min \ \ f(x) = 1/2(x_1-1)^2 + 1/2x_2^2 \\ s.t. \ \ h(x) = -x_1 + \beta x_2^2 = 0, \beta\text{为常数} \end{cases}$

取$x^* = (0,0)^T$，取$v = -1$，有

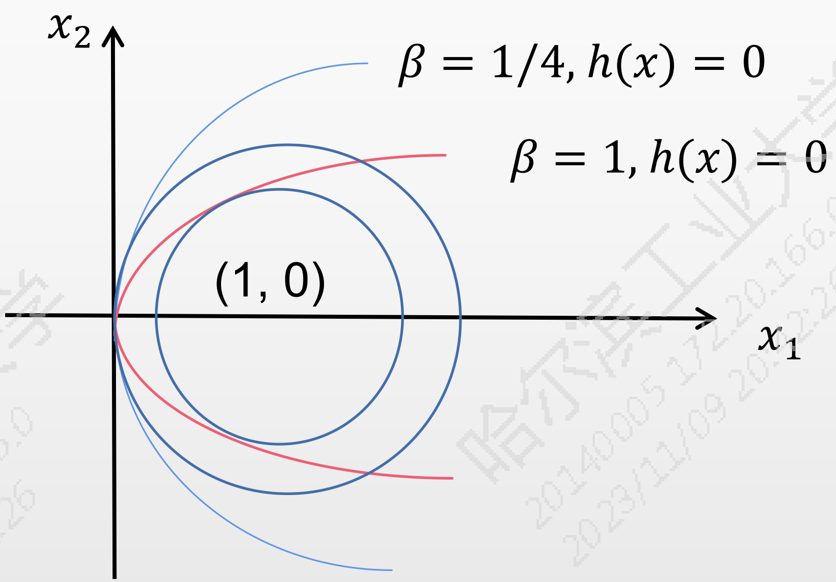
$\nabla f(x^*) - \nabla h(x^*) = 0.$

故x^*为（fh）的K-T点。

且有$\nabla^2 f(x^*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$正定

但，当$\beta = 1/4$时，x^*是opt.

当$\beta = 1$时，x^*不是opt.



注：

1.函数f的Hesse矩阵正定并不能保证是l.opt.

2.K-T点只是l.opt.的必要条件，并不是充分条件

考虑Lagrange函数$L(x,v)=f(x)+vh(x)$

取$v=-1$，则Lagrange函数的Hesse矩阵为

$$\nabla^2 L(x^*,-1)=\nabla^2 f(x^*)-\nabla^2 h(x^*)=\begin{pmatrix}1 & 0\\ 0 & 1-2\beta\end{pmatrix}$$

当$\beta<1/2$时，此时Lagrange函数的Hesse矩阵为正定的，x^*为opt.（较强的性质）

进一步取较弱的性质：取x^*处的可行方向$d=(0,d_2)^T, d_2\in\mathbb{R}, d\neq 0$

由$d^T\nabla^2 L(x^*,-1)d=(1-2\beta)d_2^2$，有：

a) 当$\beta<1/2$时，$d^T\nabla^2 L(x^*,-1)d>0$

b) 当β=1/2时，$d^T\nabla^2 L(x^*,-1)d$=0

c) 当$\beta>1/2$时，$d^T\nabla^2 L(x^*,-1)d<0$

因此，当$\beta\leq 1/2$时，x^*是opt.

当$\beta>1/2$时，x^*不是opt.

定理：（二阶必要条件）

问题（fgh）,设$x^* \in S, f(x), g_i(x)(i \in I), h_j(x), j=0,\cdots,l$,在$x^*$处二次可微，$g_i(x)(i \notin I)$在$x^*$处连续。$\{\nabla g_i(x^*)(i \in I), \nabla h_j(x^*), j=0,\cdots,l\}$线性无关。

如果x^*为l.opt.,那么存在$u_i \geq 0(i \in I)$,$v \in \mathbb{R}^l$使K-T条件式

$\nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) + \sum_{j=1}^{l} v_j \nabla h_j(x^*) = 0$成立。

由乘子u, v构造Lagrange函数

$$L(x,u,v) = f(x) + \sum_{i \in I} u_i \nabla g_i(x) + \sum_{j=1}^{l} v_j \nabla h_j(x)$$

令$S_h = \{d | \nabla h_j^T(x^*) d = 0, j = 1,2,\cdots,l\}$,

$S_{g0} = \{d | \nabla g_i^T(x^*) d = 0, i \in I\}$,

则$\forall d \in S_{g0} \cap S_h$有

$$d^T \nabla_x^2 L(x^*,u,v) d \geq 0$$

定理：（二阶充分条件）

问题（fgh）,设x^*是$K-T$点$u_i(i \in I), v_j(j = 0, \cdots, l)$为相应乘子,Lagrange函数为$L(x, u, v) = f(x) + \sum_{i \in I} u_i \nabla g_i(x) + \sum_{j=1}^{l} v_j \nabla h_j(x)$。

如果$d \in \{d | \nabla h_j^T(x^*) d = 0, j = 1, 2, \cdots, l\} \cap \{d | \nabla g_i^T(x^*) d = 0, i \in I\}$, 且$d \neq 0$，均有

$$d^T \nabla_x^2 L(x^*, u, v) d > 0$$

则x^*是问题（fgh）的严格局部最优解。

6.2 罚函数法

1.罚函数法：

$$(fgh)\left(\begin{array}{ll} \text{Min} \quad f(x) & f: \mathrm{R}^n \to \mathrm{R} \\ \text{s.t.} \quad g(x) \le 0 & g: \mathrm{R}^n \to \mathrm{R}^m \\ \quad h(x) = 0 & h: \mathrm{R}^n \to \mathrm{R}^l \end{array}\right)$$

原理：将违背约束作为对求最小值的一种惩罚，将约束并入目标函数，从而得到一个辅助的无约束最优化问题

构造罚函数：

$$\alpha(x) = \sum_{i=1}^{m} \phi(g_i(x)) + \sum_{j=1}^{l} \varphi(h_j(x))$$

目的：使满足约束的 x有 $\alpha(x) = 0$

不满足约束的 x有 $\alpha(x) > 0$

其中： $\phi(t)\begin{cases} >0 & ,\text{当}t>0\text{时} \\ =0 & \text{当}t\le 0\text{时} \end{cases}$

$\varphi(t)\begin{cases} >0 & ,\text{当}t\ne 0\text{时} \\ =0 & \text{当}t=0\text{时} \end{cases}$

取$\mu>0$, 可构造

$$\mu\alpha(x)\text{——惩罚项}\begin{cases} =0 & \text{可行} \\ \to\infty & \text{不可行} \end{cases}$$

$$f(x)+\mu\alpha(x)\text{——辅助函数}$$

$$\text{Min}\quad f(x)+\mu\alpha(x)\text{——辅助问题}$$

$\phi(t),\varphi(t)$的典型取法：

$\phi(t)=[\text{Max}\ \{0,t\}]^p \qquad \varphi(t)=|t|^p \qquad p$为正整数。

当$p=2$时，称2次罚函数.(常用：因2次是最低次的光滑函数)

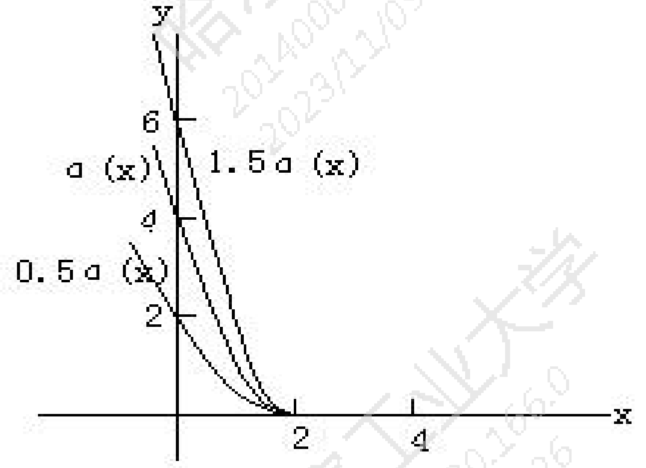
例. $\begin{cases} \text{Min} \quad x \\ \text{s.t.} \quad -x+2 \le 0 \end{cases}$

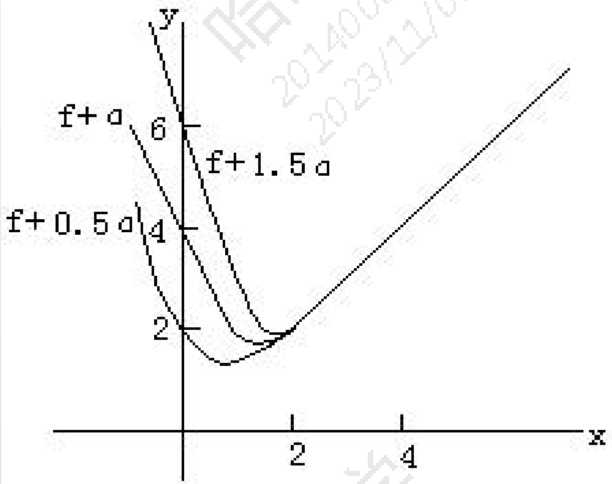
二次罚函数 $:\alpha(x) = [\text{Max}\{0, -x+2\}]^2 = \begin{cases} (x-2)^2, x<2 \\ 0, x \ge 2 \end{cases}$

- 当$\mu \to \infty$时, Min $\quad f(x)+\mu\alpha(x) \to f(x^*) = x^* = 2$
- 解析解：辅助函数

$$g(x,\mu) = f(x)+\mu\alpha(x) = \begin{cases} x+\mu(-x+2)^2 = \mu x^2 - (4\mu-1)x + 4\mu, & x<2 \\ x \quad , x \ge 2 \end{cases}$$

$$\left.\begin{array}{l} \text{当} x<2 \text{时}, g(x,\mu) \text{的驻点} \bar{x} = \dfrac{4\mu-1}{2\mu} \xrightarrow{\mu\to\infty} 2 \\ \text{当} x \ge 2 \text{时}, g(x,\mu) \text{的最小值点} \tilde{x} = 2 \end{array}\right\} \text{故} x^* = 2 \text{--opt.}$$





罚函数法的收敛性：

定义：$\theta(\mu)=\inf_{x}\{f(x)+\mu\alpha(x)\}$ 下确界

引理：设f,g,h连续，$\alpha(x)$为罚函数，连续。

再设$\forall\mu\geq 0$，$\exists x_{\mu}\in D$，使：$\theta(\mu)=f(x_{\mu})+\mu\alpha(x_{\mu})$

则 1° $\inf_{x}\{f(x)\mid x\in S\}\geq\sup\{\theta(\mu)\mid\mu\geq 0\}$

2° $f(x_{\mu}),\theta(\mu)$ 关于$\mu\geq 0$的单调非降函数；

3° $\alpha(x_{\mu})$ 关于$\mu\geq 0$的单调非增函数

定理:$(fgh), S=\{x \mid g(x) \le 0, h(x)=0\} \ne \Phi$,

在引理假设下，设存在单调增加的正数列$\{\mu_k\}$

即$0<\mu_1<\mu_2<\cdots<\mu_k<\cdots$.有$\{x_{\mu_k}\} \xrightarrow{k\to\infty} x^*$

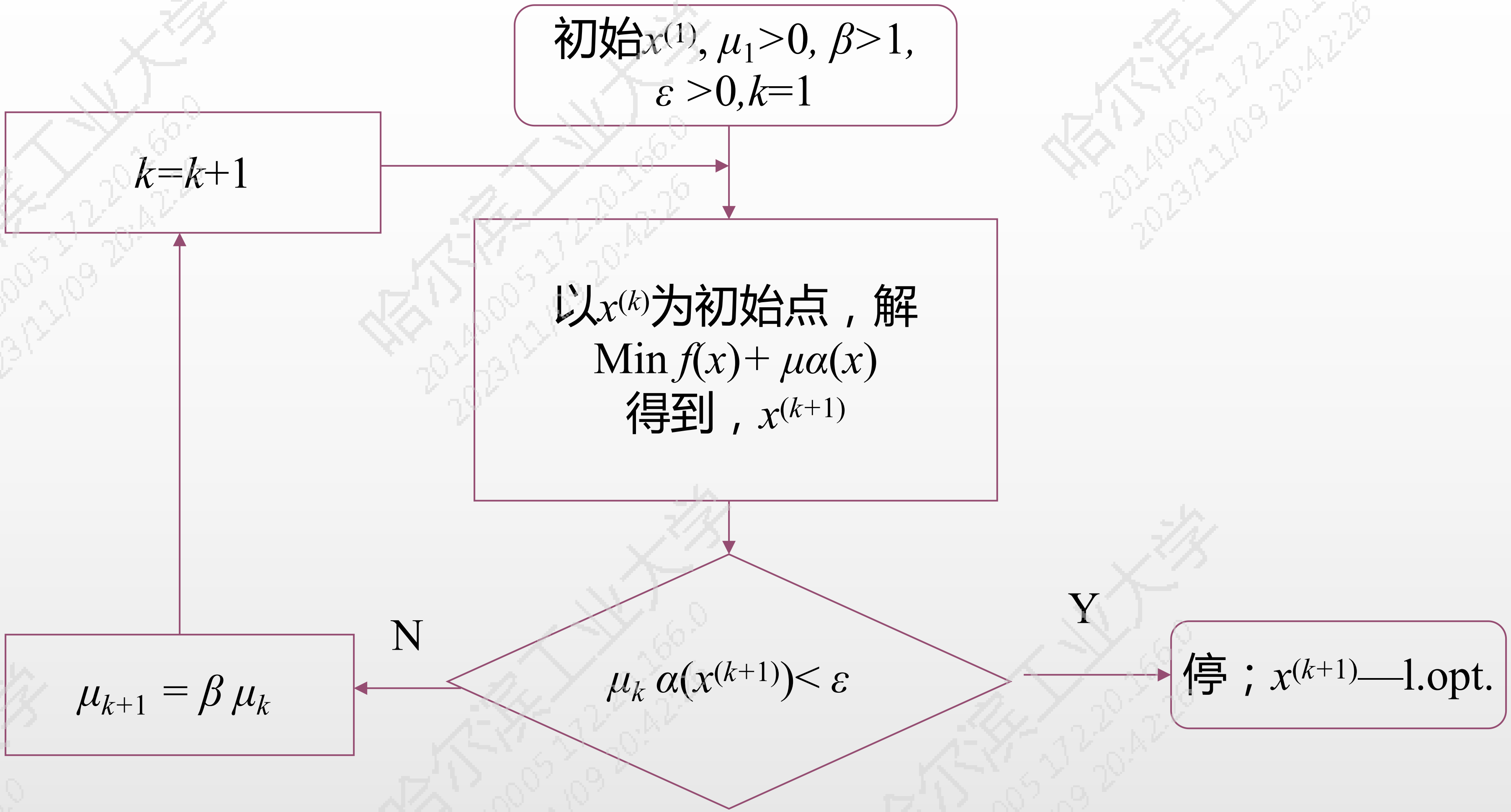
那么,$1^\circ \inf\{f(x) \mid x \in S\} = \sup\{\theta(\mu) \mid \mu \ge 0\} = \lim\limits_{\mu\to\infty} \theta(\mu)$

$2^\circ\ x^* \sim opt.$且$\mu\alpha(x_\mu) \xrightarrow{\mu\to\infty} 0$

推论：在定理条件下，若$\exists\mu \ge 0$使$\alpha(x_\mu)=0$,

则$x_\mu --opt.$

算法流程：



初始$x^{(1)}, \mu_1>0, \beta>1, \varepsilon>0, k=1$
以$x^{(k)}$为初始点，解
$\text{Min } f(x)+\mu\alpha(x)$
得到，$x^{(k+1)}$
$\mu_k\,\alpha(x^{(k+1)})<\varepsilon$
Y
停；$x^{(k+1)}$—l.opt.
N
$\mu_{k+1}=\beta\,\mu_k$
$k=k+1$

2.闸函数法（内点罚函数法）

$$(fg)\begin{cases} \text{Min} \quad f(x) & f: \mathrm{R}^n \to \mathrm{R} \\ \text{s.t.} \quad g(x) \le 0 & g: \mathrm{R}^n \to \mathrm{R}^m \end{cases}$$

记$S = \{x \mid g(x) \le 0\}, S_0 = \{x \mid g(x) < 0\} \ne \Phi$

基本思想：

从S_0中的一个点（内点）出发，在目标函数中加入惩罚项，使迭代保持在S_0内。内点法适用于只有不等式约束的问题。

构造闸函数(Barrier Function)：$B(x) = \sum_{i=1}^{m} \phi(g_i(x))$

使 $B(x)\begin{cases} > 0, & x \in S_0 \\ \to \infty, & x \to \partial S_0(\text{边界}) \end{cases}$ 为方便应有$B(x)$连续。

典型取法： $\phi(t) = -\dfrac{1}{t}$

或： $\phi(t) = |\ln(-t)|$

惩罚项： $\mu B(x)\begin{cases} \to 0 & x \in S_0 \\ +\infty & x \in \partial S_0 \end{cases}$

〈由于当 $x \in S_0$时 $B(x) > 0$且 $B(x) \xrightarrow{x \to \partial S_0} \infty$

〈故需要随着 $x \to \partial S_0, \mu \to 0^+$

辅助问题 $\min \quad f(x) + \mu B(x)$

例．
$$\begin{cases} \min & x \\ s.t. & -x+2 \le 0 \end{cases}$$

解 闸函数 $B(x)=\dfrac{1}{x-2},\ x>2$

$$\begin{cases} \text{Min} & g(x,\mu)=x+\dfrac{\mu}{x-2} \\ \text{s.t.} & x>2 \end{cases}$$

目标函数关于 x 是凸的，求驻点： $x_\mu = 2+\sqrt{\mu}$

$$x_\mu = 2+\sqrt{\mu} \xrightarrow{\mu\to 0^+} 2 = x^*$$

$g(x_\mu,\mu) \xrightarrow{\mu\to 0^+} 2$ ——最优值 （原问题 ）

定义 $\theta(\mu)=\inf\{f(x)+\mu B(x) \mid x\in S_0\}$

有类似于罚函数法的理论结果：

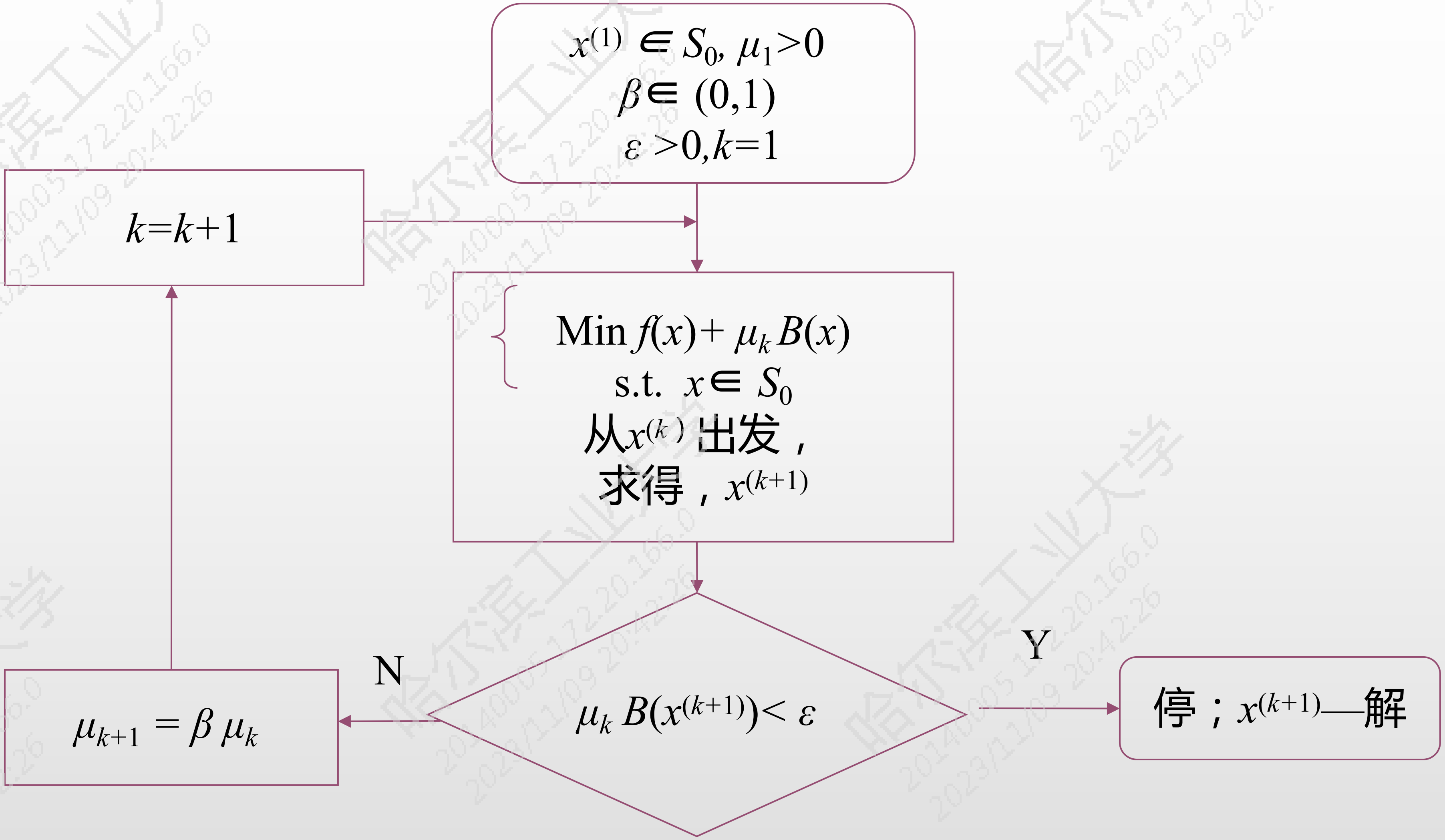
定理：(fg), f,g连续， $S_0\neq\Phi$, 最优解 $x^*\in S_0$

则 1° $\min\{f(x) \mid x\in S\}=\inf\{\theta(\mu) \mid \mu>0\}=\lim\limits_{\mu\to0^+}\theta(\mu)$

2° 若 $\forall\mu>0,\exists x_\mu\in S_0$, 使 $\theta(\mu)=f(x_\mu)+\mu B(x_\mu)$

那么, $\{x_\mu\}$的极限点是 (fg)的 opt. 且 $\lim\limits_{\mu\to0^+}\mu B(x_\mu)=0$

算法流程：


$x^{(1)} \in S_0,\ \mu_1>0$
$\beta \in (0,1)$
$\varepsilon>0, k=1$
$k=k+1$
Min $f(x)+\mu_k B(x)$
s.t. $x \in S_0$
从$x^{(k)}$出发，
求得，$x^{(k+1)}$
$\mu_k B(x^{(k+1)})<\varepsilon$
N
Y
$\mu_{k+1}=\beta\mu_k$
停；$x^{(k+1)}$—解

- 求初始内点：

$1^\circ\ \forall x^{(1)}, k=1$, 转 2°；

2° 令 $I_k=\{i \mid g_i(x^{(k)})\ge 0\}$

$$\left\{\begin{array}{l} \text{若 } I_k=\Phi, \text{则 } x^{(k)}\text{为初始内点。} \\ \text{否则，取 } j \text{使 } g_j(x^{(k)})=\max\{g_i(x^{(k)}) \mid i\in I_k\} \text{ 转 } 3^\circ; \end{array}\right.$$

3° 用闸函数法求解：

$$\left\{\begin{array}{ll} \text{Min} & g_j(x) \\ \text{s.t.} & g_i(x)<0, i\notin I_k \end{array}\right.$$

以 $x^{(k)}$ 为初始内点，得到解 $x^{(k+1)}$ 转 4°；

$$4^\circ\left\{\begin{array}{l} \text{若 } g_j(x^{(k+1)})\ge 0, \text{停，说明 } S_0=\Phi. \\ \text{否则 置 } k=k+1, \text{转 } 2^\circ\text{。} \end{array}\right.$$

罚函数法与闸函数法的缺点：

1°当罚函数法（闸函数法）的$\mu \to \infty$ （ $\mu \to 0^+$）时，惩罚项趋于 $+\infty \cdot 0$（$0 \cdot +\infty$）形式，导致计算困难。

2°计算一系列无约束问题，故计算量大。

3.乘子法

$$
(fhD)\begin{cases} \min \quad f(x) & f: R^n \to R \\ s.t.\ h(x)=0 & h: R^n \to R^l \\ x\in D & D\subseteq R^n\text{是一个集合，常由简单约束构成。} \end{cases}
$$

用增广Lagrange函数代替$f(x)$:

$$
\phi(x,v,\mu) = f(x) + \sum_{i=1}^{l} v_i h_i(x) + \sum_{i=1}^{l} \mu_i h_i^2(x)
$$

其中：　$v \in \mathrm{R}^l$为乘子，　$\mu \in \mathrm{R}^l$为罚因子。

优点：不用 μ趋向无穷，只要 μ足够大，则通过求 ϕ的极小就得到 opt.

求解 $\begin{cases} \mathrm{Min} \quad \phi(x, v^{(k)}, \mu^{(k)}) \\ \mathrm{s.t.} \quad x \in D \end{cases}$ 得到 $x^{(k+1)}, k = 0,1,2,\cdots$

若$h(x^{(k+1)}) = 0$，得到解 $x^{(k+1)}$及乘子 $v^{(k)}$；否则，调整 $v^{(k)}$及$\mu^{(k)}$。

可以证明：

存在μ'，当$\mu > \mu'$时　存在v^*

$$\begin{cases} \text{Min} \quad \phi(x, v^*, \mu) \\ \text{s.t.} \quad x \in D \end{cases}$$

的最优解，即原问题的解。

这个结论说明v的选取与μ无关，即不用很大的μ

一般问题：

$$(fghD)\begin{cases} \text{Min} \quad f(x) \\ \text{s.t.} \ \ g(x) \le 0 \\ \quad h(x) = 0 \\ \quad x \in D \end{cases} \xrightarrow{\text{引入松弛变量}} \begin{cases} \text{Min} \ \ f(x) \\ \text{s.t.} \ \ g(x) + z = 0 \\ h(x) = 0 \\ x \in D' = \{(x, z) \mid x \in D, z \ge 0\} \end{cases}$$

1.解线性约束问题的既约梯度法

(1)问题： $(P)\begin{cases} \min \quad f(x) \\ s.t. \quad Ax=b \\ \quad x\ge 0 \end{cases}$ $A_{m\times n}$, 秩$(A)=m, b\in \mathrm{R}^m$

基本思想：利用等式约束条件，把一部分变量（基变量），用另一部分独立变量（非基变量）来表示，然后代入目标函数，得到维数降低了的问题。降维问题的梯度为既约梯度。

可行集： $S=\{x\mid Ax=b, x\ge 0\}$ 多面体同(LP)的S.

(2)非退化假设：

1° A的任意m列线性无关；

2° S的每个极点都有 m个正分量。（$B^{-1}b>0$）

(3)既约梯度及搜索方向：

$\forall x\in S$, 存在分解 $A=(B,N)$, $B_{m\times m}$非奇异，

$x=\begin{pmatrix} x_B \\ x_N \end{pmatrix}$使$x_B>0, x_N\ge 0$ $\quad x_B$ ——基变量， x_N ——非基变量

相应 $\nabla f(x)=\begin{pmatrix} \nabla_B f(x) \\ \nabla_N f(x) \end{pmatrix}$, 称$r_N^{\mathrm{T}}=\nabla_N f^{\mathrm{T}}(x)-\nabla_B f^{\mathrm{T}}(x)B^{-1}N$为既约梯度

$$Ax=b$$
$$\Downarrow$$
$$x_B=B^{-1}b-B^{-1}Nx_N$$
$$\Downarrow$$
$$f(x)=f(x_B,x_N)$$
$$=f(B^{-1}b-B^{-1}Nx_N,x_N)=g(x_N)$$
$$\Downarrow$$
$$r_N=\nabla_N g(x_N)$$

$\triangleleft$ 寻找下降可行方向： d

(1)引理： d为可行方向 $\Leftrightarrow \begin{cases} Ad = 0 \\ d_j \geq 0, \text{当} x_j = 0\text{时}. \end{cases}$

证明 ." $\rightarrow$ " d为可行方向，即 $\exists \delta > 0$, 当 $\lambda \in (0, \delta)$ 时，

$$A(x + \lambda d) = Ax + \lambda Ad = b, \text{又} Ax = b, \lambda > 0, \text{故} Ad = 0$$

由于 $x_j + \lambda d_j \geq 0$, 故当 $x_j = 0$时，有 $d_j \geq 0$。

" $\leftarrow$ " $\forall \lambda > 0$, 由 $Ad = 0$, $A(x + \lambda d) = Ax = b$。

取 $\theta = \min\{ -\frac{x_j}{d_j} \mid d_j < 0\}$ 则 $\lambda \in (0, \theta)$ 时

有 $x + \lambda d \geq 0$ 即 $x + \lambda d \in S$.故 d为可行方向。

考虑分解$d=\begin{pmatrix} d_B \\ d_N \end{pmatrix}$.

根据$Ad=[B \quad N]\begin{pmatrix} d_B \\ d_N \end{pmatrix}=Bd_B+Nd_N=0$

得到$d_B=-B^{-1}Nd_N$

故要使d可行，可取在d_N中，对应$x_j=0, d_j\geq 0$及$d_B=-B^{-1}Nd_N$

(2)下降方向d:要求$\nabla f^T(x)d<0$

分解：$\nabla f^T(x)d=\nabla_B f^T(x)d_B+\nabla_N f^T(x)d_N$

$$=\nabla_B f^T(x)(-B^{-1}Nd_N)+\nabla_N f^T(x)d_N$$

$$=(\nabla_N f^T(x)-\nabla_B f^T(x)B^{-1}N)d_N$$

$$=r_N^T d_N<0$$

(3)　结合(1)、(2)的一种产生下降可行方 向d的方案：

$$d_N : d_j = \begin{cases} -r_j & \text{当} r_j \le 0 \text{时} \\ -x_j r_j & \text{当} r_j > 0 \text{时} \end{cases}$$

其中：r_j为r_N的分量。

$$d_B = -B^{-1} N d_N$$

定理：按上述方案产生 方向d,那么

1°若$d \neq 0$,d为(P)的下降可行方向；

2°若$d = 0 \Leftrightarrow x$是$K-T$点。

证明 .

1° 对 $x_j = 0$ $\begin{cases} r_j \leq 0 \Rightarrow d_j = -r_j \geq 0 \\ r_j > 0 \Rightarrow d_j = -x_j r_j = 0 \end{cases}$

故总有 $d_j \geq 0$. $d_B = -B^{-1}Nd_N$ 保证 $Ad = 0$

又 $r_N^T d_N = \sum_{j\in N} r_j d_j$ $\quad r_j d_j = \begin{cases} -r_j^2 & ,r_j \leq 0 \\ -x_j r_j^2 & ,r_j > 0 \end{cases}$

$\downarrow$

≤ 0

由于 $d \neq 0$,至少一个 r_j或$x_j r_j$非零，于是 $r_N^T d_N < 0$.证毕。

2° "$\to$" $d=0$

$i)$可得$r_N \geq 0$:反证。若存在$r_j<0(j\in N)$

那么，$d_j=-r_j>0$与$d_j=0$矛盾；

$ii)$取$u_B=0, u_N=r_N\geq 0$,则$u^Tx=u_N^Tx_N=0$

原因：$u_N=r_N$，当$r_j>0$时，$d_j=-x_jr_j=0$故$x_j=0$;

$iii)$取$v=-(\nabla_B f^T(x)B^{-1})^T\in R^m$,可得

$$K-T\text{条件}:\begin{cases}\nabla f(x)+A^Tv-u=0\text{即}\begin{cases}\nabla_B f^T(x)-\nabla_B f^T(x)B^{-1}B-u_B^T=0\\ \nabla_N f^T(x)-\nabla_B f^T(x)B^{-1}N-u_N^T=0\end{cases}\\ u\geq 0\\ u^Tx=0\end{cases}$$

得证。

"$\Leftarrow$" x是K-T点

则有

$$\begin{cases} \nabla f(x) + A^T v - u = 0 \\ u \geq 0 \\ u^T x = 0 \end{cases}$$

由于$x \geq 0$，$u \geq 0$, 于是$u^T x = 0 \Rightarrow u_B^T x_B = 0$

再由非退化假设$x_B > 0$，故$u_B = 0$.

由K-T条件的第一个式子有$\nabla_B f(x) + B^T v - u_B = 0 \Rightarrow \nabla_B f(x) + B^T v = 0$

得到$v^T = -\nabla_B f^T(x) B^{-1}$

将v带入$\nabla_N f(x) + N^T v - u_N = 0$,得

$$u_N^T = \nabla_N f^T(x) - \nabla_B f^T(x) B^{-1} N = r_N^T$$

即

$$\begin{cases} r_N^T = u_N^T \geq 0 \\ r_N^T x_N = u_N^T x_N = 0 \Rightarrow r_j x_j = 0 \end{cases}$$

由于

$$d_j = \begin{cases} -r_j, & 当 r_j \leq 0 \\ -x_j r_j, & 当 r_j > 0 \end{cases}$$

故$d = 0$

算法流程：

$x^{(1)} \in S,\ k=1$

$J_k=\{j \mid x_j$为$x^{(k)}$中最大m个正分量之一$\}$

$B=[\ldots,a_j,\ldots],(j\in J_k)$

$N=[\ldots,a_j,\ldots],(j\notin J_k)$

$r_N^{T}=\nabla_N f^{T}(x^{(k)})-\nabla_B f^{T}(x^{(k)})B^{-1}N$

$$d_j=\begin{cases}-r_j, & 当r_j\le 0\\ -x_j r_j, & 当r_j>0\end{cases}$$

$d_B=-B^{-1}Nd_N$

$d=0$?

Y: Stop; $x^{(k)}$是K-T点

N:

$$\hat{\lambda}=\begin{cases}\min\ \{-x_j^{(k)}/d_j \mid d_j<0\}, & 当d<0\\ +\infty, & 否则\ d\ge 0\end{cases}$$

为使$x^{(k+1)}=x^{(k)}+\lambda_k d$不违背非负限制

解
$$\begin{cases}\min\ f(x^{(k)}+\lambda d)\\ s.t.\ \ 0\le\lambda\le\hat{\lambda}\end{cases}$$
得
$x^{(k+1)}=x^{(k)}+\lambda_k d$

$k=k+1$

例：

$$\min 2x_1^2 + x_2^2$$

$$s.t.\begin{cases} x_1 - x_2 \le 2 \\ -2x_1 + x_2 \le 1 \\ x_1, x_2 \ge 0 \end{cases}$$

取初始点$x^{(1)} = (1,3)^T$.

解：先化为等式约束问题

$$\begin{cases} x_1 - x_2 + x_3 = 2 \\ -2x_1 + x_2 + x_4 = 1 \\ x_1, x_2, x_3, x_4 \ge 0 \end{cases}$$

初始点$x^{(1)} = (1,3,4,0)^T$

第一次迭代：

$J_1 = \{2,3\}$.

$$A = \begin{pmatrix} 1 & -1 & 1 & 0 \\ -2 & 1 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}, \quad N = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$$

$r_N^T = \nabla_N f^T(x) - \nabla_B f^T(x) B^{-1} N = (16, -6)^T$.

$d_j = \begin{cases} -r_j, & 当r_j \le 0 \\ -x_j r_j, & 当r_j \ge 0 \end{cases}$, $d_N = (-16, 6)^T$

$d_B = -B^{-1} N d_N = (-38, -22)^T$

$d^{(1)} = (-16, -38, -22, 6)^T$

$\hat{\lambda} = \min\left\{-\frac{1}{16}, -\frac{3}{-38}, -\frac{4}{-22}\right\} = \frac{1}{16}$

$$f(x^{(1)}+\lambda d^{(1)}) = 2(1-16\lambda)^2 + (3-38\lambda)^2$$

求 $min\ f(x^{(1)}+\lambda d^{(1)})$

s.t. $0 \le \lambda \le 1/16$

求导得 $\frac{d}{d\lambda} f(x^{(1)}+\lambda d^{(1)})=0 \Rightarrow \ \lambda = \frac{1}{14} \ge \frac{1}{16}$

故$\lambda_1 = \frac{1}{16}$

$$x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = \left(0, \frac{5}{8}, \frac{21}{8}, \frac{3}{8}\right)^T$$

第二次迭代得$x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = (0, 0, 2, 1)^T$

其中$\lambda_2 = \frac{1}{2}$，$d^{(2)} = \left(0, -\frac{5}{4}, -\frac{5}{4}, \frac{5}{4}\right)^T$

第二次迭代得$d^{(3)} = (0,0,0,0)^T$

由于问题为凸规则，所以$x^{(3)}$为最优解。

2.广义既约梯度法（GRG）

$$\text{标准形式}\quad (P)\begin{cases}\text{Min}\ \ f(x)\\ \text{s.t.}\quad h(x)=0\\ \quad a\le x\le b\end{cases}$$

其中：$f:R^n\to R, h:\mathrm{R}^n\to R^l$连续可微。

$a,b\in\mathrm{R}^n, a,b$的分量允许$\pm\infty$，且$a<b$.

记 $S=\{x\mid h(x)=0, a\le x\le b\}$

非退化假设：

$$\begin{cases}1^\circ\ \forall x\in S, \exists\text{分解}\quad x=\begin{pmatrix}y\\ z\end{pmatrix}, \text{使 } y\in\mathrm{R}^l\\ \qquad z\in\mathrm{R}^{n-l}, \text{记 } a=\begin{pmatrix}a_y\\ a_z\end{pmatrix}, b=\begin{pmatrix}b_y\\ b_z\end{pmatrix}\\ \qquad\qquad \text{使 } a_y<y<b_y\\ 2^\circ\ \dfrac{\partial h(x)}{\partial y}\text{非奇异}\end{cases}$$

广义既约梯度：

$$r_z^{\mathrm{T}} = \nabla_z f^{\mathrm{T}}(x) - \nabla_y f^{\mathrm{T}}(x)\left(\frac{\partial h^{\mathrm{T}}(x)}{\partial y}\right)^{-1}\left(\frac{\partial h(x)}{\partial z}\right)^{\mathrm{T}}$$

取方向：

令 $J = \{i \mid z_i = a_i \text{且} (r_z)_i > 0 \text{或} \mid z_i = b_i \text{且} (r_z)_i < 0\}$

$$(d_z)_i = \begin{cases} 0 & i \in J \\ -(r_z)_i \text{或} 0 & i \notin J \end{cases}$$

$$d_y = -\left(\frac{\partial h^{\mathrm{T}}(x)}{\partial y}\right)^{-1}\left(\frac{\partial h^{\mathrm{T}}(x)}{\partial z}\right) d_z$$

同样有结论：

1° 当 $d \neq 0$时为下降可行方向；

2° $d = 0 \Leftrightarrow x$是K－T点。