



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

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

分量形式略

约束集 $S = \{x | g(x) \leq 0, h(x) = 0\}$

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，

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

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：

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

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

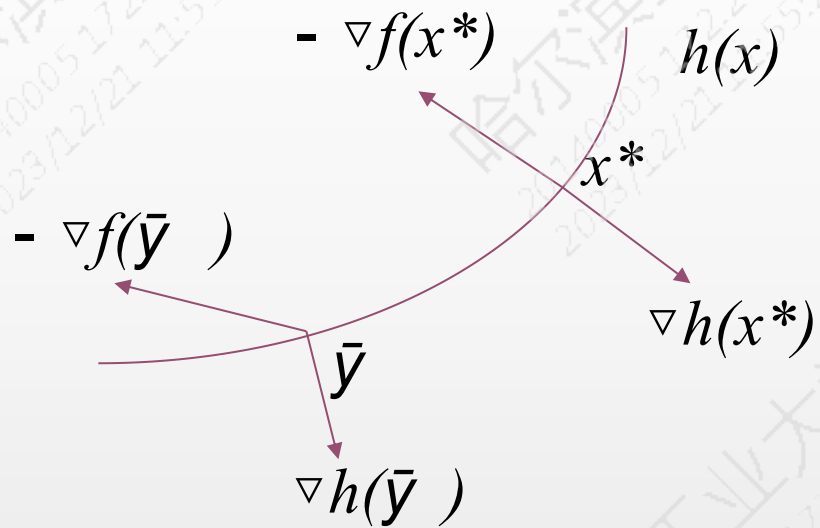
$$\nabla f(x^*) + \sum_{j=1}^l v_j^* \nabla h_j(x^*) = 0$$

矩阵形式：

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



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

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



2.不等式约束问题的Kuhn-Tucker条件

考虑问题

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i=1, 2, \dots, m \end{aligned}$$

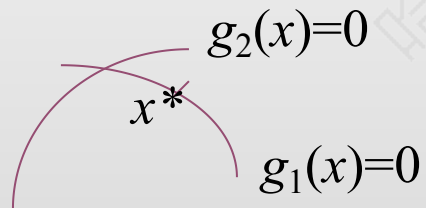
设 $x^* \in S = \{x | g_i(x) \leq 0, \quad i=1, 2, \dots, m\}$

令

$$I = \{i | g_i(x^*) = 0, \quad i=1, 2, \dots, m\}$$

称 I 为 x^* 点处的起作用集（紧约束集）。

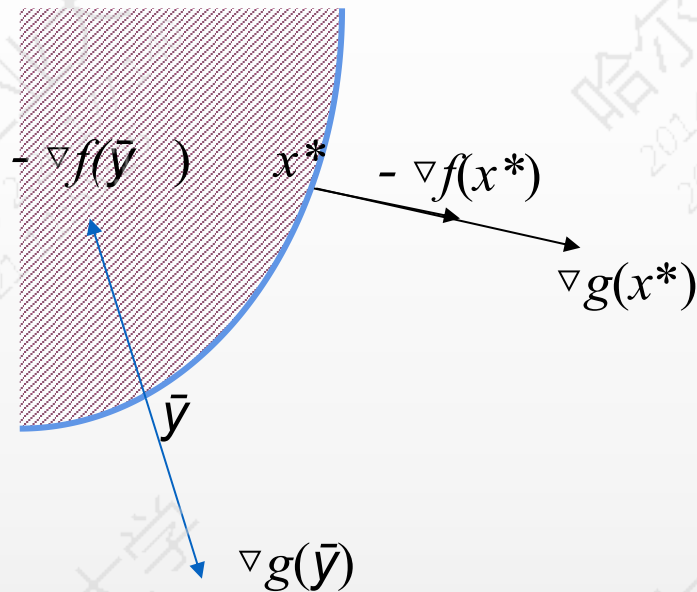
如果 x^* 是 l.opt., 对每一个约束函数来说, 只有当它是起作用约束时, 才产生影响, 如:



$g_1(x^*)=0$, 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 $S=\{x|g_i(x) \leq 0\}$, $x^* \in S$, 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 $u_i \geq 0, i \in I$ 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

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

$$u_i \geq 0, i = 1, \dots, m$$

$$u^T g(x^*) = 0, (\text{互补松弛条件})$$

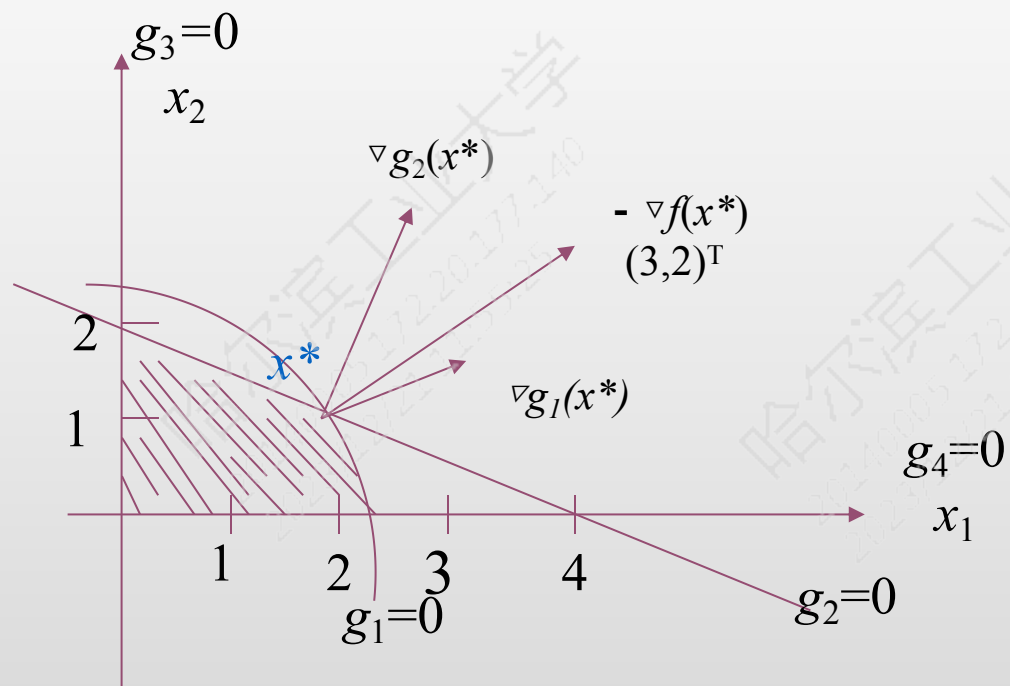
$$\Leftrightarrow u_i g_i(x^*) = 0, i = 1, \dots, m$$

满足K-T条件的点 x^* 被称为K-T点。



例

$$\left\{ \begin{array}{ll} \text{Min} & f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{s.t.} & g_1(x_1, x_2) = x_1^2 + x_2^2 - 5 \leq 0 \\ & g_2(x_1, x_2) = x_1 + 2x_2 - 4 \leq 0 \\ & g_3(x_1, x_2) = -x_1 \leq 0 \\ & g_4(x_1, x_2) = -x_2 \leq 0 \end{array} \right.$$





在 x^* 点

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

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

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

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

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

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

下面从使用K-T条件的角度来求解

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

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



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



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

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

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

$$\text{得 } u_1 = \frac{1}{3}, u_2 = \frac{2}{3} > 0$$

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 $x = (0, \pm\sqrt{5})^T$

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

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



g_3, g_4 交点: $x = (0, 0)^T \in S$ $I = \{3, 4\}$ 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 $u_2 = u_3 = u_4 = 0$

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

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

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

定理： 问题 (fgh) , $x^* \in S = \{x \mid g(x) \leq 0, h(x) = 0\}$, 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 ($j = 1, 2, \dots, l$)

在 x^* 的某邻域内连续可微。 (CQ , 约束规格)。

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



如果 x^* – 1.opt. 那么 $\exists u_i^* \geq 0, i \in I,$

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

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

如果还有 $g_i(x)(i \notin I)$ 在 x^* 亦可微, 那么

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



4.凸规划的一阶充分性条件

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

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

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 $u_i \geq 0, i \in I, v_j \in \mathbb{R}, j = 1, \dots, l$. 则 x^* 为 (f, g, h) 的 g.opt.



考虑 (LP) 的标准形式：

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

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

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

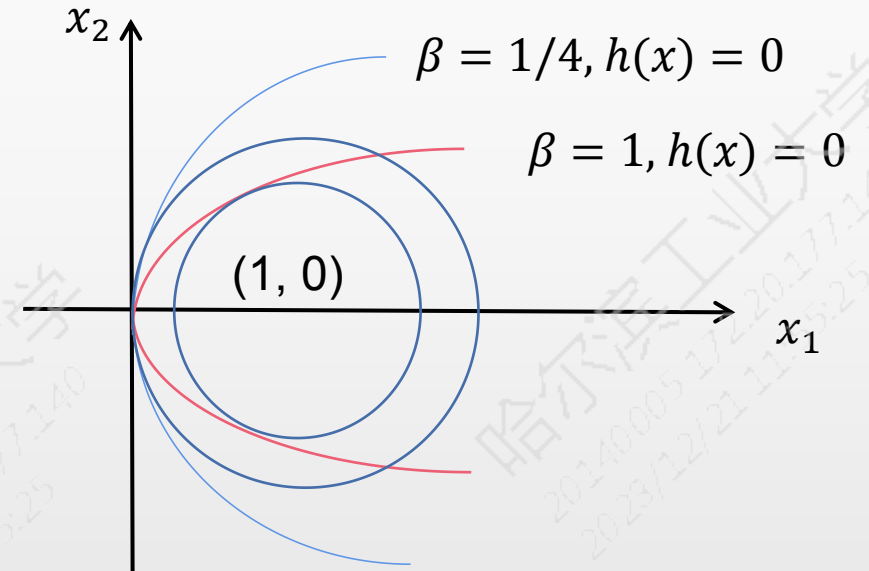
但，当 $\beta = 1/4$ 时， x^* 是 opt.

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 $L(x, v) = f(x) + v h(x)$

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 $\beta = 1/2$ 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 $x^* \in S$, $f(x)$, $g_i(x) (i \in I)$, $h_j(x)$, $j = 0, \dots, l$, 在 x^* 处二次可微, $g_i(x) (i \notin I)$ 在 x^* 处连续。 $\{\nabla g_i(x^*) (i \in I), \nabla h_j(x^*), j = 0, \dots, l\}$ 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 $S_h = \{d | \nabla h_j^T(x^*) d = 0, j = 1, 2, \dots, l\}$,

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

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 $u_i (i \in I)$ ， $v_j (j = 0, \dots, l)$ 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 $d \in \{d | \nabla h_j^T(x^*)d = 0, j = 1, 2, \dots, l\} \cap \{d | \nabla g_i^T(x^*)d = 0, i \in I\}$ ，**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:

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

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

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

取 $\mu > 0$, 可构造

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

$f(x) + \mu\alpha(x)$ -- 辅助函数

Min $f(x) + \mu\alpha(x)$ -- 辅助问题

$\phi(t), \varphi(t)$ 的典型取法:

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

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

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

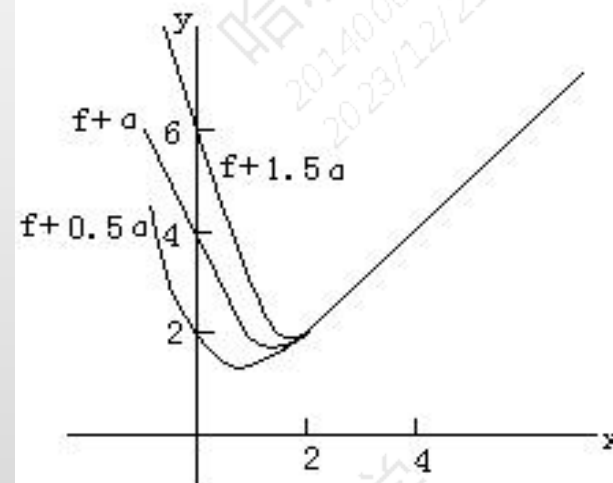
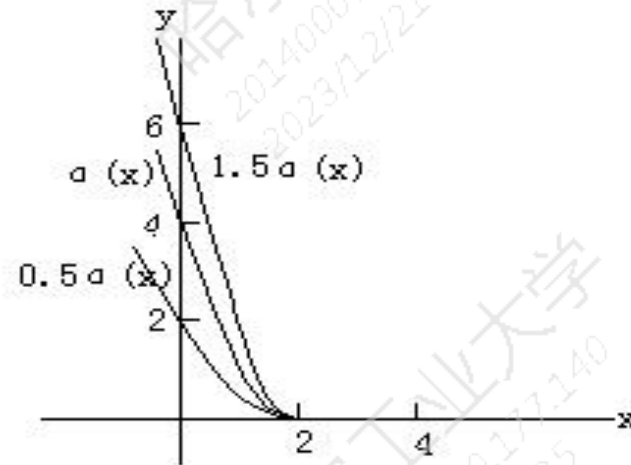
• 当 $\mu \rightarrow \infty$ 时, $\text{Min } f(x) + \mu\alpha(x) \rightarrow f(x^*) = x^* = 2$

• 解析解 : 辅助函数

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

当 $x < 2$ 时, $g(x, \mu)$ 的驻点 $\bar{x} = \frac{4\mu - 1}{2\mu} \xrightarrow{\mu \rightarrow \infty} 2$ 故 $x^* = 2$ --opt.

当 $x \geq 2$ 时, $g(x, \mu)$ 的最小值点 $\tilde{x} = 2$





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 : $(fgh), S = \{x \mid g(x) \leq 0, h(x) = 0\} \neq \Phi,$

在引理假设下, 设存在单调增加的正数列 $\{\mu_k\}$

即 $0 < \mu_1 < \mu_2 < \dots < \mu_k < \dots$ 有 $\{x_{\mu_k}\} \xrightarrow{k \rightarrow \infty} x^*$

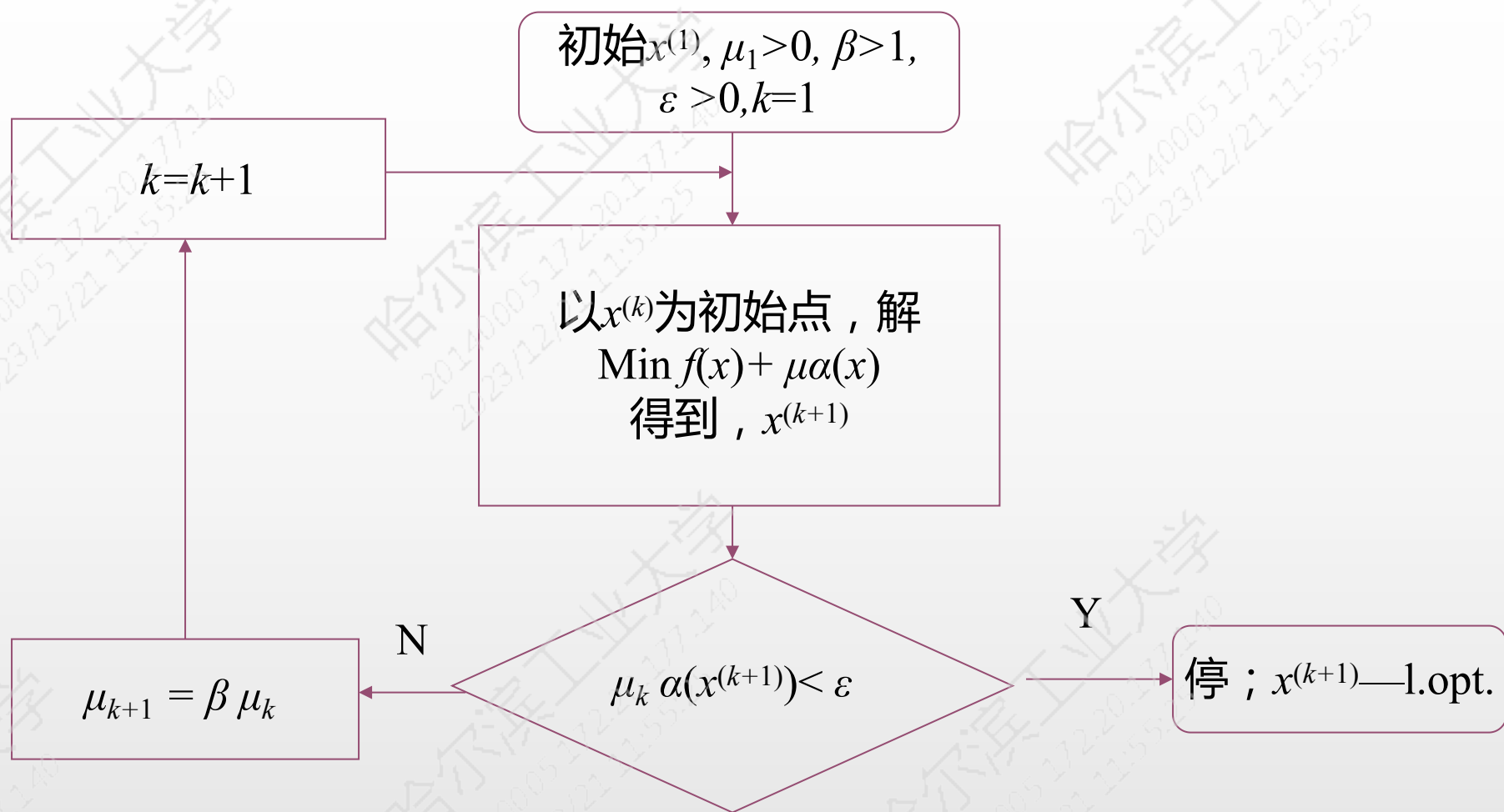
那么, $1^\circ \inf\{f(x) \mid x \in S\} = \sup\{\theta(\mu) \mid \mu \geq 0\} = \lim_{\mu \rightarrow \infty} \theta(\mu)$

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

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

则 $x_\mu = opt.$

算法流程：





2. 闸函数法 (内点罚函数法)

$$(fg) \begin{cases} \text{Min } f(x) & f: \mathbb{R}^n \rightarrow \mathbb{R} \\ \text{s.t. } g(x) \leq 0 & g: \mathbb{R}^n \rightarrow \mathbb{R}^m \end{cases}$$

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

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



典型取法： $\phi(t) = -\frac{1}{t}$

或： $\phi(t) = |\ln(-t)|$

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

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

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

辅助问题 $\min f(x) + \mu B(x)$



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

解 闸函数
$$B(x) = \frac{1}{x-2}, x > 2$$

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

目标函数关于 x 是凸的, 求驻点:

$$x_{\mu} = 2 + \sqrt{\mu}$$

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

$$g(x_{\mu}, \mu) \xrightarrow{\mu \rightarrow 0^+} 2 \text{ -- 最优值 (原问题)}$$



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:

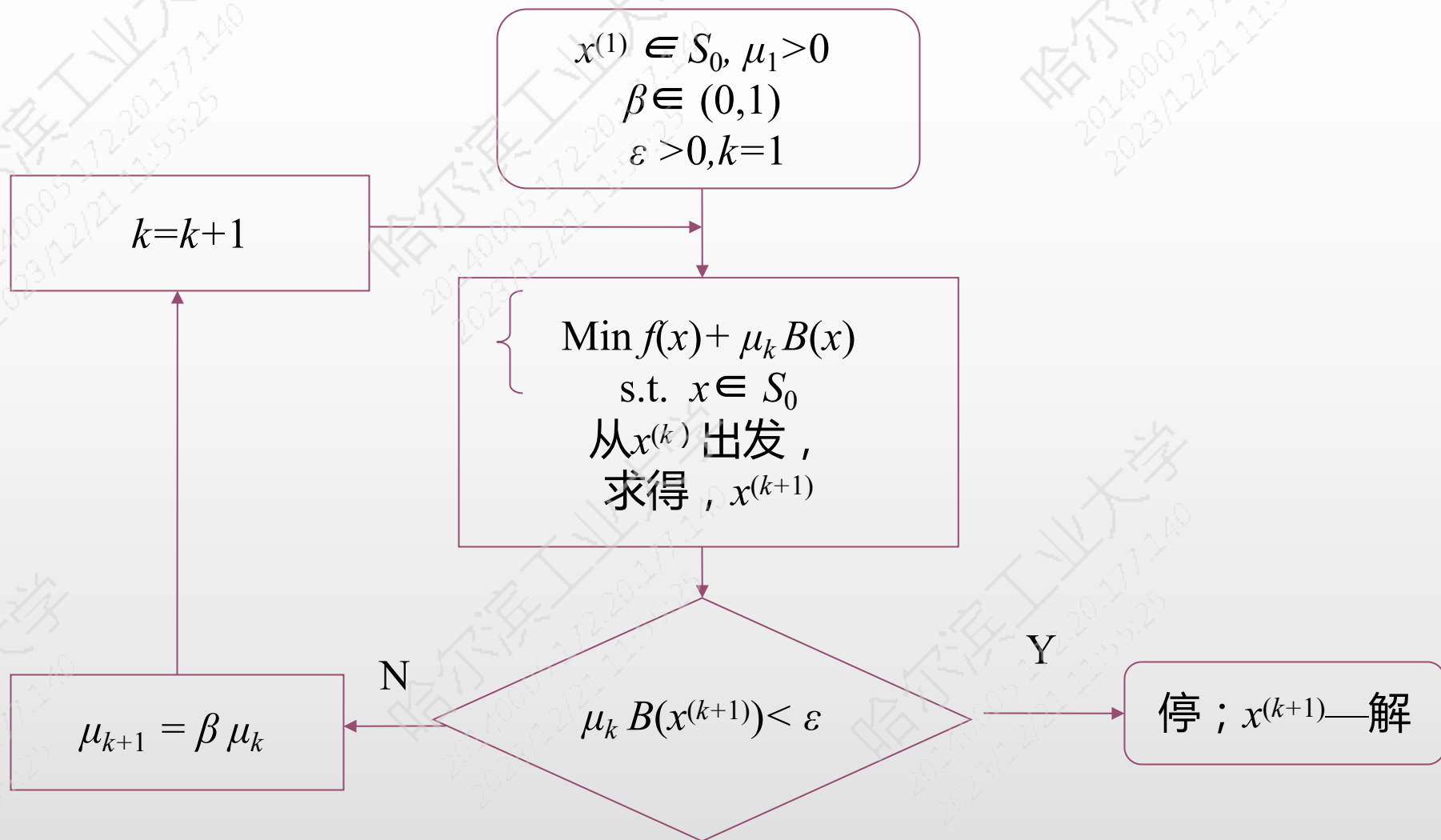
定理: $(fg), f, g$ 连续, $S_0 \neq \Phi$, 最优解 $x^* \in S_0$

则 $1^\circ \min\{f(x) \mid x \in S\} = \inf\{\theta(\mu) \mid \mu > 0\} = \lim_{\mu \rightarrow 0^+} \theta(\mu)$

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 $\lim_{\mu \rightarrow 0^+} \mu B(x_\mu) = 0$

算法流程：





- 求初始内点:

1° $\forall x^{(1)}, k = 1$, 转 2°;

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

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

3° 用罚函数法求解:

$$\begin{cases} \text{Min } g_j(x) \\ \text{s.t. } g_i(x) < 0, i \notin I_k \end{cases}$$

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

4° $\begin{cases} \text{若 } g_j(x^{(k+1)}) \geq 0, \text{ 停, 说明 } S_0 = \Phi. \\ \text{否则 置 } k = k + 1, \text{ 转 } 2^\circ. \end{cases}$



罚函数法与闸函数法的缺点：

1°当罚函数法（闸函数法）的 $\mu \rightarrow \infty$ （ $\mu \rightarrow 0^+$ ）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

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

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: $v \in R^l$ 为乘子, $\mu \in R^l$ 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 .

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

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^*

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

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：

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



6.3 既约梯度法

1. 解线性约束问题的既约梯度法

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

可行集: $S = \{x \mid Ax = b, x \geq 0\}$ 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,

$$x = \begin{pmatrix} x_B \\ x_N \end{pmatrix} \text{使 } x_B > 0, x_N \geq 0 \quad x_B \text{ -- 基变量, } x_N \text{ -- 非基变量}$$

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

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。

$$\begin{aligned} Ax &= b \\ \downarrow \\ x_B &= B^{-1}b - B^{-1}Nx_N \\ \downarrow \\ f(x) &= f(x_B, x_N) \\ &= f(B^{-1}b - B^{-1}Nx_N, x_N) = g(x_N) \\ \downarrow \\ r_N &= \nabla_N g(x_N) \end{aligned}$$



◁ 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 $d = \begin{pmatrix} d_B \\ d_N \end{pmatrix}$.

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

$$\text{得到 } d_B = -B^{-1}Nd_N$$

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 $\nabla f^T(x)d < 0$

$$\begin{aligned} \text{分解: } \nabla f^T(x)d &= \nabla_B f^T(x)d_B + \nabla_N f^T(x)d_N \\ &= \nabla_B f^T(x)(-B^{-1}Nd_N) + \nabla_N f^T(x)d_N \\ &= (\nabla_N f^T(x) - \nabla_B f^T(x)B^{-1}N)d_N \\ &= r_N^T d_N < 0 \end{aligned}$$



(3) 结合(1)、(2)的一种产生下降可行方向 d 的方案：

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

其中： r_j 为 r_N 的分量。

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



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 .

$$1^\circ \text{ 对 } x_j = 0 \quad \begin{cases} r_j \leq 0 \Rightarrow d_j = -r_j \geq 0 \\ r_j > 0 \Rightarrow d_j = -x_j r_j = 0 \end{cases}$$

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

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

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。



$$2^\circ \rightarrow d = 0$$

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 $u^T x = u_N^T x_N = 0$

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

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

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

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

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

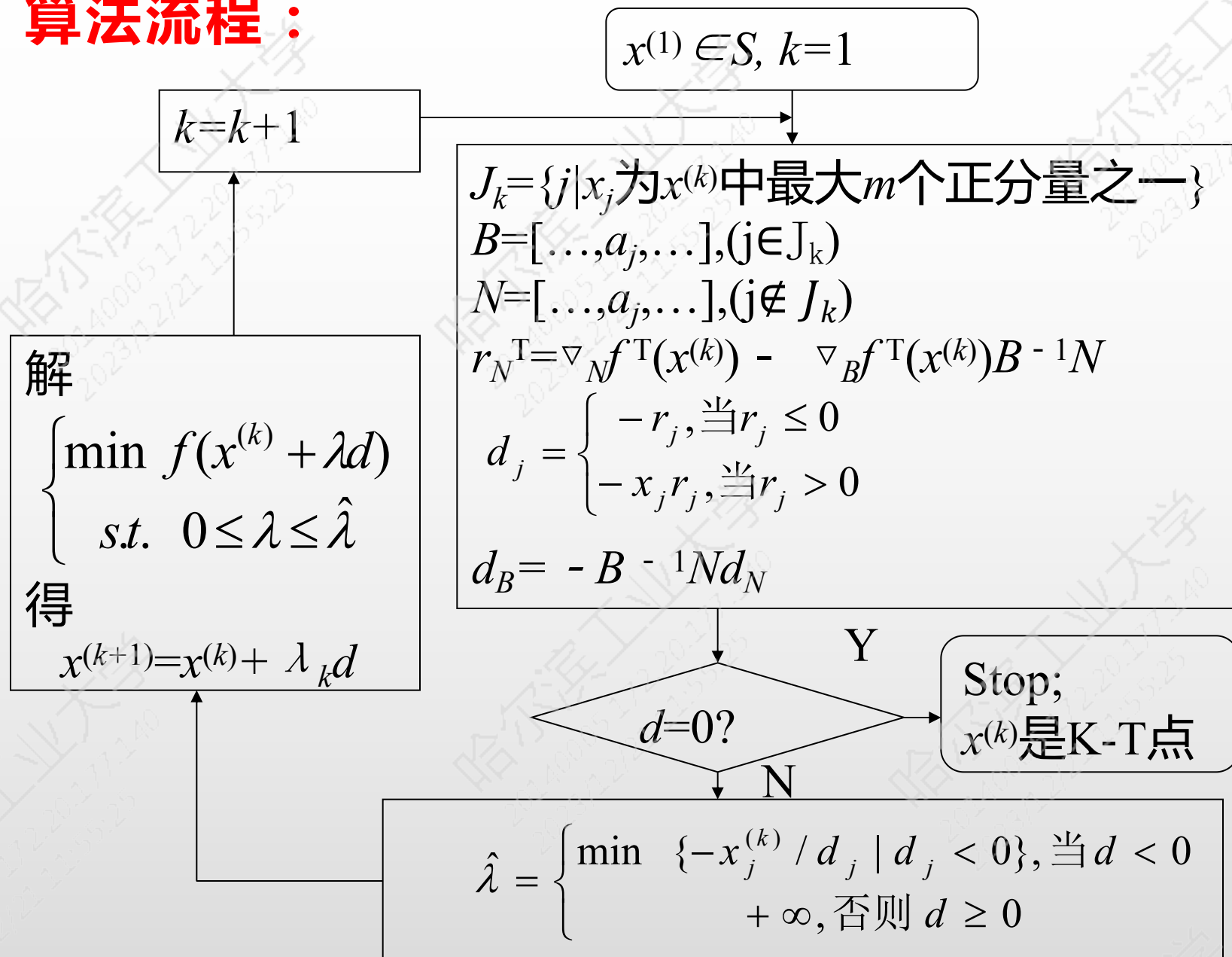
由于

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

故 $d = 0$



算法流程：



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



例：

$$\begin{aligned} \min & 2x_1^2 + x_2^2 \\ \text{s. t. } & \begin{cases} x_1 - x_2 \leq 2 \\ -2x_1 + x_2 \leq 1 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

取初始点 $x^{(1)} = (1, 3)^T$.

解：先化为等式约束问题

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

初始点 $x^{(1)} = (1, 3, 4, 0)^T$

第一次迭代：

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

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

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

$$d_j = \begin{cases} -r_j, & \text{当 } r_j \leq 0 \\ -x_j r_j, & \text{当 } r_j \geq 0 \end{cases}, \quad d_N = (-16, 6)^T$$

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

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

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



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

$$\begin{aligned} &\text{求 } \min f(x^{(1)} + \lambda d^{(1)}) \\ &\text{s.t. } 0 \leq \lambda \leq 1/16 \end{aligned}$$

$$\begin{aligned} &\text{求导得 } \frac{d}{d\lambda} f(x^{(1)} + \lambda d^{(1)}) = 0 \Rightarrow \lambda = \frac{1}{14} \geq \frac{1}{16} \\ &\text{故 } \lambda_1 = \frac{1}{16} \end{aligned}$$

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

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

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

$$\text{第二次迭代得 } d^{(3)} = (0, 0, 0, 0)^T$$

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)

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

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

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

记 $S = \{x \mid h(x) = 0, a \leq x \leq b\}$

非退化假设:

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





广义既约梯度:

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

取方向:

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

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

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

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。