第四章 约束非线性规划

自动化学院

§ 4.1 最优性条件

本章主要研究如下优化模型：

$$
\begin{aligned}
\min\ & f(x) \\
s.t.\ & h_j(x)=0,\ \ j=1,\cdots,l \\
& g_i(x)\ge 0,\ \ i=1,\cdots,m \\
& x\in R^n.
\end{aligned}
$$

其中可行域为

$$D=\{x\mid x\in R^n, h_j(x)=0, g_i(x)\ge 0, j=1,\cdots,l, i=1,\cdots,m\}.$$

§ 4.1 最优性条件

设$x' \in D$, 若有$g_i(x') = 0$，则称不等式约束$g_i(x) \geq 0$为点x'处的起作用约束。且将下标集$I = \{i \mid g_i(x') = 0, 1 \leq i \leq \mathrm{m}\}$，称为点$x'$的起作用下标集。

若有$g_i(x') > 0$，则称不等式约束$g_i(x) \geq 0$为点x'处的不起作用约束。

显然，等式约束$h_j(x) = 0$都是起作用约束。

§ 4.1 最优性条件

只含不等式约束问题的最优性条件

$$\begin{aligned} &\min \quad f(x) \\ &s.t. \quad g_i(x) \ge 0 \\ &\qquad\quad x \in R^n \end{aligned}$$

其中可行域

$$D = \{x \mid x \in R^n, g_i(x) \ge 0, i = 1, \cdots, m\}.$$

设*x**是问题的极小点，则*x**可在可行域内或可行域的边界上。

(1) *x**在可行域内部，约束不起作用，变为无约束问题，*x**必满足$\nabla f(x^*) = 0.$

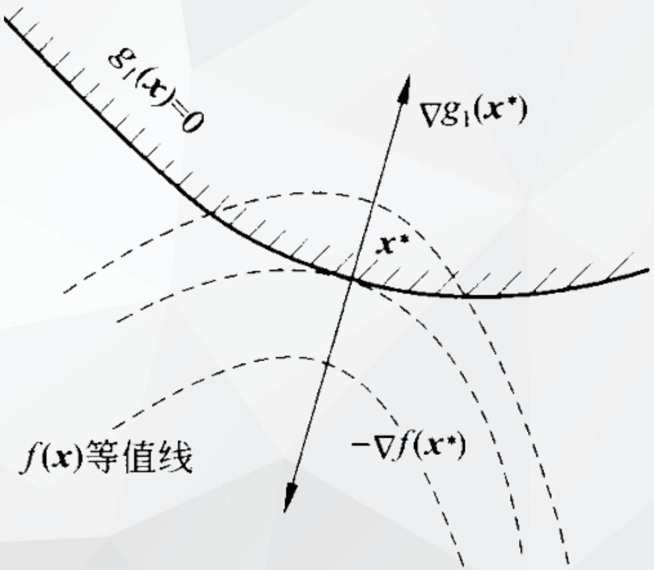
(2) *x**在可行域边界上，约束起作用。

§ 4.1 最优性条件

(i) 假设x^*位于一个约束条件的边界上，即x^*只有一个起作用约束。不失一般性，令$g_1(x) \geq 0$为x^*的起作用约束，故有$g_1(x^*) = 0$。

若x^*是局部最优解，则必有 $-\nabla f(x^*)$ 与 $\nabla g_1(x^*)$ 在同一直线上且方向相反，则存在实数$\gamma_1 \geq 0$，使得

$$\nabla f(x^*) - \gamma_1 \nabla g_1(x^*) = 0$$



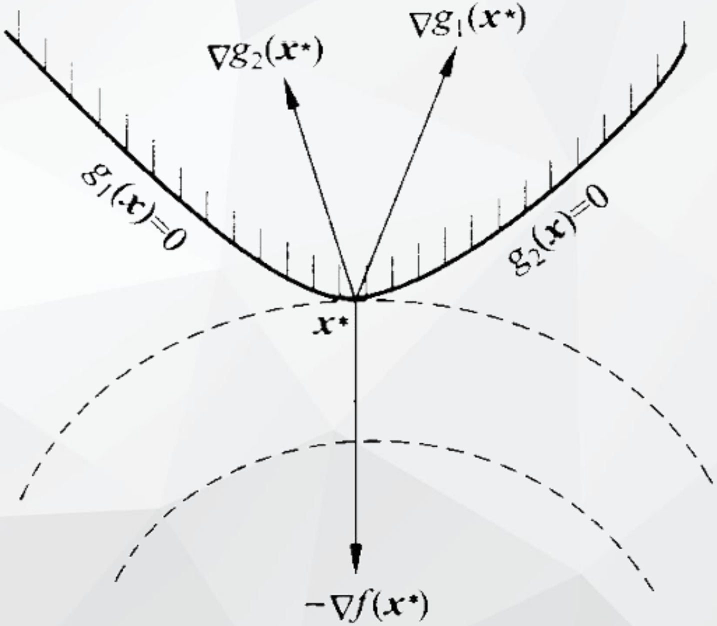
§ 4.1 最优性条件

(ii) 假设x^*位于两个约束条件的边界上，即x^*有两个起作用约束，$g_1(x^*)=0$，$g_2(x^*)=0$。

此时, $\nabla f(x^*)$**必位于** $\nabla g_1(x^*)$ **与** $\nabla g_2(x^*)$**所形成的夹角内。**

若x^*是局部最优解，且 $\nabla g_1(x^*)$ **与**$\nabla g_2(x^*)$ **线性无关，则** $\nabla f(x^*)$ **可由** $\nabla g_1(x^*)$ **与** $\nabla g_2(x^*)$ **的线性组合表示，即存在实数$\gamma_1,\gamma_2\geq 0$，使得**

$$\nabla f(x^*)-\gamma_1\nabla g_1(x^*)-\gamma_2\nabla g_2(x^*)=0$$



§ 4.1 最优性条件

推广到多个不等式约束，有

$$\begin{cases} \nabla f(x^*) - \sum_{i \in I} \gamma_i \nabla g_i(x^*) = 0 \\ \gamma_i \geq 0 \end{cases} \quad (*)$$

为了将不起作用约束也加入到上式中，增加一个条件：

$$\gamma_i g_i(x^*) = 0,\ i = 1, \cdots, m.$$

则(*)式可改写为

$$\begin{cases} \nabla f(x^*) - \sum_{i=1}^{m} \gamma_i \nabla g_i(x^*) = 0 \\ \gamma_i g_i(x^*) = 0,\ i = 1, \cdots, m. \\ \gamma_i \geq 0,\ i = 1, \cdots, m. \end{cases}$$

§ 4.1 最优性条件

定理1 (不等式约束问题的最优性条件)

设$x^* \in D$, f, g_i在x^*处可微，且$\nabla g_i(x^*)(i\in I)$线性无关。

若x^*是局部最优解，则存在 $\gamma^* = (\gamma_1^*, \gamma_2^*, \cdots, \gamma_m^*)$

使得

$$\begin{cases} \nabla f(x^*) - \sum_{i=1}^{m} \gamma_i^* \nabla g_i(x^*) = 0 \\ \gamma_i^* g_i(x^*) = 0,\ i = 1, \cdots, m. \\ \gamma_i^* \geq 0,\ i = 1, \cdots, m. \\ g_i(x^*) \geq 0. \end{cases}$$

$\longleftarrow$ KKT条件

§ 4.1 最优性条件

➢ 只含等式约束问题的最优性条件

$$\begin{aligned} &\min \quad f(x) \\ &s.t. \quad h(x)=0. \end{aligned}$$

设x^*是问题的极小点，在x^*处梯度$\nabla f(x)$和$\nabla h(x)$的方向必相同或相反，即存在$\mu \neq 0$，使得

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

§ 4.1 最优性条件

定理2 (等式约束问题的最优性条件)

设$x^*\in D$, f, h_j (j=1, ..., l) 在x^*的某一邻域内连续可微，

且$\nabla h_j(x^*)$线性无关。若x^*是局部最优解，则存在

$\mu^*=(\mu_1^*,\mu_2^*,\cdots,\mu_l^*)$，使得

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

KKT条件

§ 4.1 最优性条件

➢ 同时含有等式和不等式约束问题的最优性条件

定理3 设 $x^*\in D$, f, g_i, h_j 在 x^* 处可微，且 $\nabla g_i(x^*)$ $(i\in I)$, $\nabla h_j(x^*)$, $j=1, ..., l$ 线性无关。若 x^* 是局部最优解，则存在 $\gamma^*=(\gamma_1^*,\gamma_2^*,\cdots,\gamma_m^*)$ 和 $\mu^*=(\mu_1^*,\mu_2^*,\cdots,\mu_l^*)$，使得

$$\begin{cases} \nabla f(x^*)-\sum_{i=1}^{m}\gamma_i^*\nabla g_i(x^*)-\sum_{j=1}^{l}\mu_j^*\nabla h_j(x^*)=0 \\ \gamma_i^* g_i(x^*)=0,\ i=1,\cdots,m. \\ \gamma_i^*\geq 0,\ i=1,\cdots,m. \\ g_i(x^*)\geq 0. \\ h_j(x^*)=0. \end{cases}$$

$\longleftarrow$ KKT条件

$\downarrow$

最优解的一阶必要条件

满足KKT条件的可行点，称为KKT点。

§ 4.1 最优性条件

例1 求下列非线性规划的KKT点

$$\begin{aligned} \max\ & 14x_1 - x_1^2 + 6x_2 - x_2^2 + 7 \\ s.t.\ & x_1 + x_2 \le 2 \\ & x_1 + 2x_2 \le 3 \end{aligned}$$

解：将非线性规划化为一般形式

$$\begin{aligned} \min\ & f(x) = -14x_1 + x_1^2 - 6x_2 + x_2^2 - 7 \\ s.t.\ & g_1(x) = -x_1 - x_2 + 2 \ge 0 \\ & g_2(x) = -x_1 - 2x_2 + 3 \ge 0 \end{aligned}$$

§ 4.1 最优性条件

设$x^* = (x_1, x_2)^T$为KKT点，则满足如下条件

$$\begin{cases} \nabla f(x^*) - \gamma_1 \nabla g_1(x^*) - \gamma_2 \nabla g_2(x^*) = 0 \\ \gamma_i g_i(x^*) = 0,\ i = 1, 2. \\ \gamma_i \geq 0,\ i = 1, 2. \\ g_i(x^*) \geq 0,\ i = 1, 2. \end{cases}$$

其中，$\nabla f(x^*) = (-14 + 2x_1,\ -6 + 2x_2)^T$ $\nabla g_1(x^*) = (-1,\ -1)^T$ $\nabla g_2(x^*) = (-1,\ -2)^T$

因此，有

$$\begin{cases} -14 + 2x_1 + \gamma_1 + \gamma_2 = 0 \\ -6 + 2x_2 + \gamma_1 + 2\gamma_2 = 0 \\ \gamma_1(-x_1 - x_2 + 2) = 0 \\ \gamma_2(-x_1 - 2x_2 + 3) = 0 \\ \gamma_i \geq 0,\ i = 1, 2. \\ -x_1 - x_2 + 2 \geq 0 \\ -x_1 - 2x_2 + 3 \geq 0 \end{cases}$$

§ 4.1 最优性条件

(i) 两个约束条件均不起作用，则有$\gamma_1 = \gamma_2 = 0$

代入KKT条件，有

$$\begin{cases} -14 + 2x_1 = 0 \\ -6 + 2x_2 = 0 \end{cases}$$

解得

$$\begin{cases} x_1 = 7 \\ x_2 = 3 \end{cases}$$

但该点不满足$g_1(x^*) > 0$**和**$g_2(x^*) > 0$**，因此不是可行点。**

§ 4.1 最优性条件

(ii) g_1为起作用约束， g_2为不起作用约束，则有 $\gamma_2=0$
代入KKT条件，有

$$\begin{cases} -14+2x_1+\gamma_1=0 \\ -6+2x_2+\gamma_1=0 \\ \gamma_1(-x_1-x_2+2)=0 \end{cases}$$

解得

$$\begin{cases} x_1=3 \\ x_2=-1 \\ \gamma_1=8 \end{cases}$$

该点满足$g_1(x^*)=0$和$g_2(x^*)>0$，因此$x^*=(3,-1)^T$为KKT点。

§ 4.1 最优性条件

(iii) g_1为不起作用约束，g_2为起作用约束，则有$\gamma_1 = 0$

代入KKT条件，有

$$\begin{cases} -14 + 2x_1 + \gamma_1 = 0 \\ -6 + 2x_2 + \gamma_1 = 0 \\ \gamma_2(-x_1 - 2x_2 + 3) = 0 \end{cases}$$

解得

$$\begin{cases} x_1 = 5 \\ x_2 = -1 \\ \gamma_1 = 4 \end{cases}$$

该点满足$g_1(x^*) > 0$，因此不是可行点。

(iv) 两个约束条件均为起作用约束，则有$g_1(x^*) = 0, g_2(x^*) = 0$ 有

$$\begin{cases} -x_1 - x_2 + 2 = 0 \\ -x_1 - 2x_2 + 3 = 0 \end{cases}$$

解得

$$\begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases}$$

代入KKT条件，有

$$\begin{cases} -14 + 2 + \gamma_1 + \gamma_2 = 0 \\ -6 + 2 + \gamma_1 + 2\gamma_2 = 0 \end{cases} \longrightarrow \begin{cases} \gamma_1 = 20 \\ \gamma_2 = -8 \end{cases}$$

$\gamma_2 = -8 < 0$ **不满足条件，因此该点不是可行点。**

综上，解得为KKT点为$x^* = (3, -1)^T$，且 $\gamma_1 = 8, \gamma_2 = 0.$

例2 考虑非线性规划问题

$$\begin{aligned} \min\ & (x_1-3)^2+(x_2-2)^2 \\ s.t.\ & x_1^2+x_2^2\le 5 \\ & x_1+2x_2=4 \\ & x_1,x_2\ge 0. \end{aligned}$$

检验$x^*=(2,1)^T$是否为KKT点。

解：将非线性规划写作一般形式

$$\begin{aligned} \min\ & (x_1-3)^2+(x_2-2)^2 \\ s.t.\ & -x_1^2-x_2^2+5\ge 0 \\ & x_1+2x_2-4=0 \\ & x_1,x_2\ge 0. \end{aligned}$$

§ 4.1 最优性条件

设$x^* = (2, 1)^T$，KKT条件如下：

$$\begin{cases} \nabla f(x^*) - \gamma_1 \nabla g_1(x^*) - \mu_1 \nabla h_1(x^*) = 0 \\ \gamma_1 g_1(x^*) = 0 \\ \gamma_1 \geq 0,\ i = 1, 2. \\ g_1(x^*) \geq 0 \\ h_1(x^*) = 0 \end{cases}$$

其中，$\nabla f(x) = (2x_1 - 6,\ 2x_2 - 4)^T \quad \nabla g_1(x) = (-2x_1,\ -2x_2)^T \quad \nabla h_1(x) = (1,\ 2)^T$

因此，有

$$\begin{cases} 2x_1 - 6 + 2\gamma_1 x_1 - \mu_1 = 0 \\ 2x_2 - 4 + 2\gamma_1 x_2 - 2\mu_1 = 0 \\ \gamma_1(-x_1^2 - x_2^2 + 5) = 0 \\ \gamma_1 \geq 0 \\ -x_1^2 - x_2^2 + 5 \geq 0 \\ x_1 + 2x_2 - 4 = 0 \end{cases} \xrightarrow{\text{代入}x^*}$$

$\gamma_1 = 1/3 > 0$ √

$\mu_1 = -2/3$

$g_1(x^*) = 0$ √

$h_1(x^*) = 0$ √

因此，x^*是KKT点。

例3 给定非线性规划问题

$$\begin{aligned} \min \quad & f(x) = (x_1 - 1)^2 + x_2 \\ s.t. \quad & g_1(x) = -x_1 - x_2 + 2 \geq 0 \\ & g_2(x) = x_2 \geq 0 \end{aligned}$$

求满足KKT条件的点。

§ 4.2 可行方向法

求解带有约束条件的非线性规划问题的方法：

- 将非线性规划化为近似的线性规划（可行方向法）
- 将约束问题化为一个或一系列的无约束优化问题（惩罚函数法）
- 将复杂的问题变为较简单的问题（序列二次规划法）

§ 4.2 可行方向法

考虑非线性规划问题

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & Ax \ge b \\ & Ex = e \end{aligned}$$

约束为线性函数

其中，$f(x)$是可微函数， $A_{m\times n}$, $E_{l\times n}$, $x \in R^n$, $b \in R^m$, $e \in R^l$

$D = \{x \mid Ax \ge b, Ex = e\}$ --- 可行域

$x^{k+1} = x^k + \lambda_k d^k$ $\longrightarrow$ 下降方向 $f\left(x^{(k)} + \lambda_k d^{(k)}\right) < f\left(x^{(k)}\right)$

可行方向 $x^k + \lambda_k d^k \in D$

§ 4.2 可行方向法

下降方向：

设$x\in R^n$，若对于非零向量d，存在$\delta>0$，使$\forall\lambda\in(0,\delta)$均有$f(x+\lambda d)<f(x)$，则$d$为$f$在$x$处的下降方向。

可行方向：

设D为可行域，$x\in D$，若对于非零向量d，存在$\delta>0$，使得对$\forall\lambda\in(0,\delta)$均有$x+\lambda d\in D$，则$d$为从$x$出发的可行方向。

可行方向法：

从可行点x^k出发，寻找一下降的可行方向d^k作为搜索方向，然后沿此方向确定移动步长，得下一迭代点x^{k+1}。

§ 4.2 可行方向法

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & Ax \ge b \qquad \textbf{(1)} \\ & Ex = e \end{aligned}$$

确定下降可行方向

在可行点x^k处, $A = [A_1, A_2]^T$, $b = [b_1, b_2]^T$,

且 $A_1x^k > b_1$, $A_2x^k = b_2$, $Ex^k = e$.

不起作用约束　起作用约束

令$x = x^k + d$, 则f 在x^k处的一阶泰勒展开为

$$f(x) \approx f(x^k) + \nabla f^T(x^k)d$$

则问题(1)变为

$$\begin{aligned} \min \quad & f(x^k) + \nabla f^T(x^k)d \\ s.t. \quad & A_1(x^k + d) \ge b_1 \\ & A_2(x^k + d) \ge b_2 \\ & E(x^k + d) = e \end{aligned} \longrightarrow \begin{aligned} \min \quad & \nabla f^T(x^k)d \\ s.t. \quad & A_2 d \ge 0 \qquad \textbf{(2)} \\ & Ed = 0 \end{aligned}$$

§ 4.2 可行方向法

$$\begin{aligned} \min \quad & \nabla f^T(x^k)d \\ s.t. \quad & A_2 d \ge 0 \\ & Ed = 0 \end{aligned} \qquad \textbf{(2)}$$

$d=0$为问题(2)的可行点，且目标函数值为0，所以若d^*为最优解，则一定有$\nabla f^T(x^k)d^* \le 0$

当 $\nabla f^T(x^k)d^* < 0$ **时，则d^*为问题(1)在x^k处的下降可行方向。**

当 $\nabla f^T(x^k)d^* = 0$ **时，则任意满足$A_2 d \ge 0, Ed = 0$的d均有**

$\nabla f^T(x^k)d \ge 0$ **，此时d^*为可行方向。**

因此，可通过求解线性规划

$$\begin{aligned} \min \quad & \nabla f^T(x^k)d \\ s.t. \quad & A_2 d \ge 0 \\ & Ed = 0 \\ & -1 \le d_i \le 1 \end{aligned} \qquad \textbf{(3)}$$

得到搜索方向(下降可行方向)。

§ 4.2 可行方向法

定理：设x^k是问题(1)的一个可行解，且$A_1x^k > b_1$，$A_2x^k = b_2$，其中$A = [A_1, A_2]^T$, $b = [b_1, b_2]^T$

则下来结论成立：

- **非零向量d是点x^k处的可行方向当且仅当$A_2d \geq 0, Ed = 0$**
- **若向量d又满足：$\nabla f^T(x^k)d < 0$，则d是一个可行下降方向。**

§ 4.2 可行方向法

确定步长

为使$f(x^{k+1})$的值尽可能小，则λ越大越好，但需保证$x^{k+1} \in D$，所以,$0 \le \lambda \le \lambda_{\max}$.

则通过求解一维搜索问题：

$$\min_{0 \le \lambda \le \lambda_{\max}} f(x^k + \lambda d^k) \qquad \textbf{(4)}$$

得到步长λ^k.

§ 4.2 可行方向法

确定$\lambda_{\max}$

$$
\begin{aligned}
\min\quad & f(x^k+\lambda d^k) \\
s.t.\quad & A(x^k+\lambda d^k)\ge b \\
& E(x^k+\lambda d^k)=e
\end{aligned}
\qquad \textbf{(5)}
$$

因d^k是可行方向，$Ed^k=0$. 又x^k为可行点，有$Ex^k=e$.

所以，等式约束一定满足。

另外，不等式约束可写成

$$A_1(x^k+\lambda d)\ge b_1 \qquad \text{(i)}$$

$$A_2(x^k+\lambda d)\ge b_2 \qquad \text{(ii)}$$

因d^k是可行方向，$A_2x^k=b_2$，$A_2d^k\ge 0$，$\lambda\ge 0$，所以(ii)也成立。

则(5)式可简化为

$$\begin{aligned} \min\ & f(x^k+\lambda d^k) \\ s.t.\ & A_1x^k+\lambda A_1d^k \geq b_1 \qquad \textbf{(iii)} \\ & \lambda \geq 0 \end{aligned}$$

(iii)式可改写为 $\lambda \underbrace{A_1 d^k}_{v} \geq \underbrace{b_1 - A_1 x^k}_{u}$

$\boldsymbol{A_1 x^k > b_1}$

$\boldsymbol{v}$ $\boldsymbol{u\ (u<0)}$

因此，有 $\lambda \boldsymbol{v} \geq \boldsymbol{u}$

$$\lambda_{\max} = \begin{cases} \min\left\{\dfrac{u_i}{v_i}\right\}, & v_i < 0 \\ \infty, & v_i \geq 0 \end{cases} \qquad \textbf{(6)}$$

§ 4.2 可行方向法

可行方向法的求解步骤：

1. 选定初始点x^0，精度$\varepsilon > 0$，令$k = 0$.

2. 在点x^k处，根据起作用约束和不起作用约束，将A和b分别分解为$A = [A_1, A_2]^T, b = [b_1, b_2]^T$, 使得$A_1 x^k > b_1, A_2 x^k = b_2$.

3. 求解线性规划问题

$$\begin{aligned} \min\ & \nabla f^T(x^k)d \\ s.t.\ & A_2 d \geq 0 \\ & Ed = 0 \\ & -1 \leq d_i \leq 1 \end{aligned}$$

得到最优解d^k.

4. 如果$|\nabla f^T(x^k)d^k| < \varepsilon$，则停止迭代，$x^k$为所求最优点；否则，转5.

§ 4.2 可行方向法

5. 若$A_1 d^k \geq 0$，则从x^k出发，沿d^k方向，求λ_k，使$f(x^k+\lambda_k d^k) = \min f(x^k+\lambda d^k)$. 令$x^{k+1}=x^k+\lambda_k d^k$，转7；否则转6.

6. 令$v=A_1 d^k$, $u=b_1-A_1 x^k$，计算$\lambda_{\max}=\min\left\{\dfrac{u_i}{v_i}\,\middle|\, v_i<0\right\}$

求解$\min\limits_{0\leq\lambda\leq\lambda_{\max}} f(x^k+\lambda d^k)$，得到最优的$\lambda_k$，令$x^{k+1}=x^k+\lambda_k d^k$，转7.

7. 确定x^{k+1}的起作用约束，修正A_1, A_2和b_1, b_2，令$k=k+1$，返回2.

§ 4.2 可行方向法

例1：利用可行方向法求解：

$$
\begin{aligned}
\min\ & f(x)=(x_1-6)^2+(x_2-2)^2 \\
s.t.\ & x_1-2x_2\ge -4 \\
& -3x_1-2x_2\ge -12 \\
& x_1\ge 0 \\
& x_2\ge 0
\end{aligned}
$$

取$x^0=(2,3)^T$, $\varepsilon=0.001$

§ 4.2 可行方向法

解：x^0 代入约束条件 $g_1(x^0) = -4$

$g_2(x^0) = -12$

$g_3(x^0) = 2 > 0$

$g_4(x^0) = 3 > 0$

因此，$g_1(x)$和$g_2(x)$为x^0处的起作用约束

故有

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad b_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1 & -2 \\ -3 & -2 \end{bmatrix} \qquad b_2 = \begin{bmatrix} -4 \\ -12 \end{bmatrix}$$

§ 4.2 可行方向法

确定搜索方向$d^0=(d_1, d_2)^T$

$\nabla f(x)=[2(x_1-6), 2(x_2-2)]^T$ $\nabla f(x^0)=[-8, 2]^T$

则线性规划为

$$\begin{aligned}
\min\ & -8d_1+2d_2 \\
s.t.\ & d_1-2d_2\ge 0 \\
& -3d_1-2d_2\ge 0 \\
& -1\le d_1, d_2\le 1
\end{aligned}$$

解得 $d_1=2/3$，$d_2=-1$

因$|\nabla f^T(x^0)d^0|=\dfrac{22}{3}>\varepsilon$，**所以$x^0$不是最优解**

§ 4.2 可行方向法

确定步长λ_0

$$v = A_1 d^0 = (\frac{2}{3}, -1)^T \qquad u = b_1 - A_1 x^0 = (-2, -3)^T$$

则 $\lambda_{\max} = \min\left\{\frac{-3}{-1}\right\} = 3$

求解

$$\min_{0 \le \lambda \le 3} f(x^0 + \lambda d^0) = \frac{13}{9}(\lambda - \frac{33}{13})^2 + \frac{100}{13}$$

解得 $\lambda_0 = 33/13$

故 $x^1 = x^0 + \lambda_0 d^0 = (48/13, 6/13)^T$

§ 4.2 可行方向法

进一步，将x^1代入约束条件

$$g_1(x^1) = 36/13 > -4$$

$$g_2(x^1) = -12$$

$$g_3(x^1) = 48/13 > 0$$

$$g_4(x^1) = 6/13 > 0$$

因此，$g_2(x)$为x^1处的起作用约束

故有

$$A_1 = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad b_1 = \begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} -3 & -2 \end{bmatrix} \qquad b_2 = -12$$

§ 4.2 可行方向法

确定搜索方向$d^1=(d_1, d_2)^T$

$\nabla f(x)=[2(x_1-6), 2(x_2-2)]^T$　　　$\nabla f(x^1)=[-\frac{60}{13}, -\frac{40}{13}]^T$

则线性规划为

$$\min -\frac{60}{13}d_1+\frac{40}{13}d_2$$
$$s.t.\quad -3d_1-2d_2\ge 0$$
$$-1\le d_1, d_2\le 1$$

解得 $d_1=2/3$，$d_2=-1$

因 $|\nabla f^T(x^1)d^1|=0<\varepsilon$ **，所以$x^1=(48/13, 6/13)^T$为最优解。**

§ 4.2 可行方向法

例2：考虑下列问题：

$$
\begin{aligned}
\min\ & x_1^2 + x_1x_2 + 2x_2^2 - 6x_1 - 2x_2 - 12x_3 \\
s.t.\ \ & -x_1 + 2x_2 \le 3 \\
& x_1 + x_2 + x_3 = 2 \\
& x_1, x_2, x_3 \ge 0
\end{aligned}
$$

求出在点$\boldsymbol{x}' = (1, 1, 0)^T$处的一个下降可行方向.

§ 4.2 可行方向法

解：先将问题化为

$$\begin{aligned} \min\ & x_1^2 + x_1 x_2 + 2x_2^2 - 6x_1 - 2x_2 - 12x_3 \\ s.t.\ & x_1 - 2x_2 \ge -3 \\ & x_1 + x_2 + x_3 = 2 \\ & x_1, x_2, x_3 \ge 0 \end{aligned}$$

进一步，将x'代入约束条件

$$g_1(x') = -1 > -3$$

$$g_2(x') = 2$$

$$g_3(x') = 1 > 0$$

$$g_4(x') = 1 > 0$$

$$g_5(x') = 0$$

因此，$g_2(x)$和$g_5(x)$为x'处的起作用约束。

§ 4.2 可行方向法

相应地，有

$$A_1=\begin{bmatrix}-1 & 2 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix} \qquad b_1=\begin{bmatrix}-3\\ 0\\ 0\end{bmatrix}$$

$$A_2=\begin{bmatrix}0 & 0 & 1\end{bmatrix} \qquad b_2=0$$

设$\boldsymbol{d}=(\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3)^T$ 为下降可行方向，其应满足如下条件：

$$\begin{cases} d_3\ge 0 & (A_2 d\ge 0)\\ d_1+d_2+d_3=0 & (Ed=0)\\ -3d_1+3d_2-12d_3<0 & (\nabla f^T(x')d<0)\end{cases}$$

如 $\boldsymbol{d}=(0, -1, 1)^T$ 满足以上条件。

§ 4.3 惩罚函数法

约束非线性规划的数学模型

$$\begin{aligned} &\min\ f(x) \\ &s.t.\ \ h_j(x)=0,\ \ j=1,\cdots,l \\ &\qquad g_i(x)\ge 0,\ \ i=1,\cdots,m \\ &\qquad x\in R^n. \end{aligned}$$

可行域 $D=\{x\,|\,x\in R^n, h_j(x)=0, g_i(x)\ge 0, j=1,\cdots,l, i=1,\cdots,m\}$.

基本思想：根据约束的特点构造某种惩罚函数，使得对约束问题的求解，转换为一系列无约束问题的求解。

§ 4.3 惩罚函数法

(一) 外点法

基本思想：对违反约束的点在目标函数中进行相应的“惩罚”，而对可行点不予惩罚。

对于只含等式约束的问题

$$\begin{aligned} &\min \ f(x) \\ &s.t. \ \ h_j(x)=0, \ j=1,\cdots,l \end{aligned}$$

定义惩罚函数为 $\varphi(x,\sigma)=f(x)+\boxed{\sigma\sum_{j=1}^{l}h_j^2(x)}$

$\sigma \geq 0$ 惩罚因子　　　　　　　　惩罚项

§ 4.3 惩罚函数法

当 $x \notin D$，即 $h_j(x) \neq 0$，则有 $h_j^2(x) > 0$.

故若σ >0 (很大)，说明 $\varphi(x,\sigma)$的极小点不会是违反约束的*x*.

因此，将求解

$$\begin{aligned} &\min \quad f(x) \\ &s.t. \quad h_j(x) = 0, \ \ j = 1, \cdots, l \end{aligned}$$

转化为对

$$\min_x \varphi(x,\sigma) = f(x) + \sigma \sum_{j=1}^{l} h_j^2(x)$$

的求解。

§ 4.3 惩罚函数法

对于只含不等式约束的问题

$$\begin{aligned} &\min \ f(x) \\ &s.t. \ \ g_i(x) \geq 0 \\ &\qquad x \in R^n \end{aligned}$$

当 $x \in D$，$g_i(x) \geq 0$，希望惩罚项为0.

当 $x \notin D$，$g_i(x) < 0$，希望惩罚项大于0.

构造惩罚项为 $\sigma\, [\max\{0, -g_i(x)\}]^2$

此时，有 $\varphi(x,\sigma) = f(x) + \sigma \sum_{i=1}^{m} [\max\{0, -g_i(x)\}]^2$

且 $\varphi(x,\sigma) = \begin{cases} f(x), & x \in D \\ f(x) + \text{很大的正数}, & x \notin D \end{cases}$

当σ很大

至少存在i_1，使$g_{i_1}(x) < 0$，从而有罚项 $\geq \sigma\, [g_{i_1}(x)]^2$

§ 4.3 惩罚函数法

设原问题的最优解为x^*，$\min\limits_{x} \varphi(x,\sigma)$ 的最优解为$x^*(\sigma)$

讨论x^*与$x^*(\sigma)$的关系

(i) 若$x^*(\sigma)\in D$，则是原问题的最优解。

(ii) 若 $x^*(\sigma)\notin D$，则σ充分大时，$x^*(\sigma)$会很靠近可行域的边界，是原问题的近似最优解。

如何选取σ，使$x^*(\sigma)$是所需的最优解？

通过迭代逐步增大σ，令$\sigma_{k+1}=c\sigma_k$

$c>1,\ c=5\sim10$

§ 4.3 惩罚函数法

对于同时包含等式及不等式约束的问题

$$\begin{aligned} &\min \ f(x) \\ &s.t. \ \ h_j(x)=0,\ j=1,\cdots,l \\ &\qquad g_i(x)\ge 0,\ i=1,\cdots,m \end{aligned}$$

定义惩罚函数为

$$\varphi(x,\sigma)=f(x)+\sigma\sum_{i=1}^{m}[\max\{0,-g_i(x)\}]^2+\sigma\sum_{j=1}^{l}h_j^2(x)$$

设原问题的最优解为***x****，$\min\limits_{x}\varphi(x,\sigma)$ 的最优解为***x****(**σ**)

则有 $x^*(\sigma)\notin D \xrightarrow{\sigma\to+\infty} \boldsymbol{x}^*$

解析法

§ 4.3 惩罚函数法

外点法的求解步骤：

1. 给定初始点 x^0，初始惩罚因子σ_1，放大系数$c>1$，精度要求$\varepsilon>0$，令$k=1$.

2. 以 $x^{k\text{-}1}$为初始点，令$\sigma_k P(x)=\sigma_k\sum_{i=1}^{m}[\max\{0,-g_i(x)\}]^2+\sigma_k\sum_{j=1}^{l}h_j^2(x)$
求解无约束问题

$$\min f(x)+\sigma_k P(x)$$

得到极小点 x^k.

3. 若$\sigma_k P(x^k)<\varepsilon$，则停止计算，得到近似解 x^k；否则，令$\sigma_{k+1}=c\sigma_k$，$k=k+1$，返回2.

§ 4.3 惩罚函数法

例：利用外点法求解下列问题：

$$\min\ f(x)=(x_1-1)^2+x_2^2$$
$$s.t.\quad g(x)=x_2\geq 1.$$

定义惩罚函数为

$$\varphi(x,\sigma)=f(x)+\sigma\sum_{i=1}^{m}[\max\{0,-g_i(x)\}]^2$$
$$=(x_1-1)^2+x_2^2+\sigma[\max\{0,-(x_2-1)\}]^2$$
$$=\begin{cases}(x_1-1)^2+x_2^2, & x_2\geq 1\\ (x_1-1)^2+x_2^2+\sigma(x_2-1)^2, & x_2<1\end{cases}$$

§ 4.3 惩罚函数法

利用解析法求解 $\min \varphi(x,\sigma)$，有

$$\frac{\partial\varphi}{\partial x_1}=2(x_1-1) \qquad \frac{\partial\varphi}{\partial x_2}=\begin{cases}2x_2, & x_2\ge 1\\ 2x_2+2\sigma(x_2-1), & x_2<1\end{cases}$$

令

$$\frac{\partial\varphi}{\partial x_1}=0, \quad \frac{\partial\varphi}{\partial x_2}=0$$

得到 $\boldsymbol{x^*=(1,\ 0)^T}$ **(舍)**

$$x^*=\begin{bmatrix}1\\ \dfrac{\sigma}{1+\sigma}\end{bmatrix}\xrightarrow{\sigma\to\infty}\begin{bmatrix}1\\ 1\end{bmatrix}$$

故 $\boldsymbol{x^*=(1,\ 1)^T}$ 为问题的最优解。

§ 4.3 惩罚函数法

例2：利用外点法求解下列问题：

$$\min\ f(x)=x_1^2+x_2^2$$
$$s.t.\quad h(x)=x_1+x_2-1=0.$$

§ 4.3 惩罚函数法

(二) 内点法

基本思想：从可行域的内点出发，并在可行域内部进行搜索。因此，只适用于求解不等式约束问题。

$$\min \ f(x)$$
$$s.t. \quad g_i(x) \geq 0, \ i=1,\cdots,m.$$

可行域为 $D=\{x \mid g_i(x)\geq 0, i=1,\cdots,m\}$.

当 $\boldsymbol{x \in D}$ 内部，$\boldsymbol{g_i(x)>0}$，希望惩罚项很小.

当 $\boldsymbol{x \to D}$ 边界，存在$\boldsymbol{g_{i_1}(x)=0}$，希望惩罚项很大.

构造惩罚项为

$$rB(x)=r\sum_{i=1}^{m}\frac{1}{g_i(x)} \qquad rB(x)=-r\sum_{i=1}^{m}\log g_i(x)$$

$\boldsymbol{r\ (r>0)}$ 是一个很小的数

§ 4.3 惩罚函数法

此时，定义障碍函数

$$\varphi(x,r)=f(x)+rB(x)$$

且 $$\varphi(x,r)=\begin{cases} f(x)+\text{有限的数}, & x\in D \text{ 内部} \\ +\infty, & x\to D \text{ 边界} \end{cases}$$

希望障碍函数最优解**$x^*(r)$**→原问题最优解**x^***

在D的内部 　　　　在D的边界上

$$\varphi(x^*(r),r)=f(x^*(r))\boxed{-r\sum_{i=1}^{m}\log g_i(x^*(r))}\to f(x^*)$$

$$\sum_{i=1}^{m}\log g_i(x^*(r))\to-\infty \Longrightarrow \quad r\to 0$$

$$\boxed{-r\sum_{i=1}^{m}\log g_i(x^*(r))}\to 0$$

$$rB(x)=r\sum_{i=1}^{m}\frac{1}{g_i(x)} \qquad rB(x)=-r\sum_{i=1}^{m}\log g_i(x)$$

§ 4.3 惩罚函数法

设原问题的最优解为x^*，$\min\limits_{x\in D}\varphi(x,r)$ 的最优解为$x^*(r)$

当$r\to 0$，则有 $x^*(r)\to x^*$

如何选取r，使$x^*(r)$是所需的最优解？

通过迭代逐步减小r，令$r_{k+1}=cr_k$

$0 < c < 1$

§ 4.3 惩罚函数法

内点法的计算步骤：

1. 给定初始点$x^0 \in D$, 精度$\varepsilon > 0$, 取$r_1 > 0$, $0 < c < 1$, 令$k = 1$.

2. 以 x^{k-1} 为初始点，求$\min\limits_{x \in D} \varphi(x, r_k)$的最优解，记为$x^k$，转3.

3. 若$r_k B(x^k) < \varepsilon$, 则停止计算，得到近似解 x^k；否则，令$r_{k+1} = c r_k$，$k = k+1$，返回2.

§ 4.3 惩罚函数法

例：用内点法求解下列问题：

$$\begin{aligned} &\min\ f(x)=\frac{1}{12}(x_1+1)^3+x_2 \\ &s.t.\quad g_1(x)=x_1-1\ge 0, \\ &\qquad\quad g_2(x)=x_2\ge 0. \end{aligned}$$

定义障碍函数为

$$\varphi(x,r)=\frac{1}{12}(x_1+1)^3+x_2+r(\frac{1}{x_1-1}+\frac{1}{x_2})$$

利用解析法求解 $\min\limits_{x\in D}\varphi(x,r_k)$，令

$$\frac{\partial\varphi}{\partial x_1}=\frac{1}{4}(x_1+1)^2-\frac{r}{(x_1-1)^2}=0 \qquad\qquad \frac{\partial\varphi}{\partial x_2}=1-\frac{r}{x_2^2}=0$$

§ 4.3 惩罚函数法

解得

$$x^* = \begin{bmatrix} \sqrt{2\sqrt{r}+1} \\ \sqrt{r} \end{bmatrix} \xrightarrow{r\to 0} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

故 $x^* = (1, 0)^T$ 为问题的最优解。

§ 4.3 惩罚函数法

(三) 乘子法

考虑只有等式约束的情形：

$$
\begin{aligned}
&\min \ f(x) \\
&s.t. \quad h_j(x)=0,\ j=1,\cdots,l.
\end{aligned}
\qquad (1)
$$

转化为

$$\min_x \varphi(x,\sigma)=f(x)+\sigma\sum_{j=1}^{l}h_j^2(x) \qquad (2)$$

设x^*是问题(1)的最优解，那么x^*是否为问题(2)的最优解？

分析：若x^*是问题(2)的最优解，则有$\nabla\varphi_x(x^*,\sigma)=0$

而$\nabla\varphi_x(x^*,\sigma)=\nabla f(x^*)+2\sigma\sum_{j=1}^{l}\boxed{h_j(x^*)}\nabla h_j(x^*)=\nabla f(x^*)\neq 0$ 矛盾

故x^*不是问题(2)的最优解。

§ 4.3 惩罚函数法

如何改进罚函数，使x^*也是其最优解？

$$\varphi(x,\sigma,\mu)=f(x)-\underbrace{\boxed{\sum_{j=1}^{l}\mu_j h_j(x)}}_{\text{乘子项}}+\underbrace{\boxed{\frac{\sigma}{2}\sum_{j=1}^{l}h_j^2(x)}}_{\text{惩罚项}} \qquad \textbf{(3)}$$

若x^*是问题(1)的最优解，则x^*是问题(3)的最优解。

证明：**x^*是问题(1)的最优解，则有** $\nabla f(x^*)-\sum_{j=1}^{l}\mu_j\nabla h_j(x^*)=0$

而 $$\nabla\varphi_x(x^*,\sigma,\mu)=\underbrace{\boxed{\nabla f(x^*)-\sum_{j=1}^{l}\mu_j\nabla h_j(x^*)}}_{\mathbf{0}}+\sigma\sum_{j=1}^{l}\underbrace{\boxed{h_j(x^*)}}_{\mathbf{0}}\nabla h_j(x^*)=0$$

得证.

§ 4.3 惩罚函数法

如何确定μ和σ?

(i) 一般σ取充分大的数

(ii) 给定$\sigma > 0$足够大，已有μ^k, 求解 $\min \varphi(x,\sigma,\mu^k)$ 得最优解x^k

则有 $\nabla \varphi_x(x^k,\sigma,\mu^k) = \nabla f(x^k) - \sum_{j=1}^{l} \mu_j^k \nabla h_j(x^k) + \sigma \sum_{j=1}^{l} h_j(x^k) \nabla h_j(x^k) = 0$

$$\nabla f(x^k) - \sum_{j=1}^{l} (\mu_j^k - \sigma h_j(x^k)) \nabla h_j(x^k) = 0$$

又知 $\nabla f(x^*) - \sum_{j=1}^{l} \mu_j^* \nabla h_j(x^*) = 0$

故有 $\mu_j^{k+1} = \mu_j^k - \sigma h_j(x^k),\ j = 1,\cdots,l.$

接下来再求解 $\min \varphi(x,\sigma,\mu^{k+1})$，可得到$x^{k+1}$，重复以上过程，得到两个点列$\{\mu^k\}$, $\{x^k\}$.

乘子法的求解步骤：

1. 给定初始点 x^0，惩罚因子σ，$\mu^1 > 0$，精度$\varepsilon > 0$，$c > 0$ ($c \in [2, 10]$)，$0 < \beta < 1$，令$k = 1$.

2. 以 x^{k-1}为初始点，求$\min \varphi(x, \sigma, \mu^k)$的最优解，记为$x^k$，

3. 若 $\|h(x^k)\| < \varepsilon$，则停止计算，得到近似解 x^k；否则，转4.

4. 若$\dfrac{\| h(x^k) \|}{\| h(x^{k-1}) \|} \geq \beta$，令$\sigma = c\sigma$，转4；否则，直接转5.

5. 计算$\mu_j^{k+1} = \mu_j^k - \sigma h_j(x^k)$，令$k = k+1$，返回2.

§ 4.3 惩罚函数法

例：用乘子法求解下列问题：

$$\min\ f(x) = 2x_1^2 + x_2^2 - 2x_1x_2$$
$$s.t.\quad h(x) = x_1 + x_2 - 1 = 0.$$

定义惩罚函数为

$$\varphi(x,\sigma,\mu) = 2x_1^2 + x_2^2 - 2x_1x_2 - \mu(x_1 + x_2 - 1) + \frac{\sigma}{2}(x_1 + x_2 - 1)^2$$

取$\sigma = 2$，在第k次迭代，利用解析法求解$\min\varphi(x,\sigma,\mu^k)$，有

$$\nabla\varphi_{x_1}(x,\sigma,\mu^k) = 4x_1 - 2x_2 - \mu^k + 2(x_1 + x_2 - 1) = 0$$
$$\nabla\varphi_{x_2}(x,\sigma,\mu^k) = 2x_2 - 2x_1 - \mu^k + 2(x_1 + x_2 - 1) = 0$$

则

$$x^k = \begin{bmatrix} \dfrac{\mu^k + 2}{6} \\ \dfrac{\mu^k + 2}{4} \end{bmatrix}$$

§ 4.3 惩罚函数法

因 $\mu^{k+1}=\mu^k-\sigma h(x^k)$

代入 $x^k=\begin{bmatrix}\dfrac{\mu^k+2}{6}\\ \dfrac{\mu^k+2}{4}\end{bmatrix}$，有 $\mu^{k+1}=\dfrac{1}{6}\mu^k+\dfrac{1}{3}$

当$k\to\infty$ 时，有 $\mu^*=\dfrac{1}{6}\mu^*+\dfrac{1}{3} \quad \to \quad \mu^*=\dfrac{2}{5}$

则 $x^k=\begin{bmatrix}\dfrac{\mu^k+2}{6}\\ \dfrac{\mu^k+2}{4}\end{bmatrix} \xrightarrow{k\to\infty} \begin{bmatrix}\dfrac{2}{5}\\ \dfrac{3}{5}\end{bmatrix}$

§ 4.3 惩罚函数法

考虑只有不等式约束的情形：

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

利用等式约束的结果，引入变量y_i，将上式化为等式约束

$$\begin{aligned} &\min \ f(x) \\ &s.t. \quad g_i(x) - y_i^2 = 0,\ i = 1, \cdots, m. \end{aligned}$$

则惩罚函数为

$$\tilde{\varphi}(x, y, \sigma, \mu) = f(x) - \sum_{i=1}^{m} \mu_i (g_i(x) - y_i^2) + \frac{\sigma}{2} \sum_{i=1}^{m} (g_i(x) - y_i^2)^2$$

$$\tilde{\varphi}(x, y, \sigma, \mu) = f(x) + \sum_{i=1}^{m} \left\{ \frac{\sigma}{2} \left[y_i^2 - \frac{1}{\sigma}(\sigma g_i(x) - \mu_i) \right]^2 - \frac{\mu_i^2}{2\sigma} \right\}$$

§ 4.3 惩罚函数法

$$\tilde{\varphi}(x,y,\sigma,\mu)=f(x)+\sum_{i=1}^{m}\left\{\frac{\sigma}{2}\left[y_i^2-\frac{1}{\sigma}(\sigma g_i(x)-\mu_i)\right]^2-\frac{\mu_i^2}{2\sigma}\right\}$$

为使$\tilde{\varphi}(x,y,\sigma,\mu)$ 取极小，有

$$y_i^2=\frac{1}{\sigma}[\max\{0,\sigma g_i(x)-\mu_i\}]^2$$

代入y_i^2 ，因此，有

$$\tilde{\varphi}(x,\sigma,\mu)=f(x)+\frac{1}{2\sigma}\sum_{i=1}^{m}\{[\max(0,\mu_i-\sigma g_i(x))]^2-\mu_i^2\}$$

§ 4.3 惩罚函数法

对于同时含有等式及不等式约束的情形：

$$
\begin{aligned}
&\min \ f(x) \\
&s.t. \quad h_j(x)=0, \ j=1,\cdots,l \\
&\qquad\ \ g_i(x)\geq 0, \ i=1,\cdots,m.
\end{aligned}
$$

可以定义惩罚函数为

$$
\varphi(x,\sigma,\mu,\lambda)=f(x)+\frac{1}{2\sigma}\sum_{i=1}^{m}\{[\max(0,\lambda_i-\sigma g_i(x))]^2-\lambda_i^2\}
$$

$$
-\sum_{j=1}^{l}\mu_j h_j(x)+\frac{\sigma}{2}\sum_{j=1}^{l}h_j^2(x)
$$

其中

$$
\mu_j^{k+1}=\mu_j^k-\sigma h_j(x^k), \ j=1,\cdots,l.
$$

$$
\lambda_i^{k+1}=\max(0,\lambda_i^k-\sigma g_i(x^k)), \ i=1,\cdots,m.
$$

§ 4.3 惩罚函数法

例：用乘子法求解下列问题：

$$\min\ x_1^2 + x_2^2$$
$$s.t.\quad x_1 \ge 1$$

定义惩罚函数为

$$\varphi(x,\sigma,\mu) = f(x) + \frac{1}{2\sigma}\{[\max(0,\mu-\sigma g(x))]^2 - \mu^2\}$$

$$= x_1^2 + x_2^2 + \frac{1}{2\sigma}\{[\max(0,\mu-\sigma(x_1-1))]^2 - \mu^2\}$$

$$= \begin{cases} x_1^2 + x_2^2 - \dfrac{\mu^2}{2\sigma}, & x_1 - 1 \ge \dfrac{\mu}{\sigma} \\ x_1^2 + x_2^2 + \dfrac{1}{2\sigma}\{[\mu-\sigma(x_1-1)]^2 - \mu^2\}, & x_1 - 1 < \dfrac{\mu}{\sigma} \end{cases}$$

§ 4.3 惩罚函数法

利用解析法求解 $\min\varphi(x,\sigma,\mu)$，有

$$\nabla\varphi_{x_1}(x,\sigma,\mu)=\begin{cases}2x_1, & x_1-1\geq\dfrac{\mu}{\sigma}\\[2ex] 2x_1-\mu+\sigma(x_1-1), & x_1-1<\dfrac{\mu}{\sigma}\end{cases}$$

$$\nabla\varphi_{x_2}(x,\sigma,\mu)=2x_2$$

在第k次迭代，令 $\nabla\varphi_{x_1}(x,\sigma,\mu^k)=0,\quad \nabla\varphi_{x_2}(x,\sigma,\mu^k)=0$

得到

$$x^k=\begin{bmatrix}\dfrac{\mu^k+\sigma}{2+\sigma}\\ 0\end{bmatrix}$$

§ 4.3 惩罚函数法

因 $\mu^{k+1}=\max(0,\mu^k-\sigma g(x^k))=\mu^k-\sigma g(x^k)$

代入$x^k=\begin{bmatrix}\dfrac{\mu^k+\sigma}{2+\sigma}\\ 0\end{bmatrix}$，有 $\mu^{k+1}=\dfrac{2\mu^k+2\sigma}{2+\sigma}$

当$k\to\infty$ 时，有 $\mu^*=\dfrac{2\mu^*+2\sigma}{2+\sigma}\quad\to\quad\mu^*=2$

则 $x^k=\begin{bmatrix}\dfrac{\mu^k+\sigma}{2+\sigma}\\ 0\end{bmatrix}\quad\xrightarrow{k\to\infty}\quad\begin{bmatrix}1\\0\end{bmatrix}$

故 $\boldsymbol{x}^*=(1,0)^T$ 为问题的最优解。

§ 4.3 惩罚函数法

外点法：初始点可在R^n空间中任意选取。

惩罚因子很大时，求解惩罚函数相对困难，且迭代的中间结果往往不是可行解，不能作为原问题的近似最优解。

内点法：初始点是可行域的内点，难以选取。

迭代在可行域内部进行，中间的迭代结果可作为近似解。

乘子法：无需惩罚因子 $\sigma \to \infty$

惩罚函数的极小点为原问题的最优解。

§ 4.4 序列二次规划法

$$
\begin{aligned}
\min\ & f(x) \\
s.t.\ & h_j(x)=0,\ j=1,\cdots,l \\
& g_i(x)\ge 0,\ i=1,\cdots,m \\
& x\in R^n.
\end{aligned}
$$

其中，f, g_i, h_j均为实值连续函数，具有二阶连续偏导数。

基本思想：在迭代点x^k处，用二次规划模型来替代约束非线性规划问题，通过求解一系列二次规划的解来逼近原问题的解。

§ 4.4 序列二次规划法

计算步骤：

1. 给定初始点x^0, 精度$\varepsilon > 0$, 令$k=0$.

2. 求在点 x^k 处二次规划模型的最优解，记为d^k，转3.

3. 若$\|d^k\| < \varepsilon$, 则停止计算，得到近似解 x^k；否则，令 $x^{k+1} = x^k + d^k$，$k = k+1$，返回2.

§ 4.4 序列二次规划法

$$\begin{aligned} \min\ & f(x) \\ s.t.\ & h_j(x)=0,\ j=1,\cdots,l \\ & g_i(x)\ge 0,\ i=1,\cdots,m \\ & x\in R^n. \end{aligned}$$

如何将原问题化为二次规划模型？

二次规划：目标函数为二次实函数，约束为线性函数。

令$x=x^k+d$，那么，在点x^k处的二次规划模型为：

$$\begin{aligned} \min\ & \frac{1}{2}d^TQ_kd+\nabla f^T(x^k)d & \text{..... } f(x)\text{的二阶泰勒展开} \\ s.t.\ & \nabla h_j^T(x^k)d+h_j(x^k)=0,\ j=1,\cdots,l & \text{...... } h_j(x)\text{的一阶泰勒展开} \\ & \nabla g_i^T(x^k)d+g_i(x^k)\ge 0,\ i=1,\cdots,m. & \text{...... } g_i(x)\text{的一阶泰勒展开} \end{aligned}$$

一般Q_k为$f(x)$在点 x^k处的海塞矩阵

§ 4.4 序列二次规划法

例：考虑非线性规划

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

取$x^0=(0,0)^T$, 列出其在x^0处的二次规划模型.

解：令 $f(x)=(x_1-2)^2+x_2^2$ $\quad g_1(x)=(x_1-1)^2+(x_2-1)^2-1$

$g_2(x)=x_1$ $\quad g_3(x)=x_2$

$\nabla f(x)=(2x_1-4,\ 2x_2)^T$ $\quad \nabla f(x^0)=(-4,\ 0)^T$ $\quad \nabla^2 f(x)=\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

$\nabla g_1(x)=(2x_1-2,\ 2x_2-2)^T$ $\quad \nabla g_1(x^0)=(-2,\ -2)^T$

$\nabla g_2(x)=(1,\ 0)^T$ $\quad \nabla g_3(x)=(0,\ 1)^T$

§ 4.4 序列二次规划法

设$d=(d_1, d_2)^T$，令$x=x^0+d$，那么，在点x^0处的二次规划模型为：

$$\begin{aligned}
\min\ & \frac{1}{2}(d_1,d_2)\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}d_1\\ d_2\end{pmatrix}+(-4\quad 0)\begin{pmatrix}d_1\\ d_2\end{pmatrix}\\
s.t.\ & (-2,-2)\begin{pmatrix}d_1\\ d_2\end{pmatrix}+1\geq 0\\
& d_1,d_2\geq 0
\end{aligned}$$

即

$$\begin{aligned}
\min\ & d_1^2+d_2^2-4d_1\\
s.t.\ & -2d_1-2d_2+1\geq 0\\
& d_1,d_2\geq 0
\end{aligned}$$

§ 4.4 序列二次规划法

如何求解二次规划问题?

考虑等式约束二次规划问题---Lagrange方法

$$\min \ \frac{1}{2}x^T Q x + c^T x$$
$$s.t. \quad A^T x = b$$

定义Lagrange函数

$$L(x,\mu) = \frac{1}{2}x^T Q x + c^T x - \mu^T (A^T x - b)$$

根据一阶必要条件，有

$$\frac{\partial L(x,\mu)}{\partial x} = Qx + c - A\mu = 0$$

$$\frac{\partial L(x,\mu)}{\partial \mu} = -A^T x + b = 0$$

$$\Longrightarrow \begin{bmatrix} Q & -A \\ -A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ \mu \end{bmatrix} = \begin{bmatrix} -c \\ -b \end{bmatrix}$$

§ 4.4 序列二次规划法

若$\begin{bmatrix} Q & -A \\ -A^T & 0 \end{bmatrix}$可逆，可直接求解得到最优解$x$*和最优的

Lagrange乘子μ*.

因一阶必要条件，也是KKT条件，所以在计算中也可直接求解方程组得到最优解。

§ 4.4 序列二次规划法

例：利用Lagrange方法求解下列问题

$$\min \ 2x_1^2 + x_2^2 + x_1x_2 - x_1 - x_2$$
$$s.t. \quad x_1 + x_2 = 1$$

解：定义Lagrange函数

$$L(x,\mu) = 2x_1^2 + x_2^2 + x_1x_2 - x_1 - x_2 - \mu^T(x_1 + x_2 - 1)$$

$$\frac{\partial L(x,\mu)}{\partial x_1} = 4x_1 + x_2 - 1 - \mu = 0$$

$$\frac{\partial L(x,\mu)}{\partial x_2} = x_1 + 2x_2 - 1 - \mu = 0 \longrightarrow x^* = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{3}{4} \end{bmatrix}$$

$$\frac{\partial L(x,\mu)}{\partial \mu} = -(x_1 + x_2 - 1) = 0$$

$$\mu^* = \frac{3}{4}$$

§ 4.4 序列二次规划法

考虑不等式约束二次规划问题---有效集方法

$$\min \ \frac{1}{2}x^T Qx + c^T x \qquad \textbf{(1)}$$
$$s.t. \quad a_i^T x \ge b_i, \ i = 1, \cdots, m.$$

定义Lagrange函数

$$L(x,\mu) = \frac{1}{2}x^T Qx + c^T x - \sum_{i=1}^{m} \mu_i (a_i^T x - b_i)$$

设x*为问题的最优解，则必有

$$Qx^* + c - \sum_{i=1}^{m} \mu_i a_i = 0 \longrightarrow \text{计算}\mu_i$$

$$\mu_i \ge 0.$$

§ 4.4 序列二次规划法

设在迭代点x处的起作用约束集为$W=\{i \mid a_i^T x=b_i\}$，则可通过求解等式约束二次规划问题

$$\begin{aligned} \min\ & \frac{1}{2}x^T Qx+c^T x \\ s.t.\ & a_i^T x=b_i,\ i\in W. \end{aligned} \qquad \textbf{(2)}$$

得到问题(1)的最优解。

当(2)的解是(1)的可行点，则还需判断相应的μ_i是否满足$\mu_i \geq 0$. 若满足，则停止计算；否则，去掉一个约束，重新求解新的等式约束二次规划问题。

如何增加或减少约束呢？

当(2)的解不是(1)的可行点，则需增加约束，然后求解所得到的等式约束二次规划问题。

§ 4.4 序列二次规划法

设第k次迭代点x^k处的起作用约束集为W_k，令$x = x^k + d$

有 $$\frac{1}{2}x^T Qx + c^T x = \frac{1}{2}(x^k + d)^T Q(x^k + d) + c^T (x^k + d)$$

$$= \underline{\frac{1}{2}(x^k)^T Qx^k} + \underline{(x^k)^T Qd} + \frac{1}{2}d^T Qd + \underline{c^T x^k} + \underline{c^T d}$$

$Qx^k + c = \nabla f(x^k)$

$$= \frac{1}{2}d^T Qd + \nabla f^T (x^k)d + C$$

且 $a_i^T x = b_i \rightarrow a_i^T (x^k + d) = b_i \rightarrow a_i^T d = 0 \ (i \in W_k)$

故问题(2)可简化为问题(3)

$$\begin{aligned} &\min \ \frac{1}{2}d^T Qd + \nabla f^T (x^k)d \\ &s.t. \quad a_i^T d = 0, \ i \in W_k. \end{aligned} \qquad (3)$$

§ 4.4 序列二次规划法

$$\min \ \frac{1}{2}d^TQd + g^Td$$
$$s.t. \quad a_i^Td = 0, \ i \in W_k. \qquad \textbf{(3)}$$

根据问题(3), 若x^k不是最优解，分析如何增加或减少约束来得到新的等式约束二次规划问题，即确定W_{k+1}.

设d^k为(3)的解，μ_i^k 为相应的乘子。

(i) 当$d^k = 0$，则x^k是问题(2)的最优解。若$\mu_i^k \geq 0$，则停止计算；否则，令 $\mu_j^k = \min\{\mu_i^k < 0, i \in W_k\}$，则去掉 μ_j^k 所对应的约束，有 $W_{k+1} = W_k \setminus j$

§ 4.4 序列二次规划法

(ii)当$d^k \neq 0$，令$x^{k+1} = x^k + \alpha_k d^k$，使 $a_i^T(x^k + \alpha_k d^k) = b_i$成立。

若$x^k + d^k$是可行点，则令$x^{k+1} = x^k + d^k$，即$\alpha_k = 1$.

若$x^k + d^k$不是可行点，则令$x^{k+1} = x^k + \alpha_k d^k$，需求解满足条件的$\alpha_k$.

当$i \in W_k$ 时，满足约束，即 $a_i^T(x^k + \alpha_k d^k) = b_i$ 总是成立的。

当 $i \notin W_k$ 时，若 $a_i^T d^k \geq 0$，则对所有的$\alpha_k \geq 0$，有 $a_i^T(x^k + \alpha_k d^k) \geq b_i$

若 $a_i^T d^k < 0$，仅当 $\alpha_k \leq \dfrac{b_i - a_i^T x^k}{a_i^T d^k}$ 时，有 $a_i^T(x^k + \alpha_k d^k) \geq b_i$

故取 $\alpha_k = \min\left\{1, \min\left\{\dfrac{b_i - a_i^T x^k}{a_i^T d^k} \,\middle|\, i \notin W_k, a_i^T d^k < 0\right\}\right\}$

当 $\alpha_k < 1$时，将x^{k+1}处的起作用约束j，加入到W_k中，形成$W_{k+1} = W_k \cup \{j\}$.

§ 4.4 序列二次规划法

有效集方法的计算步骤：

1. 给点初始可行点x^0，令x^0处的起作用集为W_0，$k=0$.

2. 求解问题(3)，得到d^k.

3. 若$d^k=0$，计算相应的μ_i^k。若$\mu_i^k\geq 0$，停止计算，得到最优解$x^*=x^k$；否则，确定j_k，令$x^{k+1}=x^k$, $W_{k+1}=W_k\backslash\{j_k\}$，转5.

4. 若$d^k\neq 0$，计算α_k，令$x^{k+1}=x^k+\alpha_k d^k$。若$\alpha^k<1$，确定x^{k+1}处的起作用约束j，并加入到W_k中，得到$W_{k+1}=W_k\cup\{j\}$；否则，令$W_{k+1}=W_k$.

5. 令$k=k+1$，转2.

§ 4.4 序列二次规划法

例：利用有效集方法求解下列问题：

$$\begin{aligned}
\min\ & (x_1-1)^2+(x_2-2.5)^2 \\
s.t.\ & x_1-2x_2+2\geq 0 \\
& -x_1-2x_2+6\geq 0 \\
& -x_1+2x_2+2\geq 0 \\
& x_1,x_2\geq 0
\end{aligned}$$